THIRD EDITION

FUNDAMENTAL STATISTICS FOR THE SOCIAL AND BEHAVIORAL SCIENCES

Howard T. **TOKUNAGA**



To Jennifer , my kids Meagan and Eika , and my parents, Katsumi and Grayce Tokunaga.

As a global academic publisher, Sage is driven by the belief that research and education are critical in shaping society. Our mission is building bridges to knowledge—supporting the development of ideas into scholarship that is certified, taught, and applied in the real world.

Sage's founder, Sara Miller McCune, transferred control of the company to an independent trust, which guarantees our independence indefinitely. This enables us to support an equitable academic future over the long term by building lasting relationships, championing diverse perspectives, and co-creating social and behavioral science resources that transform teaching and learning.

▶ Long Description:

FUNDAMENTAL STATISTICS FOR THE SOCIAL AND BEHAVIORAL SCIENCES

Third Edition

Howard T. Tokunaga

San Jose State University



eISBN: 978-1-0718-6160-8

Copyright © 2027 by Sage.

All rights reserved. Except as permitted by U.S. copyright law, no part of this work may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without permission in writing from the publisher.

All third party trademarks referenced or depicted herein are included solely for the purpose of illustration and are the property of their respective owners. Reference to these trademarks in no way indicates any relationship with, or endorsement by, the trademark owner.



FOR INFORMATION:

2455 Teller Road, Thousand Oaks, California 91320, E-mail: orders@sagepub.com

3rd floor, HYLO, 103-105 Bunhill Row, London, EC1Y 8LZ, United Kingdom

10th Floor, Emaar Capital Tower 2, MG Road Sikanderpur, Sector 26, Gurugram, Haryana – 122002, India

18 Cross Street #10-10/11/12, China Square Central, Singapore 048423

Library of Congress Control Number: 2026943277

ISBN: 978-1-0718-6144-8

Printed in the United States of America

This book is printed on acid-free paper.

26 27 28 29 30 10 9 8 7 6 5 4 3 2 1

Acquisitions Editor: Anna Villarruel

Content Development Editor: Kenzie Offley

Editorial Assistant: Daria Terry

Production Editor: Syeda Aina Rahat Ali

Copy Editor: Michelle Ponce

Typesetter: Lumina Datamatics

Cover Designer: Candice Harman

Marketing Manager: Victoria Velasquez

CONTENTS

Preface
Acknowledgments
About the Author
Chapter 1 Introduction to Statistics
Chapter 2 Examining Data: Tables and Figures
Chapter 3 Measures of Central Tendency
Chapter 4 Measures of Variability
Chapter 5 Normal Distributions
Chapter 6 Probability and Introduction to Hypothesis Testing
Chapter 7 Testing One Sample Mean
Chapter 8 Estimating the Mean of a Population
Chapter 9 Testing the Difference Between Two Means
Chapter 10 Errors in Hypothesis Testing, Statistical Power, and Effect Size
Chapter 11 One-Way Analysis of Variance (ANOVA)
Chapter 12 Two-Way Analysis of Variance (ANOVA)
Chapter 13 Correlation
Chapter 14 Linear Regression and Multiple Correlation
Chapter 15 Chi-Square
Appendix A: Tables
Appendix B: Review of Basic Mathematics
Glossary
References
Index

DETAILED CONTENTS

- Preface
- Acknowledgments
- About the Author
- Chapter 1 Introduction to Statistics
  - 1.1 What Is Statistics?
  - 1.2 Why Learn Statistics?
  - 1.3 Introduction to the Stages of the Research Process
    - Developing a Research Hypothesis to Be Tested
      - Identifying a Question or Issue to Be Examined
      - Reviewing and Evaluating Relevant Theories and Research
      - Stating a Research Hypothesis: Independent and Dependent Variables
    - Collecting Data
      - Drawing a Sample From a Population
      - Determining How Variables Will Be Measured: Levels of Measurement
      - Selecting a Method to Collect the Data: Experimental and Nonexperimental Research Methods
    - Analyzing the Data
      - Calculating Descriptive Statistics
      - Calculating Inferential Statistics
    - Drawing a Conclusion Regarding the Research Hypothesis
    - Communicating the Findings of the Study
  - 1.4 Plan of the Book
  - 1.5 Looking Ahead
  - 1.6 Summary
  - 1.7 Important Terms
  - 1.8 Exercises
  - Answers to Learning Checks
  - Learning Check 1
  - Learning Check 2

Answers to Odd-Numbered Exercises
Chapter 2 Examining Data: Tables and Figures
2.1 An Example From the Research: Winning the Lottery
2.2 Why Examine Data?
Gaining an Initial Sense of the Data
Identifying Outliers
Evaluating Research Methodology
Determining Whether Data Meet Statistical Criteria and Assumptions
2.3 Examining Data Using Tables
What Are the Values of the Variable?
How Many Participants in the Sample Have Each Value of the Variable?
What Percentage of the Sample Has Each Value of the Variable?
2.4 Grouped Frequency Distribution Tables
Cumulative Percentages
2.5 Examining Data Using Figures
Displaying Nominal and Ordinal Variables: Bar Charts and Pie Charts
Displaying Interval and Ratio Variables: Histograms and Frequency Polygons
Drawing Inappropriate Conclusions From Figures
2.6 Examining Data: Describing Distributions
Modality
Symmetry
Variability
2.7 Looking Ahead
2.8 Summary
2.9 Important Terms
2.10 Formula Introduced in This Chapter
Percentage (%)
2.11 Using IBM SPSS Software *
Creating Data Files: Defining Variables and Entering Data
Examining Data: Frequency Distribution Tables and Figures: The Lottery Study (2.1)
2.12 Exercises

Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 3 Measures of Central Tendency
3.1 An Example From the Research: The 10% Myth
3.2 Understanding Central Tendency
3.3 The Mode
Determining the Mode
3.4 The Median
Calculating the Median: Odd Number of Scores
Calculating the Median: Even Number of Scores
Calculating the Median When Multiple Scores Have the Median Value
3.5 The Mean
Calculating the Mean
Reporting Calculated Numbers
Rounding Calculated Numbers
Calculating the Mean From Frequency Distribution Tables
The Mean as a Balancing Point in a Distribution
The Sample Mean vs. the Population Mean: Statistics vs. Parameters
3.6 Comparison of the Mode, Median, and Mean
Strengths and Weaknesses of the Mode
Strengths and Weaknesses of the Median
Strengths and Weaknesses of the Mean
Choosing a Particular Measure of Central Tendency
3.7 Measures of Central Tendency: Drawing Conclusions
3.8 Looking Ahead
3.9 Summary
3.10 Important Terms
3.11 Formulas Introduced in This Chapter
Median (Odd Number of Scores)
Median (Even Number of Scores)

Mean
Mean (Frequency Distribution Table)
Mean (Balancing Point in a Distribution)
3.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 4 Measures of Variability
4.1 An Example From the Research: How Many Sometimes" in an Always"?
4.2 Understanding Variability
4.3 The Range
Strengths and Weaknesses of the Range
4.4 The Interquartile Range
Strengths and Weaknesses of the Interquartile Range
4.5 The Variance (s^2)
Definitional Formula for the Variance
Computational Formula for the Variance
Why Not Use the Absolute Value of the Deviation in Calculating the Variance?
Why Divide by N − 1 Rather Than N in Calculating the Variance?
4.6 The Standard Deviation (s)
Definitional Formula for the Standard Deviation
Computational Formula for the Standard Deviation
Eyeball Estimating" the Mean and Standard Deviation
4.7 Measures of Variability for Populations
The Population Variance (σ^2)
The Population Standard Deviation (σ)
Measures of Variability for Samples vs. Populations
4.8 Measures of Variability for Nominal Variables
The Index of Qualitative Variation
4.9 Measures of Variability: Drawing Conclusions

4.10 Looking Ahead
4.11 Summary
4.12 Important Terms
4.13 Formulas Introduced in This Chapter
Range
Interquartile Range
Variance (s^2) (Definitional Formula)
Variance (s^2) (Computational Formula)
Standard Deviation (s) (Definitional Formula)
Standard Deviation (s) (Computational Formula)
Population Variance (σ^2) (Definitional Formula)
Population Standard Deviation (σ) (Definitional Formula)
Index of Qualitative Variation (IQV)
4.14 Using SPSS
Calculating Measures of Central Tendency and Variability: The Frequency Rating Study (4.1)
4.15 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 5 Normal Distributions
5.1 Normal Distributions
What Are Normal Distributions?
Why Are Normal Distributions Important?
Is There Only One Normal Distribution?
5.2 The Standard Normal Distribution
What Is the Standard Normal Distribution?
Interpreting z-Scores: Distance From the Mean
Interpreting z-Scores: Relative Position Within the Distribution
Finding the Area Between the Mean and a z-Score

Finding the Area Less or Greater Than a z-Score
Finding the Area Between Two z-Scores
5.3 Applying z-Scores to Normal Distributions
Transforming Normal Distributions Into the Standard Normal Distribution
Understanding the z-Score Formula
Example: Transforming SAT Scores Into z-Scores
Interpreting z-Scores: Distance From the Mean
Interpreting z-Scores: Relative Position Within Distribution
Finding the Area Between the Mean and a Score
Finding the Area Less or Greater Than a Score
Finding the Area Between Two Scores
Transforming Scores Into z-Scores: Drawing Conclusions
5.4 Standardizing Frequency Distributions
Example: Measuring Politicians' Behavior
Standardizing a Frequency Distribution
Interpreting Standardized Scores
Using Standardized Scores to Compare or Combine Variables
Characteristics of Standardized Distributions
When Can You Apply the Normal Curve Table to a Standardized Distribution?
5.5 Looking Ahead
5.6 Summary
5.7 Important Terms
5.8 Formulas Introduced in This Chapter
z-Score
Standardized Score
5.9 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 6 Probability and Introduction to Hypothesis Testing

6.1 A Brief Introduction to Probability
- What Is Probability?
- Why Is Probability Important to Researchers?
- Applying Probability to Normal Distributions
- Applying Probability to Binomial Distributions

6.2 Example: Making Heads or Tails of the Super Bowl

6.3 Introduction to Hypothesis Testing
- State the Null and Alternative Hypotheses (H_0 and H_1)
  - The Null Hypothesis (H_0)
  - The Alternative Hypothesis (H_1)
- Make a Decision About the Null Hypothesis
  - Defining a Low" Probability of a Statistic: Setting Alpha (α)
  - Identifying the Values of the Statistic With a Low Probability
  - Stating a Decision Rule
  - Calculating a Value of a Statistic
  - Making the Decision Whether to Reject the Null Hypothesis
- Draw a Conclusion From the Analysis
- Relate the Result of the Analysis to the Research Hypothesis
- Summary of Steps in Hypothesis Testing

6.4 Issues Related to Hypothesis Testing: An Introduction
- The Issue of Proof" in Hypothesis Testing
- Errors in Decision Making
- Factors Influencing the Decision About the Null Hypothesis
  - Sample Size
  - Alpha (α)
  - Directionality of the Alternative Hypothesis

6.5 Looking Ahead

6.6 Summary

6.7 Important Terms

6.8 Formula Introduced in This Chapter

Probability

6.9 Exercises

Answers to Learning Checks

Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 7 Testing One Sample Mean
7.1 An Example From the Research: Do You Read Me?
7.2 The Sampling Distribution of the Mean
Characteristics of the Sampling Distribution of the Mean
7.3 Inferential Statistics: Testing One Sample Mean (σ Known)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: z-Test for One Mean
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the z-Test for One Mean
Summary
7.4 A Second Example From the Research: Unique Invulnerability
7.5 Introduction to the t-Distribution
Characteristics of the Student t-Distribution
7.6 Inferential Statistics: Testing One Sample Mean (σ Not Known)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: t-Test for One Mean
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the t-Test for One Mean

Summary
7.7 Factors Affecting the Decision About the Null Hypothesis
Sample Size
Alpha (α)
Directionality of the Alternative Hypothesis
7.8 Looking Ahead
7.9 Summary
7.10 Important Terms
7.11 Formulas Introduced in This Chapter
z-Test for One Mean
Population Standard Error of the Mean $\left(\mathbf{i}t\sigma_{\mathbf{i}t\overline{X}}\right)$
Degrees of Freedom (df) for One Sample
t-Test for One Mean
Standard Error of the Mean $\left(\mathbf{i}ts_{\mathbf{i}t\overline{X}}\right)$
7.12 Using SPSS
Testing One Sample Mean (σ Not Known): The Unique Invulnerability Study (7.6)
7.13 Exercises
Answers to Learning Checks
Learning Check 2
Learning Check 3
Learning Check 5
Learning Check 6
Answers to Odd-Numbered Exercises
Chapter 8 Estimating the Mean of a Population
8.1 An Example From the Research: Salary Survey
8.2 Introduction to the Confidence Interval for the Mean
Using the Sampling Distribution of the Mean to Create the Confidence Interval for the Mean
8.3 The Confidence Interval for the Mean (σ Not Known)
State the Desired Level of Confidence
Calculate the Confidence Interval and Confidence Limits
Calculate the Standard Error of the Mean $\left(s_{\overline{X}}\right)$

Calculate the Degrees of Freedom (df) and Identify the Critical Value of t
Calculate the Confidence Interval and Confidence Limits
Draw a Conclusion About the Confidence Interval
Probability of an Interval or the Probability of a Population Mean?
Summary
8.4 The Confidence Interval for the Mean (σ Known)
State the Desired Level of Confidence
Calculate the Confidence Interval and Confidence Limits
Calculate the Population Standard Error of the Mean $\left(\sigma_{\overline{X}}\right)$
Identify the Critical Value of z
Calculate the Confidence Interval and Confidence Limits
Draw a Conclusion About the Confidence Interval
Summary
8.5 Factors Affecting the Width of the Confidence Interval for the Mean
Sample Size
Estimating the Sample Size Needed for a Desired Confidence Interval
Level of Confidence
Choosing Between Different Levels of Confidence: Probability Versus Precision
8.6 Interval Estimation and Hypothesis Testing
Concerns About Hypothesis Testing
The Interpretation of Statistical Significance and Nonsignificance
The Influence of Sample Size on the Decision About the Null Hypothesis
Interval Estimation as an Alternative to Hypothesis Testing
8.7 Looking Ahead
8.8 Summary
8.9 Important Terms
8.10 Formulas Introduced in This Chapter
Confidence Interval for the Mean (σ Not Known)
Confidence Interval for the Mean (σ Known)
Estimated Sample Size for the Confidence Interval for the Mean

8.11 Using SPSS
Calculating the confidence interval for the mean (The salary survey study (8.1))
8.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 9 Testing the Difference Between Two Means
9.1 An Example From the Research: You Can Just Wait
9.2 The Sampling Distribution of the Difference
Characteristics of the Sampling Distribution of the Difference
9.3 Inferential Statistics: Testing the Difference Between Two Sample Means
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: t-Test for Independent Means
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the t-Test for Independent Means
Summary
9.4 Inferential Statistics: Testing the Difference Between Two Sample Means (Unequal Sample Sizes)
An Example From the Research: Kids' Motor Skills and Fitness
Inferential Statistics: Testing the Difference Between Two Sample Means (Unequal Sample Sizes)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis

Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the t-Test for Independent Means (Unequal Sample Sizes)
Summary
9.5 Inferential Statistics: Testing the Difference Between Paired Means
Research Situations Appropriate for Within-Subjects Research Designs
An Example From the Research: Web-Based Family Interventions
Inferential Statistics: Testing the Difference Between Paired Means
Calculate the Difference Between the Paired Scores
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the t-Test for Dependent Means
Summary
9.6 Looking Ahead
9.7 Summary
9.8 Important Terms
9.9 Formulas Introduced in This Chapter
Degrees of Freedom (df), Difference Between Two Sample Means
t-Test for Independent Means
Standard Error of the Difference $\left(s_{\overline{X}_1-\overline{X}_2}\right)$
Standard Error of the Difference (Unequal Sample Sizes)
Degrees of Freedom (df), Difference Between Paired Means
t-Test for Dependent Means
Standard Error of Difference Scores (s D)
9.10 Using SPSS
Testing the Difference Between Two Sample Means: The Parking Lot Study (9.1)
Testing the Difference Between Paired Means: The Web-Based Intervention Study (9.5)
9.11 Exercises
Answers to Learning Checks
Learning Check 1

Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Answers to Odd-Numbered Exercises
Chapter 10 Errors in Hypothesis Testing, Statistical Power, and Effect Size
10.1 Hypothesis Testing vs. Criminal Trials
10.2 An Example From the Research: Truth or Consequences
Correct Decisions and Mistakes in Criminal Trials
10.3 Two Errors in Hypothesis Testing: Type I and Type II Error
Type I Error: The Risk in Rejecting the Null Hypothesis
What Is the Probability of Making a Type I Error?
What Is the Probability of Not Making a Type I Error?
Why Is Type I Error a Concern?
Type II Error: The Risk in Not Rejecting the Null Hypothesis
What Is the Probability of Making a Type II Error?
What Is the Probability of Not Making a Type II Error?
Why Is Type II Error a Concern?
Why Does Type II Error Occur?
Type I and Type II Error: Summary
10.4 Controlling Type I and Type II Error
Controlling Type I Error
Concerns About Controlling Type I Error
Controlling Type II Error
Increasing Sample Size
Raising Alpha (α)
Using a Directional Alternative Hypothesis (H1)
Increasing Between-Group Variability
Decreasing Within-Group Variability
Controlling Type I and Type II Error: Summary
10.5 Measures of Effect Size
Interpreting Inferential Statistics: The Influence of Sample Size
The Inappropriateness of “Highly Significant”

Effect Size as Variance Accounted for

Measure of Effect Size for the Difference Between Two Sample Means: r^2

Interpreting Measures of Effect Size

Presenting Measures of Effect Size

Reasons for Calculating Measures of Effect Size

A Second Measure of Effect Size for the Difference Between Two Sample Means: Cohen's d

10.6 Looking Ahead

10.7 Summary

10.8 Important Terms

10.9 Formulas Introduced in This Chapter

Probability of Type I Error

Probability of Type II error

r Squared (r^2)

Cohen's d

10.10 Exercises

Answers to Learning Checks

Learning Check 1

Learning Check 2

Learning Check 3

Learning Check 4

Answers to Odd-Numbered Exercises

Chapter 11 One-Way Analysis of Variance (ANOVA)

11.1 An Example From the Research: It's Your Move

11.2 Introduction to Analysis of Variance (ANOVA)

Review: Testing the Difference Between the Means of Two Samples

Understanding Between-Group and Within-Group Variability

Relating Between-Group and Within-Group Variability

Introduction to Analysis of Variance (ANOVA) and the F-Ratio

Characteristics of the F-Ratio Distribution

11.3 Inferential Statistics: One-Way Analysis of Variance (ANOVA)

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Value, and State a Decision Rule
Calculate a Statistic: F-Ratio for the One-Way ANOVA
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (R^2)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Summary
11.4 A Second Example: The Parking Lot Study Revisited
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Value, and State a Decision Rule
Calculate a Statistic: F-Ratio for the One-Way ANOVA
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (R^2)
Draw a Conclusion From the Analysis
The Relationship Between the t-Test and the F-Ratio
11.5 Analytical Comparisons Within the One-Way ANOVA
Planned Versus Unplanned Comparisons
Between-Group and Within-Group Variance in Analytical Comparisons
Conducting Planned Comparisons
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Conducting Unplanned Comparisons
Concerns Regarding Unplanned Comparisons
Methods of Controlling Familywise Error

Controlling Familywise Error: The Tukey Test
11.6 Looking Ahead
11.7 Summary
11.8 Important Terms
11.9 Formulas Introduced in This Chapter
Degrees of Freedom for Between-Group Variance, One-Way ANOVA (df_{BG})
Degrees of Freedom for Within-Group Variance, One-Way ANOVA (df_{WG})
Between-Group Variance, One-Way ANOVA (MS_{BG})
Total Sample Mean, One-Way ANOVA $\left(\overline{X}_T\right)$
Within-Group Variance, One-Way ANOVA (MS_{WG})
F-Ratio, One-Way ANOVA (F)
Measure of Effect Size, One-Way ANOVA (R^2)
Relationship Between t-Test and F-Ratio for the One-Way ANOVA
F-Ratio, Analytical Comparison (F_{comp})
Between-Group Variance, Analytical Comparison $\left(MS_{\mathrm{BG}_{\mathrm{comp}}}\right)$
Measure of Effect Size, Analytical Comparison $\left(R^2_{\mathrm{comp}}\right)$
Critical Value, Tukey Test for Unplanned Comparisons (F_T)
11.10 Using SPSS
One-Way Analysis of Variance (ANOVA): The Chess Study (11.1)
Analytical Comparisons: Planned Comparisons—Explanation vs. Prediction (11.5)
11.11 Exercises
Answers to Learning Checks
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 12 Two-Way Analysis of Variance (ANOVA)
12.1 An Example From the Research: Vote—or Else!
12.2 Introduction to Factorial Research Designs
Testing Interaction Effects

An Example of Interaction Effect
An Example of No Interaction Effect
Testing Main Effects
The Relationship Between Main Effects and Interaction Effects
Factorial Designs vs. Single-Factor Designs: Advantages and Disadvantages
12.3 The Two-Factor (A × B) Research Design
Notational System for Components of the Two-Factor Research Design
Analyzing the Two-Factor Research Design: Descriptive Statistics
Notational System for Means in the Two-Factor Research Design
Examining the Descriptive Statistics
12.4 Introduction to Analysis of Variance (ANOVA) for the Two-Factor Research Design
F-Ratios in the Two-Factor Research Design
12.5 Inferential Statistics: Two-Way Analysis of Variance (ANOVA)
State the Null and Alternative Hypotheses (H_0 and H_1)
Main Effect of Factor A
Main Effect of Factor B
A × B Interaction Effect
Make Decisions About the Null Hypotheses
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State Decision Rules
Calculate Statistics: F-Ratios for the Two-Way ANOVA
Make Decisions Whether to Reject the Null Hypotheses
Determine the Levels of Significance
Calculate Measures of Effect Size (R^2)
Draw Conclusions From the Analyses
Relate the Results of the Analyses to the Research Hypothesis
12.6 Investigating a Significant A × B Interaction Effect: Analysis of Simple Effects
12.7 Looking Ahead
12.8 Summary
12.9 Important Terms
12.10 Formulas Introduced in This Chapter

Degrees of Freedom for Main Effect of Factor A, Two-Way ANOVA (df_A)
Degrees of Freedom for Main Effect of Factor B, Two-Way ANOVA (df_B)
Degrees of Freedom for A × B Interaction Effect, Two-Way ANOVA ($df_{A \times B}$)
Degrees of Freedom for Within-Group Variance, Two-Way ANOVA (dfWG)
Between-Group Variance for Main Effect of Factor A, Two-Way ANOVA (MS_A)
Between-Group Variance for Main Effect of Factor B, Two-Way ANOVA (MS_B)
Sums of Squares for Between-Group Variance, Two-Way ANOVA (SS_{BG})
Between-Group Variance for A × B Interaction, Two-Way ANOVA ($MS_{A \times B}$)
Within-Group Variance, Two-Way ANOVA (MS_{WG})
F-Ratio for Main Effect of Factor A, Two-Way ANOVA (F_A)
F-Ratio for Main Effect of Factor B, Two-Way ANOVA (F_B)
F-Ratio for A × B Interaction, Two-Way ANOVA ($F_{A \times B}$)
Measure of Effect Size for Main Effect of Factor A, Two-Way ANOVA(R^2_A)
Measure of Effect Size for Main Effect of Factor B, Two-Way ANOVA(R^2_B)
Measure of Effect Size for A × B Interaction, Two-Way ANOVA(R^2_{AB})
12.11 Using SPSS
Two-Way Analysis of Variance (ANOVA): The Voting Message Study (12.1)
12.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 13 Correlation
13.1 An Example From the Research: It's Good for You!
13.2 Introduction to the Concept of Correlation
Describing the Relationship Between Variables
Nature of the Relationship
Direction of the Relationship
Strength of the Relationship

Measuring the Relationship Between Variables: Correlational Statistics
Introduction to the Pearson Correlation Coefficient (r)
Calculating Correlational Statistics: The Role of Variance and Covariance
13.3 Inferential Statistics: Pearson Correlation Coefficient
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: Pearson Correlation Coefficient (r)
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (r^2)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the Pearson Correlation
A Computational Formula for the Pearson r
Represent the Covariance Between X and Y (SPXY)
Represent the Variance of Variable X (SSX)
Represent the Variance of Variable Y (SSY)
Calculate the Pearson Correlation Coefficient (r)
13.4 Correlating Two Sets of Ranks: The Spearman Rank-Order Correlation
Introduction to Ranked Variables
An Example From the Research: Do Students and Teachers Think Alike?
Inferential Statistics: The Spearman Rank-Order Correlation Coefficient
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
13.5 Correlational Statistics vs. Correlational Research
Does ANOVA = Causality?
“Correlation Is Not Causality”—What Does That Mean?
13.6 Looking Ahead

13.7 Summary
13.8 Important Terms
13.9 Formulas Introduced in This Chapter
Degrees of Freedom for Pearson Correlation (df)
Sum of Products (Definitional Formula) (SPXY)
Sum of Squares for the X Variable (Definitional Formula) (SSX)
Sum of Squares for the Y Variable (Definitional Formula) (SSY)
Pearson Correlation (Definitional Formula) (r)
Measure of Effect Size for the Pearson Correlation Coefficient (r^2)
Sum of Products (Computational Formula) (SPXY)
Sum of Squares for the X Variable (Computational Formula) (SSX)
Sum of Squares for the Y Variable (Computational Formula) (SSY)
Pearson Correlation (Computational Formula) (r)
Spearman Rank-Order Correlation (rs)
Measure of Effect Size for the Spearman Rank-Order Correlation $\left(r_s^2\right)$
13.10 Using SPSS
Pearson Correlation: The Picky Eater Study (13.1)
Spearman Rank-Order Correlation: The Business Student Study (13.4)
13.11 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Answers to Odd-Numbered Exercises
Chapter 14 Linear Regression and Multiple Correlation
14.1 Predicting One Variable From Another: Linear Regression
The Linear Regression Equation
Calculating the Linear Regression Equation
Calculate the Slope of the Equation (b)
Calculate the Y-Intercept of the Equation (a)

Report the Linear Regression Equation
Drawing the Linear Regression Equation
A Second Example: Put Down Your Pencils
Calculating the Linear Regression Equation
Drawing the Linear Regression Equation
14.2 Correlation With Two or More Predictors: Introduction to Multiple Correlation and Regression
The Limitations of Using Only One Predictor
Introduction to Multiple Correlation
Multiple Correlation With Two Predictors: An Example
Measuring the Relationship Between Two Predictors and a Criterion: The Multiple Correlation (R)
Testing the Multiple Correlation for Statistical Significance
Predicting a Criterion From Two Predictors: The Multiple Regression Equation
Conclusion
14.3 Looking Ahead
14.4 Summary
14.5 Important Terms
14.6 Formulas Introduced in This Chapter
Linear Regression Equation
Estimate of Slope (b)
Estimate of Y-Intercept (a)
Multiple Correlation Coefficient (R)
F-Ratio for the Multiple Correlation Coefficient (F)
Multiple Regression Equation
Regression Coefficient
Estimate of Y-Intercept (a) (Multiple Regression)
14.7 Using SPSS
Linear Regression: The Picky Eater Study (14.1)
Multiple Regression: The College Admission Example (14.2)
14.8 Exercises
Answers to Learning Checks
Learning Check 1

Learning Check 2

Answers to Odd-Numbered Exercises

Chapter 15 Chi-Square

15.1 An Example From the Research (One Categorical Variable): Are You My Type?

15.2 Introduction to the Chi-Square Statistic

Observed and Expected Frequencies: The Basis of the Chi-Square Statistic

Characteristics of the Distribution of Chi-Square Statistics

15.3 Inferential Statistic: Chi-Square Goodness-of-Fit Test

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Value, and State a Decision Rule

Calculate a Statistic: Chi-Square (χ^2)

Make a Decision Whether to Reject the Null Hypothesis

Determine the Level of Significance

Calculate a Measure of Effect Size (Cramér's φ)

Draw a Conclusion From the Analysis

Relate the Result of the Analysis to the Research Hypothesis

Assumptions Underlying the Chi-Square Statistic

Conducting the Chi-Square Goodness-of-Fit Test With Unequal Hypothesized Proportions

Calculate Expected Frequencies (f_e)

Calculate the Chi-Square Statistic (χ^2)

15.4 An Example From the Research (Two Categorical Variables): Seeing Red

15.5 Inferential Statistic: Chi-Square Test of Independence

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Value, and State a Decision Rule

Calculate a Statistic: Chi-square (χ^2)

Make a Decision Whether to Reject the Null Hypothesis

Determine the Level of Significance
Calculate a Measure of Effect Size (Cramér's φ)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
15.6 Parametric and Nonparametric Statistical Tests
Parametric vs. Nonparametric Statistical Tests
Reasons for Using Nonparametric Statistical Tests
Examples of Nonparametric Statistical Tests
15.7 Looking Ahead
15.8 Summary
15.9 Important Terms
15.10 Formulas Introduced in This Chapter
Degrees of Freedom for the Chi-Square Goodness-of-Fit Test
Expected Frequency for the Chi-Square Goodness-of-Fit Test
Chi-Square Statistic (χ^2)
Cramér's φ for the Chi-Square Goodness-of-Fit Test
Degrees of Freedom for the Chi-Square Test of Independence
Expected Frequency for the Chi-Square Test of Independence
Cramér's φ for the Chi-Square Test of Independence
15.11 Using SPSS
Chi-Square Goodness-of-Fit Test—Equal Hypothesized Proportions: The Personality Study (15.1)
Chi-Square Goodness-of-Fit Test—Unequal Hypothesized Proportions: The Personality Study (15.3)
Chi-Square Test of Independence: The Color Preference Study (15.4)
Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Learning Check 6

Answers to Odd-Numbered Exercises

Appendix A: Tables

Appendix B: Review of Basic Mathematics

I. Symbols

II. Terms

III. Addition, Subtraction, Multiplication, Division

A. All Positive Numbers

B. Adding All Negative Numbers

C. Adding a Combination of Positive and Negative Numbers

D. Subtracting Negative Numbers

E. Multiplying Two Negative Numbers

F. Multiplying a Combination of Positive and Negative Numbers

G. Dividing Two Negative Numbers

H. Dividing One Positive Number and One Negative Number

IV. Order of Operations

V. Fractions, Decimals, Percentages

A. Fraction

B. Decimal

C. Percentage

VI. Solving Simple Equations With One Unknown (X)

A. X Has a Value Subtracted From It (Solve for X Using Addition)

B. X Has a Value Added to It (Solve for X Using Subtraction)

C. X Is Multiplied by a Value (Solve for X Using Division)

D. X Is Divided by a Value (Solve for X Using Multiplication)

VII. Solving More Complex Equations With One Unknown (X)

A. X Is Multiplied by a Value and Then Has a Value Either Added to It or Subtracted From It (Solve for X Using Either Subtraction or Addition and Then Division)

B. X Is Divided by a Value and Then Has a Value Either Added to It or Subtracted From It (Solve for X Using Either Addition or Subtraction and Then Multiplication)

C. X Has a Value Added to It or Subtracted From It and Is Then Multiplied by a Value (Solve for X Using Division and Then Either Subtraction or Addition)

D. X Has a Value Either Added to It or Subtracted From It and Is Then Divided by a Value (Solve for X Using Multiplication and Then Either Subtraction or

Addition)
Glossary
References
Index

PREFACE

A Few (?) Words About Me

When I entered college, I thought my path would take me to law school; however, taking an Intro to Psych class made me realize I enjoy the challenge of trying to understand the human mind. The college I attended, U.C. Santa Cruz, was a fairly unconventional university, but somehow in the midst of a sea of humanistic psychologists I became attracted to the empirical and methodological aspects of psychology by taking classes from instructors such as David Harrington and Dane Archer, who showed me that statistics could appeal to students if taught with a gentle, guiding approach that addresses questions relevant to students' lives. After graduating from college, I was able to get a job as a research assistant for a human resource consulting firm. Despite my lack of work experience, I was hired primarily as a result of having taken statistics and research methods courses, which taught me that learning statistics has benefits both inside and outside of the classroom.

Several years later, I started grad school at U.C. Berkeley, where two events critical to this book took place. First, serving as a teaching assistant, I found I really enjoyed helping students, particularly in statistics and research methods classes often approached with fear and suspicion. Second, I took graduate classes from Geoff Keppel, who had developed his own unique system for analyzing experimental research designs. His lectures and books showed me that statistics can be taught in a systematic way that highlights similarities rather than differences between different research situations. Geoff managed to transform something as daunting-sounding as a "3 x 2 x 4 research design" into the mathematical equivalent of playing with wooden toy alphabet blocks labeled "A," "B," and "C."

After leaving Cal, I was hired by San Jose State, where I teach an introductory statistics course aimed at students with a wide range of backgrounds, ability, and motivation. When I went to select a textbook for the course, for the first time I looked carefully at the wide range of offerings. What I found striking (and still find striking) was that the majority of books focused on providing formulas and very small sets of data designed to demonstrate how to correctly calculate the correct numbers from these formulas. Little emphasis, however, was given to what these numbers meant or implied. Given my own experiences learning and using statistics, I thought a book was needed that focusses on how statistics may be used to answer questions and test ideas within the larger research process.

Goal of this Book

This textbook, aimed primarily at students majoring in the behavioral and social sciences, includes many of the same topics and chapters as the majority of existing introductory statistics books. However, what I believe distinguishes this book is its thematically-based approach that places statistics within the research process. The primary purpose of this

book is to not just teach students how to calculate statistics but how to interpret the results of statistical analyses in light of a study's research hypothesis, and to communicate one's results and interpretations to a broader audience. Hopefully, this book will not only help students understand the purpose and use of statistics, but it will also give them a greater appreciation of how research studies are conceived, conducted, and communicated.

The book is intended for students with a limited knowledge of mathematics, statistics, and research design and methodology. Some common obstacles faced by students are their math anxiety, fear of conducting arithmetic calculations, and an inability to see the relevance of the material to their daily lives. To address these concerns, this books attempts to judiciously use mathematical terminology, write in a conversational manner, and utilize actual research situations rather than contrived sets of numbers.

Organization of the Book

The 15 chapters of this book may be placed into three general categories. Chapters 1 to 4 are designed to introduce students to the research process and how data that have been collected may be organized, presented, and summarized:

Chapter 1 (Introduction to Statistics) defines the concept of statistics, explains why statistics are important to learn, and describes the steps in the research process that are the framework for the rest of the book.

Chapter 2 (Examining Data: Tables and Figures) discusses the importance of examining data that have been collected and describes how tables and figures may be created to represent data.

Chapter 3 (Measures of Central Tendency) presents measures of central tendency (the mode, median, and the mean) and compares them in terms of their relative strengths and weaknesses.

Chapter 4 (Measures of Variability) presents measures of variability (range, interquartile range, variance, standard deviation, index of qualitative variation) and the concept of parameters (population variance and standard deviation).

Chapters 5 through 10 discuss the process of conducting statistical analyses to test research questions and hypotheses, as well as issues and controversies regarding this process:

Chapter 5 (Normal Distributions) introduces theoretical distributions, starting with normal distributions, before moving on to the standard normal distribution (normal curve).

Chapter 6 (Probability and Introduction to Hypothesis Testing) presents the structure for later chapters by introducing, defining, and illustrating steps in hypothesis testing.

Chapter 7 (Testing One Sample Mean) tests one sample mean using the *z* -test and *t* -test and introduces sampling distributions, the central limit theorem, and the Student *t* distribution.

Chapter 8 (Estimating the Mean of a Population) introduces interval estimation by developing the confidence interval for the mean and discusses interval estimation as an alternative to null hypothesis significance testing (NHST).

Chapter 9 (Testing the Difference Between Two Means) covers the *t* -test for independent means for both equal and unequal sample size and introduces within-subjects research designs using the *t* -test for dependent means.

Chapter 10 (Errors in Hypothesis Testing, Statistical Power, and Effect Size) discusses Type I and Type II error and introduces measures of effect size as a way of drawing appropriate conclusions about an analysis.

Chapters 11 to 15 discuss different statistical procedures used in research situations that vary in the number of independent variables as well as how the variables have been measured:

Chapter 11 (One-way Analysis of Variance (ANOVA)) discusses the one-way ANOVA, the relationship between the *t* -statistic and the F-ratio, and planned and post-hoc comparisons.

Chapter 12 (Two-way Analysis of Variance (ANOVA)) introduces factorial research designs, the calculation of F-ratios for main effects and interaction effects, and illustrates the use of simple effects analyses to test research hypotheses.

Chapter 13 (Correlation) defines and illustrates the concept of correlation, discusses the Pearson and Spearman correlation coefficients, and makes the distinction between correlational statistics and correlational research methods.

Chapter 14 (Linear Regression and Multiple Correlation) introduces the concept of regression, discusses linear regression, and works through an example of multiple regression with two predictor variables.

Chapter 15 (Chi-Square) describes the analysis of categorical variables using the chi-square goodness-of-fit test and the chi-square test of independence before ending the book with a brief introduction to nonparametric statistics.

In most cases, the chapters in this book share a uniform format and structure, centered on the research process, based on examples of actual published research studies addressing topics relevant to students' lives. For each example, the research hypothesis is clearly identified; after a brief description of how the data in the study were collected, the calculation and interpretation of both descriptive and inferential statistics are presented,

ending with the extent to which the results of the statistical analyses support the research hypothesis and how these findings may be communicated to a broader audience.

Pedagogical Features

The chapters in this textbook contain several reoccurring features designed to enhance students' comprehension of the material. These include the following:

Periodic learning checks within chapters. Short learning checks, in the form of conceptual review questions and computational exercises, are provided throughout each chapter.

Explicit presentation of formulas and calculations. This book provides both definitional formulas to highlight their relation to the concept being measured and computational formulas to facilitate calculations. In applying these formulas, all of the steps in the calculation of statistics such as the *t* -test and F-ratio are included to help students in completing exercise and homework assignments.

Illustration of the presentation and communication of analyses. Every chapter includes a discussion and illustration of how statistical analyses may be reported as well how they may be communicated to a larger audience in both written and visual form.

Highlighted definition of terms. Every term and concept introduced in this book is boldfaced, immediately followed by an explicit definition.

End of chapter summaries and exercises. The end of each chapter contains a summary of critical points, a list of terms and formulas introduced in the chapter, and a large number of exercises, many of which refer to published research studies.

Calculation of statistics using SPSS. Many chapters include end-of-chapter content that provides an annotated presentation of how statistics may be calculated using the SPSS software. Screen shots of data files, menu selections, and output files are included and annotated to locate and identify relevant aspects of the analysis.

What's New to the 3rd Edition

I'd like to start by expressing my most heartfelt appreciation to the instructors and reviewers of the first two editions of this book as their evaluation, insights, criticisms, and suggestions were absolutely vital to the changes made in this third edition. I'd like to particularly thank the instructors who've adopted the book who were kind enough to spontaneously send me their thoughts and suggestions. It is my sincere hope that all of you see the incorporation of your input in this edition.

There were two main goals in this edition. The first goal was to make the writing as brief and concise as possible, manifested in the goal of reducing the length of each chapter by 5% to

10%. This goal was achieved by simplifying the language and shortening or eliminating what I, the book's adopters, and the book's reviewers felt was unnecessary detail or discussion. The second goal was to increase and update the exercises at the end of many of the chapters, including the use of recent published research as the basis of these exercises. Another goal was to update the demonstration of the use of SPSS to analyze data using the most recent version (Version 31 at the time of this book's publication).

More specifically, in terms of content, the following changes were made to these chapters:

Chapter 1 (Introduction to Statistics): The published research studies used to illustrate concepts in the text, learning checks, and exercises were added to and updated using topics relevant to students (e.g., fear of missing out (FOMO), cryptocurrency, legalization of cannabis, Gen Z, COVID, fake news).

Chapter 6 (Probability and Introduction to Hypothesis Testing): The Super Bowl example used to introduce and illustrate the process of hypothesis testing was updated.

Chapter 8 (Estimating the Mean of a Population): The main research example, which involves data from an actual national salary survey, has been updated throughout this chapter with more current salary data.

As with the first two editions, any and all questions, concerns, suggestions, and complaints you have about the book are most welcome and can be sent to me at howard.tokunaga@sjsu.edu .

A Few Tips for Students

To you, the college student about to read this book as part of taking a statistics course: "Welcome!" and "Great job!" We welcome you because you're about to embark on a semester-long journey that we hope enhances your skills and widens your perspective; we congratulate you because it's a journey not everyone is willing to take. We say to our students that by taking just one statistics course they are ahead of 95% of the population in their knowledge of statistics.

We also tell our students that our primary goal for them is not to become a statistics genius but just to become comfortable with the process of analyzing and interpreting data. In that regard, we agree with a 2012 *Forbes* magazine article entitled *The Six Classes That Will Make Any College Grad Employable* that noted "a good statistics class will help the student become comfortable pushing numbers around and avoiding false interpretations."

At the present moment, we know our encouragement may be of little comfort because you might be anxious about having to learn statistics. Some of you might be anxious about having to learn statistical concepts and formulas; others might be anxious because the research process seems pretty complicated. Talking with students who have taken our courses over the years has helped us assemble the following advice:

Master the Material Presented in the Early Chapters

Chapters 1 through 6 discuss how research is typically conducted, how data are summarized and described, and the process by which researchers conduct statistical analyses to test their ideas. It is absolutely critical for you to have a firm understanding of these chapters as they lay the groundwork for later chapters that discuss a variety of statistical procedures used by researchers to test hypotheses.

What does it mean to master this material? First, *read the chapters both before and after they're discussed in class* . Reading the chapters beforehand helps you identify anything that's unclear and have your questions addressed by your instructor; re-reading the chapters afterward helps confirm your understanding of the material. Next, *be able to define and explain key conceptspresented in these chapters* . These concepts are highlighted in bold-faced letters, and it's important to learn them when they're first introduced because they'll appear throughout the remainder of the book. Next, *do the learning checks within the chapters and the exercises at the end of the chapters* , because they include exercises and review questions you can ask yourself to assess your understanding of the material. Most importantly, *do not miss class during the early part of the semester* ! It's been our experience that students who miss critical lectures at the beginning of the term often have difficulty keeping up as the semester continues.

Review High School Algebra

If you happen to move beyond the introductory statistics course for which you're reading this book, you'll find that statistics are often calculated using computers and statistical software rather than hand calculations. Consequently, some statistics textbooks no longer include mathematical formulas but instead have students use statistical software or Microsoft Excel. However, this book takes a different approach for two main reasons. First, we believe that in order to understand the results of statistical analyses, one must understand their underlying foundation. Learning statistics via computer software can lead to a "brain-dead" approach in which students are at the mercy (rather than control) of their computers. We refer to this as the "I'm only as smart as my printout" method of learning statistics, also known as "Because this is what my computer told me."

Second, this book contains mathematical formulas and calculations for one simple reason: The statistics in this book are not hard to calculate. To assess whether you're prepared to read this book, take the following test:

1. Do you know how to add, subtract, multiply, and divide?
2. Do you know how to calculate the square or square root of a number?
3. Do you understand the order of operations and parentheses within mathematical formulas?

If your answers to these questions are "Yes," congratulations! You possess the ability to conduct *every* mathematical calculation in this book because *none* of the formulas involve geometry, trigonometry, or calculus. However, if you're unsure of your mathematical ability, we suggest you review your high school algebra. Also, there's an appendix at the end of this

book that includes the mathematical concepts and operations needed to conduct the statistical analyses in this book. We highly recommend you read this appendix to review and assess your mathematical skills.

Learn to Use a Statistical Calculator

Although learning the statistics in this book doesn't require the use of computers or software, you will need a hand calculator. Calculators with statistical capabilities are typically identified as "statistical" or "scientific" calculators. Although there's a wide selection of brands and models, know that the calculators most appropriate for this book cost only $10 to $15. We do *not* recommend you use or purchase a more expensive calculator! We prefer these simpler calculators because students are able to do their calculations more quickly and with fewer errors. We do suggest you purchase a calculator similar to the one your instructor uses as they will be more familiar with its features and idiosyncrasies.

In selecting a calculator, look closely at the keyboard; you'll need a calculator whose keys have labels such as $\bar{X}$, σ, and ΣX (Chapters 3 and 4 discuss what these symbols represent). When you go calculator shopping, you'll find calculators that have keys with labels such as ΣY and XY. Although calculators with these Y symbols are only slightly more expensive than calculators with just the X symbols, we do *not* recommend their purchase because these features are not needed to perform the calculations in this book. In fact, we've found students with Y calculators have a more difficult time performing simpler calculations.

Ask Questions

Conducting research is the process of asking and answering questions. Accordingly, throughout this book issues are often framed in the form of questions, which we hope encourages you to ask questions as well. The first person you should direct any questions to is yourself. As you work your way through the chapters, get into the habit of asking yourself, "What does this mean?" A good test of your level of comprehension is whether you can explain the material to yourself in a meaningful way.

You should also be sure to work with your instructor to confirm your level of understanding and clarify unanswered questions. Instructors often say to students, "The only bad question is the one you don't ask." Don't be hesitant about asking questions in class. Instructors will tell you they use one student's question to assess an entire class's level of understanding. This is what we call the "pencil test"—when one student asks a question, we look to see how many other pencils get raised, which reflects the number of students who had the same question. Asking questions enhances the learning experience for yourself, your classmates, and your instructor.

Form a Study Group

No matter how much we encourage students to ask questions in class, it seems they hesitate for fear of drawing attention to themselves. As a result, we turn to another well-

traveled saying: “Misery loves company.” Forming a study group with other students will help you keep up with reading assignments, confirm your comprehension of the material, clarify any confusion you have, prepare for exams, and provide social support. We do suggest that your group contain at least one person who will raise the group’s questions and concerns to the instructor during class or office hours.

Practice, Practice, Practice

The best way to learn the topics covered by this book is to obtain as much practice as possible. Throughout the chapters are examples in which the calculations have been worked out for you; in the middle and end of each chapter are a number of exercises. The answers to some of the exercises are provided at the back of this book, and your instructor has access to the answers to the other exercises. Do the end-of-chapter exercises even if your instructor does not assign them as homework in order to identify any reoccurring mistakes you make. We find that the vast majority of errors students make are simple computational errors, committed when students do the calculations too quickly.

Also note that the publisher of this book, Sage Publications, has developed a webpage for this book that contains a huge amount of supplemental tools such as videos and quizzes for each chapter. Find it by going to the Sage Publishing website **collegepublishing.sagepub.com** , searching for my name (Tokunaga), and looking for the **Resources** tab.

Enough cheerleading and preaching – let’s get started!

WELCOME!

GREAT JOB!

Additional Teaching Resources

Visit **collegepublishing.sagepub.com** , and navigate to the Resources tab on your book’s page to find the teaching materials designed to accompany this textbook. On this site you will find an array of materials that will save you time and to help you keep students engaged, including the following:

Learning management system cartridges that easily integrate with your course management system so student test results and graded assignments seamlessly flow into your gradebook

Test banks , aligned to Bloom’s Taxonomy, that provide a diverse range of test items, including multiple choice, true/false, and essay questions

Lecture notes that provide an outline and the key concepts in each chapter to aid in lecture preparation

PowerPoint ® slides that offer a flexible, accessible, and customizable solution for creating multimedia lectures

Tables and Figures from the book to support lecture preparation and class discussions

Sample course syllabi , including suggested models for structuring your course

Answers to even-numbered questions from the text help facilitate grading

SPSS datasets for use with exercises from the text.

Multimedia content , including original SAGE videos featuring tutorials with author Howard T. Tokunaga that bring concepts to life and appeal to diverse learners

ACKNOWLEDGMENTS

This book has traveled a long journey, with many people playing both transient and persevering roles in its development. First and foremost, I'd like to thank my kids, Meagan and Will, for providing me perspective and inspiration amid all of the twists and turns we faced while this book was being written. I'm so very blessed to have you as my "burden." Away from home, I'm grateful to my graduate students at San Jose State—working on this book has given me much empathy for their thesis-related travails. However, they should be aware that the completion of this book ends the unspoken agreement that I not ask, "So how's your thesis?" if they don't ask, "So how's your book?" A number of friends have provided most welcome relief after long days staring at a computer screen, with particular thanks going to Ellie, Eileen, Susan, Bonnie, Ann, Lisa, and Terry. I would like to thank Charles Linsmeier and David Shirley for their many helpful comments and advice on earlier versions of this book, Julie West for her help in developing chapter exercises, and Youn-Jeong Choi and Naomi Ran for their careful examination of the chapters and exercises.

I would like to express my appreciation to everyone at SAGE who has played a role in shepherding this book to completion. First and foremost, I must express my gratitude to my editor, Vicki Knight. It has been a joy and an honor to work with someone with such a truly exceptional combination of experience, knowledge, perspective, and humor. In production I was fortunate enough to work with others on the SAGE team who made the process a pleasure: Yvonne McDuffee, editorial assistant; Katie Guarino, assistant editor; Nicole Elliott, executive marketing manager; Candice Harman, cover designer; and the copyeditor, Michelle Ponce. Furthermore, I am grateful to the reviewers who took the time to read the chapters, raise questions, and provide constructive comments and suggestions to assist me in clarifying and strengthening the manuscript. The following are the reviewers for third edition:

Erin M. Fekete, University of Indianapolis

Brooke Magnus, Boston College

Loretta Neal McGregor; Arkansas State University

Michael T. Moore, Adelphi University

Megan M. Ringel, St. Edward's University

Bettina Viereck, University of Hartford

Janice Weaver, Community College of Allegheny County

ABOUT THE AUTHOR

Howard T. Tokunaga is professor of Psychology at San Jose State University, where he serves as coordinator of the MS Program in Industrial/Organizational (I/O) Psychology and teaches undergraduate and graduate courses in statistics, research methods, and I/O psychology. He received his bachelor's degree in psychology at U.C. Santa Cruz and his PhD in psychology at U.C. Berkeley. In addition to his teaching, he has consulted with a number of public sector and private sector organizations on a wide variety of management and human resource issues. He is author of *Moving from IBM SPSS to R and RStudio: A Statistics Companion* , and coauthor (with G. Keppel) of *Introduction to Design and Analysis* : *A Student's Handbook* . In his spare time, he enjoys the outdoors, museums, live theater, and being as physically, socially, and politically active as possible.

CHAPTER 1 INTRODUCTION TO STATISTICS

Learning Objectives

1.1 Define the concept of statistics.

1.2 Describe different reasons for learning statistics.

1.3 Describe the main steps in the research process and the tasks and concepts within each step.

In introducing this book to you, we assume you're a college student taking what is perhaps your first course in statistics to fulfill a requirement for your major or a general education requirement. If so, you may be asking yourself two questions:

What is statistics?

Why learn statistics?

The ultimate goal of this book is to help you begin to answer these two questions.

1.1 What Is Statistics?

You may not know it, but you encounter a variety of "statistics" in your day-to-day activities: the typical cost of going to college, the yearly income of the average college graduate, the average price of a car, and so on. So what exactly is "statistics"?

The Merriam-Webster dictionary defines **statistics** as *a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data* . When most people think about statistics, they focus on the "analysis" aspect of this definition—that is to say, they focus on numbers that result from analyzing data. However, statistics not only is concerned about how data are *analyzed* but also how data are *collected* and how the results of statistical analyses are *interpreted* and *communicated* . The main purpose of this book is to introduce, describe, and illustrate the role of statistics within the larger research process.

1.2 Why Learn Statistics?

There are many reasons why you should learn statistics. First, not only do you currently encounter statistics in your daily activities, but *throughout your life you are affected by the results of research and statistical analyses* . Which college you attend is based in part on test scores developed by researchers. You may have to take a personality or intelligence test to get a job. The choices of drugs and medicines available to you are based on medical research. Learning about statistics will help you become a more informed and aware consumer of research and statistical analyses that affect many aspects of your life.

A second reason for learning statistics is that *you may be required to interpret the results of statistical analyses* . Many college courses require students to read academic research journal articles. Evaluating published research is complicated by the fact that different people studying the same topic may come up with diverse or even opposing conclusions. Understanding statistics and their role in the research process will help you decide whether conclusions drawn in research articles are appropriate and justified.

Another reason for learning statistics is that *you might be required to do your own research* . College courses sometimes have students design and conduct mini-research studies; undergraduate majors might encourage students to do senior honors theses; graduate programs often require masters' theses and doctoral dissertations. Learning to collect and analyze data will help you address your own questions in an objective, systematic manner.

A final reason for learning statistics is that it may *help you in your future career* . The website Careercast.com conducts an annual survey in which they evaluate jobs on four dimensions: environmental factors, income, employment outlook, and stress. Recently, their #1 rated job was "data scientist," and the #2 rated job was "statistician." They noted, "A statistician's skill set can be used to break down and analyze large quantities of data. The demand for these skills spans a variety of industries, including marketing, banking, government, sports, retail, and even healthcare."

It's generally a good idea for students to maintain a healthy level of curiosity or even skepticism in regard to their education. However, we find that when it comes to learning statistics, students' feelings may be characterized as fear or anxiety. We understand these feelings, and we hope the benefits of learning statistics will become clear to you and help overcome any concerns you have.

1.3 Introduction to the Stages of the Research Process

Much of scientific research involves asking questions. Throughout this book, we'll examine a broad range of questions regarding human attitudes and behavior that contemporary researchers have asked and attempted to answer. The following are research questions we'll address in this chapter to introduce the stages of the research process:

> Does appealing to one's fear of missing out (FOMO) affect the intention to invest in cryptocurrencies?

Do pro-cannabis messages on social media lead to cannabis use?

Do parents differ in their awareness of their children's online activities?

Do nonbinary college students have the same experiences as cis men or cis women?

How might you try to answer questions such as these? You could base your answers on your personal beliefs, or you could adopt the answers given to you by others. But rather than relying on subjective beliefs and feelings, researchers in many academic disciplines test their ideas using the scientific method. The **scientific method** is a method of investigation that uses the objective and systematic collection and analysis of empirical data to test theories and hypotheses.

At its simplest, this book will portray the scientific method as consisting of five main steps or phases:

developing a research hypothesis to be tested,

collecting data,

analyzing the data,

drawing a conclusion regarding the research hypothesis, and

communicating the findings of the study.

Accomplishing each of these five steps requires completing a number of tasks, as shown in Figure 1.1 . Because this sequence of steps serves as the model for the research process we'll discuss throughout this book, each step is briefly introduced in the following sections. It's important to understand that the research process depicted in Figure 1.1 represents an ideal way of doing research. The "real" way, as you may discover in your own efforts, is often anything but a smooth ride but rather is filled with false starts, dead ends, and wrong turns.

Developing a Research Hypothesis to Be Tested

The initial stage—and the first step—of the research process is to develop a research hypothesis to be tested. A **research hypothesis** may be defined as a statement regarding an expected or predicted relationship between variables. A **variable** is a property or characteristic of an object, event, or person that can take on different values. One example of a variable is "U.S. state," a variable with 50 possible values (Alabama, Arkansas, etc.).

Research hypotheses are usually developed through the completion of several tasks:

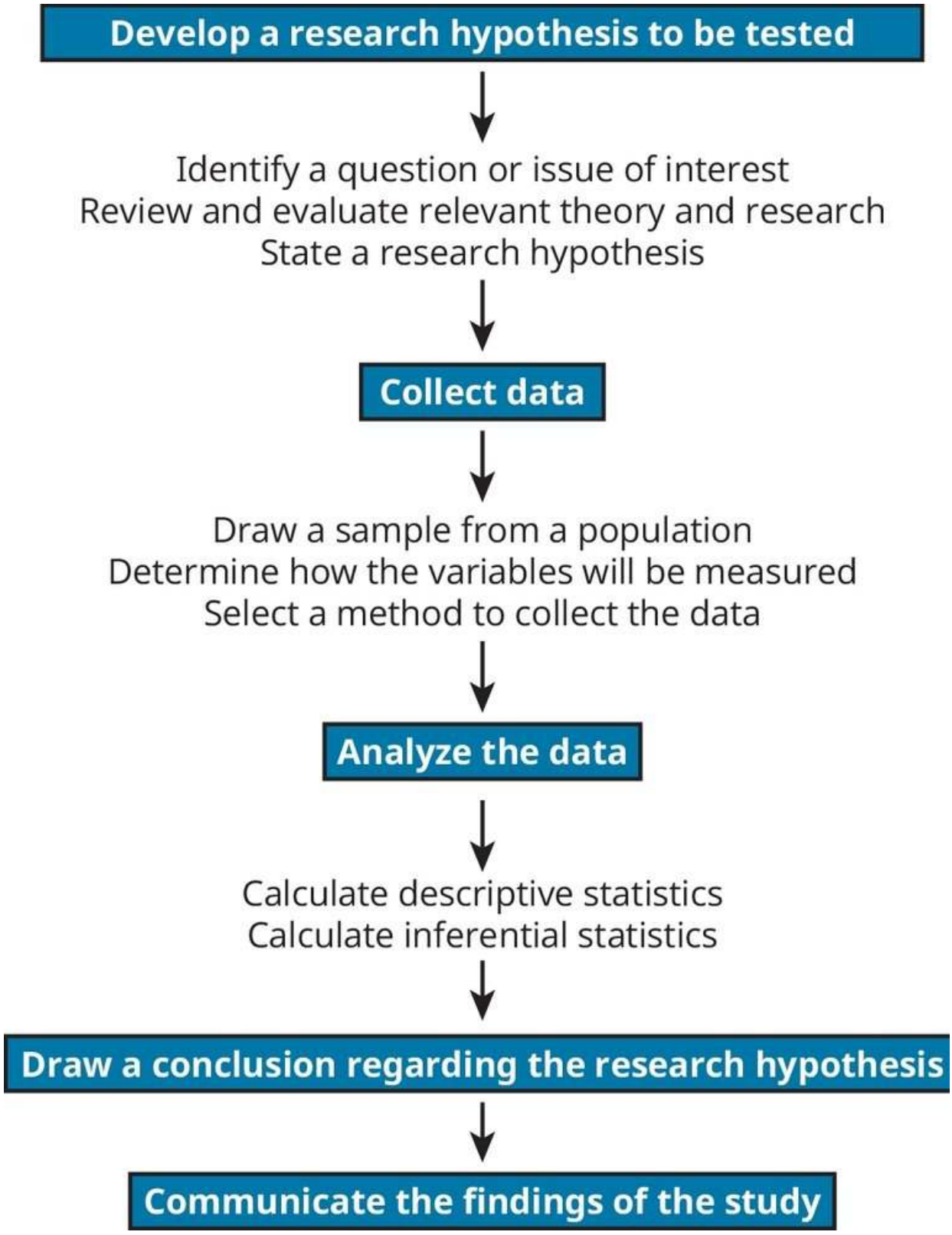
identifying a question or issue to be examined,

reviewing and evaluating relevant theories and research, and

stating a research hypothesis.

Identifying a Question or Issue to Be Examined

Most research starts with a question posed by the researcher that often comes from the researcher's own ideas and daily observations. Although this may not seem terribly scientific, there's an advantage in using one's own experience as a starting point: People are generally much more motivated to explore issues that concern them personally. In teaching statistics, we advise students developing their own research projects to study something of interest to them. Conducting research can be tedious, difficult, and frustrating. At various points, you may ask yourself, "Why am I doing this?" Being able to provide a satisfactory answer to this question will help you overcome obstacles you encounter along the way.


Develop a research hypothesis to be tested
Identify a question or issue of interest
Review and evaluate relevant theory and research
State a research hypothesis
Collect data
Draw a sample from a population
Determine how the variables will be measured
Select a method to collect the data
Analyze the data
Calculate descriptive statistics
Calculate inferential statistics
Draw a conclusion regarding the research hypothesis
Communicate the findings of the study

▶ Long Description: Figure 1.1 Steps in the Research Process Within the Scientific Method

Figure 1.1 Steps in the Research Process Within the Scientific Method

Reviewing and Evaluating Relevant Theories and Research

Beyond your own curiosity, research questions often arise from studying the theories, ideas, and research of others. A **theory** is a set of propositions used to describe or explain a phenomenon. The purpose of a theory is to summarize and explain specific facts using a logically consistent framework. Placing a question within a theoretical framework provides guidance and structure to research.

Reviewing and evaluating existing theories and research helps you decide whether it's worth the time and energy required to conduct a study. By seeing what others have done, you may conclude that the current way of thinking on a topic is incomplete or mistaken. By doing this review, you're able to ensure that the research you conduct adds to and improves upon an existing body of knowledge.

Stating a Research Hypothesis: Independent and Dependent Variables

Understanding and evaluating an existing literature not only helps develop a question of interest, but it may also lead to a predicted answer to that question. Within the scientific method, this answer is stated as a research hypothesis, defined earlier as a statement regarding an expected or predicted relationship between variables. Table 1.1 lists the research questions and research hypotheses for the studies listed at the beginning of this section. For example, the first research hypothesis states that "consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos."

One characteristic of research hypotheses such as those listed in Table 1.1 is that they identify the variables that are the focus of their research. As mentioned earlier, a variable is a property or characteristic with different values. Variables can be classified in several ways. In specifying a research hypothesis, researchers often speak in terms of "independent" and "dependent" variables. An **independent variable** may be defined as a variable manipulated or controlled by the researcher. A **dependent variable** , on the other hand, is a variable measured by the researcher that is expected to change or vary as a function of the independent variable. The independent variable is seen as the "cause" or "treatment," whereas the dependent variable is the "effect" or "outcome."

Researchers are interested in examining the effect of the independent variable on the dependent variable. Consider the first research hypothesis provided in Table 1.1 : "Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos." Here, the independent variable is the type of appeal consumers are exposed to: FOMO or non-FOMO. The dependent variable is the intention to invest in cryptos. In this study, the effect of the independent variable on the dependent variable is that

differences in the intention to invest in cryptos (the dependent variable) may “depend” upon which type of appeal (FOMO or non-FOMO) a consumer receives. Table 1.2 lists the independent and dependent variables for each of the research hypotheses in Table 1.1 .

In addition to identifying variables, a second characteristic of research hypotheses is that they specify the nature and direction of the relationship between variables. For example, the first research hypothesis in Table 1.1 states that “Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *higher* intention to invest in cryptos.” The word *higher* indicates the nature and direction of the relationship between the independent variable (type of appeal) and the dependent variable (intention to invest in cryptos). The direction of the relationship would not have been stated if the hypothesis had included the less specific phrase “will have a *different* intention,” which simply indicates that the two groups of consumers are not expected to be the same. Table 1.3 provides directional and nondirectional research hypotheses for the research studies from Table 1.1 .

Table 1.1 Examples of Research Questions and Hypotheses

Research Question	Research Hypothesis
Does appealing to one’s fear of missing out (FOMO) affect the intention to invest in cryptocurrencies? (Friederich et at., 2024)	Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.
Do pro-cannabis messages on social media lead to cannabis use? (Willoughby et al., 2024)	Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.
Do parents differ in their awareness of their children’s online activities? (Symons et al., 2017)	Mothers have more accurate knowledge of their children’s online activities than do fathers.
Do nonbinary college students have the same experiences as cis men or cis women? (Marx et al., 2024)	Nonbinary college students experience higher levels of victimization than do cis men or cis women.

The ability to form a directional research hypothesis is dependent on the state of the literature on the question of interest. If little or perhaps conflicting research has been conducted, researchers may not be able to form a directional hypothesis. For example, one study examining the relationship between exercise deprivation (not allowing people to exercise) and depression. reported, “We did not have a directional hypothesis, and when participants asked what we expected to find in this study, we replied: ‘We really are not sure since the results of earlier work on exercise deprivation are mixed’” (Mondin et al., 1996, p. 1200).

Table 1.2 Research Hypotheses and Their Independent and Dependent Variables

Research Hypothesis	Independent Variable	Dependent Variable
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher	Type of appeal	Intention to invest in cryptos

Research Hypothesis	Independent Variable	Dependent Variable
intention to invest in cryptos.		
Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.	Exposure to pro-cannabis messages	Intention to use cannabis
Mothers have more accurate knowledge of their children's online activities than do fathers.	Parent	Knowledge of children's online activities
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	Gender identity	Amount of victimization

Collecting Data

Once a research hypothesis has been formulated, researchers are ready to proceed to the second stage in the research process: collecting data. This step is seen as being composed of three tasks:

drawing a sample from a population,

determining how the variables will be measured, and

selecting a method by which to collect the data.

Table 1.3 Directional and Nondirectional Research Hypotheses

Directional Research Hypothesis	Nondirectional Research Hypothesis
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *higher* intention to invest in cryptos.	Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *different* intention to invest in cryptos.
Exposure to pro-cannabis messages on social media is related to *increased* intentions to use cannabis.	Exposure to pro-cannabis messages on social media is *related to* intentions to use cannabis.
Mothers have *more* accurate knowledge of their children's online activities than do fathers.	Mothers have *different* knowledge of their children's online activities than do fathers.
Nonbinary college students experience *higher* levels of victimization than do cis men or cis women.	Nonbinary college students experience *different* levels of victimization than do cis men or cis women.

Drawing a Sample From a Population

The first step in collecting data is to identify the group of participants to which the research hypothesis applies. The group to which the results of a study may be applied or generalized is called a population. A **population** is the total number of possible units or elements that could potentially be included in a study. Although people often think the word "population" refers to everyone in the world, researchers typically define their populations more narrowly. For example, the population for the first research hypothesis in Table 1.1 could be "consumers," "U.S. consumers," "Georgia consumers," or "University of Georgia consumers." Researchers typically try to define their populations as broadly as possibly (e.g., "consumers" rather than "University of Georgia consumers") to maximize the applications or implications of their research.

Learning Check 1

Reviewing What You've Learned So Far



1. Review questions
    a. What are the five steps involved in the research process?
    b. Why is it useful to review and evaluate theories and research before conducting a study?
    c. What are the two main characteristics of research hypotheses?
    d. What is the difference between an independent variable and a dependent variable?
2. The following are several research hypotheses from published studies. For each research hypothesis, identify the independent and dependent variable.
    a. "College students will rate instructors who dress formally (i.e., business suit and tie) as having more expertise than instructors who dress casually (i.e., slacks and shirt)" (Sebastian & Bristow, 2008, p. 197).
    b. "It was expected tha. . . . greater amounts of television viewin. . . . would predict greate. . . . posttraumatic stress symptoms" (McLeish & Del Ben, 2008, p. 417).
    c. "Do borrowers who did not complete college have greater financial distress compared with . . . non-borrowers who did not complete college [or] borrowers who earned a degree?" (Jabbari et al., 2023, p. 6).
    d. "We expected that drivers would believe that texting while driving is more detrimental to their driving performance than browsing social media" (Hashash et al., 2019, p. 70).

It's extremely difficult to collect data from all members of a population. Imagine, for example, the time and money needed to collect information from every consumer in the United States! For this reason, researchers typically draw conclusions about populations based on information collected from a sample of the population. A **sample** is a subset or portion of a

population. Table 1.4 describes the samples used in the four studies introduced in Table 1.1 . As you can see from Table 1.4 , samples vary greatly in terms of their targeted population and size (the number of participants). The extent to which the results of a study can be generalized to the target population depends on the extent to which the sample is representative of the population.

Determining How Variables Will Be Measured: Levels of Measurement

The research hypotheses described in Table 1.1 involve variables such as type of appeal, intentions to use cannabis, and gender identity. To conduct a research study, the researcher must decide how to measure the variables stated in the research hypothesis. **Measurement** is the assignment of numbers or categories to objects or events according to rules.

For example, to measure the variable "height," a researcher might use "number of inches from the ground in bare feet" as a form of measurement. To measure "success in college," a student's grade point average (GPA) may be obtained from school transcripts. "Self-esteem" might be measured by having people complete a questionnaire and be categorized as having either "low" or "high" self-esteem. As these examples demonstrate, different types of variables have different forms of measurement. Consequently, researchers have identified four distinct levels for measuring variables: nominal, ordinal, interval, and ratio.

The values of variables measured at the **nominal level of measurement** differ in category or type. The word *nominal* implies having to do with "names," such that we use first and last names to distinguish between people. Gender identity is an example of a nominal variable, in that it consists of categories or types (such as cis male and cis female) rather than numeric values. In the first research hypothesis in Table 1.1 , the independent variable of "type of appeal" is measured at the nominal level with two categories: FOMO and non-FOMO.

Variables measured at the **ordinal level of measurement** have values that can be placed in an order relative to the other values. Rankings (such as finishing first, second, or third in a race), size (small, medium, large, or extra-large), and ratings (below average, average, or above average) are examples of ordinal variables. The values of ordinal variables indicate that one value represents relatively more or less of the variable than other values, but they don't specify the precise difference between values. For example, although you can say that a "medium" drink is smaller than a "large" drink which is smaller than an "extra-large" drink, you can't specify the exact difference in size between the three drinks.

Table 1.4 Research Hypotheses and Study Samples

Research Hypothesis	Study Sample
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.	"We recruited 150 participants from Amazon MTurk" (Friederich et at., 2024, p. 105).
Exposure to pro-cannabis messages on social media is	"Teens (N = 350) . . . needed to be between 13 and 17 years old and from Washington state" (Willoughby et

Research Hypothesis	Study Sample
related to increased intentions to use cannabis.	al., 2024, p. 185).
Mothers have more accurate knowledge of their children's online activities than do fathers.	"triadic data including reports from children aged 13 to 18, mothers and fathers . . . recruited in Flanders, the northern, Dutch speaking region of Belgium . . . a total of 357 valid triads was achieved" (Symons et al., 2017, p. 405).
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	"undergraduate college students aged 18–30 who were enrolled at 4-year institutions. This yielded a sample of 46,510 students" (Marx et al., 2024, p. 334).

The values of variables measured at the **interval level of measurement** are numbers equally spaced along a continuum. One example of an interval variable is the Fahrenheit scale of temperature. Here, a difference of 5 degrees has the same meaning anywhere along the scale; for example, the difference between 45°F and 50°F is the same as the difference between 65°F and 70°F. Many variables studied in the behavioral sciences (e.g., personality characteristics or attitudes) are considered to be measured at the interval level of measurement.

Variables at the **ratio level of measurement** are identical to interval variables, with one exception: Ratio scales possess what is known as a true zero point, for which the value of zero (0) represents the complete absence of the variable. Variables that describe a physical dimension (such as height, weight, distance, and time duration) typically have a true zero point. In the last research hypothesis in Table 1.1 , the variable "number of times victimized in last 12 months" is a ratio variable because it has a true zero point, where zero would indicate the complete absence of victimizations.

One advantage of ratio measurement versus interval is that ratio variables allow for a greater number of comparisons among values. Consider, for example, the value 6 for the ratio variable "inches." Not only is the difference between 4 and 6 inches the same as the difference between 6 and 8 inches (thereby involving addition and subtraction, which also can be used with interval variables), 6 inches is twice as long as 3 inches and half as long as 12 inches (involving multiplication and division). Multiplication and division comparisons cannot be made with interval variables. You cannot, for example, say that a temperature of 90°F is three times as much temperature as 30°F.

The boldfaced and italicized text in Table 1.5 illustrates how the four studies described in the earlier tables measured their independent and dependent variables. As you can see, the variables in these studies have been measured in a variety of ways at different levels of measurement.

Table 1.5 Research Hypotheses and Measurement of Variables

Research Hypothesis	Independent Variable	Dependent Variable

Research Hypothesis	Independent Variable	Dependent Variable
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.	Type of appeal **FOMO, non-FOMO (nominal)**	Intention to invest in cryptos **Likelihood of investing** **1-7 scale (1 = Highly unlikely to 7 = Highly likely) (ordinal)**
Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.	Exposure to pro-cannabis messages on social media **How often saw pro-cannabis messages (ratio)**	Intentions to use cannabis **Likelihood of using cannabis in the future (ratio)**
Mothers have more accurate knowledge of their children's online activities than do fathers.	Type of parent **Mother, Father (nominal)**	Accuracy of online activities **Accurate, somewhat accurate, inaccurate (nominal)**
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	Gender identity **Nonbinary, cis man, cis woman** **(nominal)**	Number of times victimized in last 12 months **(ratio)**

What difference does the level of measurement make?

A variable's level of measurement has important implications for researchers in that it influences how research hypotheses are stated and how data are analyzed. For example, imagine you study the variable "success in college." You could measure this variable at the *nominal* level of measurement by classifying students into one of two groups: successful or unsuccessful. Researchers typically use nominal variables when they are interested in studying differences between groups, such as, "Do successful and unsuccessful students differ in their study habits?"

You could instead measure success in college as an *ordinal* variable by ranking students from top to bottom, which enables a researcher to study relative differences with less concern for the precise size of these differences. For example, using ranked data, you could ask, “How similar are younger and older faculty members’ rankings of their students?”

To study success in college at the *interval* level, you could rate each student on a scale from 1 (low) to 100 (high). You could then use these ratings to address a research question such as, “Is there a relationship between faculty ratings of their students and students’ ratings of themselves?”

Finally, to measure success in college using the *ratio* level of measurement, you could use students’ salary after graduation (in dollars) as an indicator of success. Salary is a ratio variable because it contains a true zero point. In practice, ratio variables can be used to ask many of the same questions as those involving interval variables. For example, you could ask, “Is there a relationship between faculty members’ ratings of students and students’ salaries after graduation?”

As you see, how researchers measure their variables influences how they state their research questions and hypotheses. The level of measurement has another important implication: It helps determine the statistics researchers use to analyze their data. Certain statistics can be applied to variables measured at the interval or ratio levels of measurement, whereas others are appropriate for nominal or ordinal variables. As we move from chapter to chapter of this book, the level of measurement of the variables being analyzed will be noted.

Selecting a Method to Collect the Data: Experimental and Nonexperimental Research Methods

In addition to drawing a sample from a population and determining how the variables will be measured, the third step in collecting data is to determine the type of research method to use to collect the data on these variables. Research methods may be classified into two main types:

experimental research methods, and

nonexperimental research methods.

Experimental research methods.

Experimental research methods are methods designed to test causal relationships between variables—more specifically, whether changes in independent variables produce or cause changes in dependent variables. To make inferences about cause-effect relationships, researchers conducting an experiment must first eliminate other possible causes or explanations for changes in the dependent variable besides the independent variable. If it can be claimed that a variable other than the independent variable created the changes in the dependent variable, the research results are said to be confounded. A **confounding variable** is a variable related to an independent variable that provides an alternative explanation for the relationship between the independent and dependent variables.

To illustrate the concept of confounding variables, consider the following question: Is there a relationship between a mother's ethnicity and the birth weight of her baby? Although research has shown differences in the birth weights of babies of different ethnicities, one study noted that some of this research did not eliminate the possibility that these differences may be partly due to differences between ethnic groups on factors associated with low birth weights, such as the mother's average age at the time of pregnancy and behaviors such as smoking and drinking (Ma, 2008). These other factors are considered confounding variables in that they provide alternative explanations for the relationship between ethnicity and babies' birth weights. Confounding variables make it difficult for researchers to draw firm conclusions regarding the relationship between the independent and dependent variable.

Learning Check 2

Reviewing What You've Learned So Far



1. Review questions
    a. What is the difference between a population and a sample?
    b. Within the research process, what is the relationship between populations and samples?
    c. What are the four levels of measurement? How do they differ?
2. For each of the following variables, name the scale of measurement (nominal, ordinal, interval, or ratio).
    a. Type of school attended (public, private)
    b. Probability of graduating college in 4 years (0% to 100%)
    c. Rating of television (good, better, best)
    d. Number of computers in home

Researchers minimize the influence of confounding variables in two main ways. First, researchers exercise experimental control by making the research setting (i.e., characteristics of the research participants, the instruments or measures administered, the methods used to collect data) the same for all participants. In the birth weight study, for example, the research design might control for the effect of parental smoking on birth weight by only including nonsmokers in the study sample, thereby excluding those who smoke.

Another way researchers exert control over the research setting is by including a condition known as a **control group** , a group of participants in an experiment not exposed to the independent variable. For example, if you conducted an experiment examining the effects of caffeine on one's health, one group (the experimental group) might be instructed to drink coffee while a second group (the control group) drinks decaffeinated coffee. By comparing the results of the two groups, researchers can then assess the impact of caffeinated coffee on the health of those who drink it.

Because researchers cannot possibly identify and control for all potential confounding variables, a second strategy used to minimize the influence of confounding variables is **random assignment** , which is assigning participants to each category of an independent variable in such a way that each participant has an equal chance of being assigned to each category. For example, a researcher could randomly assign a participant to the experimental or control condition by flipping a coin: heads for experimental, tails for control. The purpose of random assignment is to equalize or neutralize the effects of confounding variables by distributing them equally over the levels of the independent variable.

Nonexperimental (correlational) research methods.

Experimental research designs are one of the best tools researchers have for making causal inferences. However, the ability to make causal inferences requires a great deal of control that may not always be possible or desirable. For this reason, researchers often employ **nonexperimental research methods** , which are research methods designed to measure relationships between variables without having the ability to infer cause-effect relationships. These research methods are sometimes referred to as *correlational* research methods. It is important to understand that correlational research methods are different from correlational statistics we'll introduce in Chapter 13. Correlational research methods pertain to how data are *collected* ; correlational statistics pertain to how data are *analyzed* . Some of the most common types of nonexperimental research designs include quasi-experiments, survey research, observational research, and archival research.

Quasi-experimental research compares naturally formed or preexisting groups rather than randomly assigning participants to conditions. One example of this type of research was a study examining the effects of the COVID-19 pandemic on three groups of students: those living on campus, those living off campus but within commuting distance to school, and those living far away in another city, state, or country (Bisconer & McGill, 2023). This study is a quasi-experiment because the students were not randomly assigned to one of the three living conditions before collecting data.

Survey research methods obtain information directly from a group of people regarding their opinions, beliefs, or behavior. The goal of survey research, which can include using questionnaires and interviews, is to obtain information from a sample to represent or estimate the views of a larger population. One example of survey research examined college students' mental health by having them fill out a survey in which they reported their spirituality (the desire for truth, meaning, and significance of one's life) and their self-acceptance (knowledge and acceptance of one's self and limitations) (Nonis et al., 2024). Survey research is nonexperimental in that the variables being measured are not manipulated by the researcher.

Observational research is the systematic and objective observation of naturally occurring behavior or events. Observational research is often used to study phenomena that the researcher cannot or should not deliberately manipulate. For example, researchers studying interpersonal conflict among children would never force children to fight each other. One study resolved this issue by videotaping groups of adolescents discussing topics such as competition and peer pressure and then counting the number of times the adolescents teased or mocked each other (Connolly et al., 2015).

Rather than observing or measuring behavior directly, **archival research** is the use of archives (records or documents of the activities of individuals, groups, or organizations) to examine research questions or hypotheses. One example of an archival research study studied the question of whether men are perceived to be more creative than women (Proudfoot et al., 2015). In this study, the 100 most viewed talks on TED.com were identified, and the researchers compared viewers' posted ratings of the talks given by men versus women.

Labeling variables in nonexperimental research.

One issue that often confuses students when learning about nonexperimental research is the practice of referring to the variables as the "independent variable" and the "dependent variable." For example, in the study described previously that involved biological sex and TED.com ratings, biological sex may be called the "independent variable." However, because biological sex isn't manipulated by the researcher, it would be inappropriate to say that biological sex "causes" differences in TED.com ratings. To avoid confusion, researchers sometimes refer to the independent variable and the dependent variable in a nonexperimental study as the **predictor variable** and **criterion variable**, respectively.

Analyzing the Data

Once data collection has been completed, the next step in the research process involves data analysis. This part of the research process addresses the primary topic of this book: statistics. Because this is the first chapter of this book, we won't describe analyzing data in detail. Instead, we introduce the notion that there are two main purposes of analyzing data: (1) to organize, summarize, and describe data that has been collected and (2) to test and draw conclusions about ideas and hypotheses. These two purposes are met by calculating two main types of statistics: descriptive statistics and inferential statistics.

Calculating Descriptive Statistics

Descriptive statistics are statistics used to summarize and describe a set of data for a variable. For example, you've probably heard of statistics used to describe or summarize aspects of our society, such as crime statistics and unemployment statistics. Using a research-related example, Caitlin Abar, a researcher at Brown University, conducted a study examining increases in students' alcohol-related behaviors after entering college (Abar, 2012). To measure students' level of alcohol use, she created a variable called "typical weekend drinking," which "was measured as the sum of drinks consumed on a typical Friday and Saturday within the past 30 days" (p. 22). One way to summarize students' responses to the typical weekend drinking variable (0, 1, 2, etc.) would be to calculate the mean, which is the mathematical average of a set of scores. The mean, described in Chapter 3, is an example of a descriptive statistic. The first part of this book describes a variety of descriptive statistics used by researchers to summarize data they've collected.

Calculating Inferential Statistics

Besides summarizing and describing data, a second purpose of statistics is to test hypotheses. **Inferential statistics** are statistical procedures used to test hypotheses and draw conclusions from data collected during research studies. Using inferential statistics, researchers make inferences about what's believed to exist in populations based on data gathered from samples of the population.

As an example of an inferential statistic, let's return to the example of students' alcohol-related behaviors (Abar, 2012). This study was interested in seeing whether these behaviors were related to aspects of these students' relationships with their parents. One aspect was "alcohol communications," defined as "how often they discussed alcohol related topics with their parents" (p. 22). To test a hypothesis regarding the relationship between students' alcohol use and their alcohol communications, we could use an inferential statistic known as the Pearson correlation coefficient, which will be discussed in Chapter 13. The chapters in the last half of this book discuss a broad variety of inferential statistics, along with detailed instructions for analyzing and drawing conclusions from statistical data and presenting research findings.

Drawing a Conclusion Regarding the Research Hypothesis

Once statistical analyses have been completed, the next step is to interpret the results of the analyses as they relate to the study's research hypothesis. More specifically, do the findings of the analyses support or not support the research hypothesis? The word *support* in the previous sentence is very important. Students and researchers are sometimes tempted to say their findings "prove" their hypotheses are either "true" or "false." However, as we'll discuss in Chapter 5, because researchers typically don't collect data from the entire population, they can't know with 100% certainty whether their hypothesis has been proven to be true or false. A *huge* part of this book is learning how to draw appropriate conclusions from statistical analyses.

Communicating the Findings of the Study

Conducting research requires a variety of different skills: conceptual skills to develop research hypotheses, methodological skills to collect data, and mathematical skills to analyze these data. Another integral part of the research process is the communication skills needed to inform others about a study. Researchers must be not only skilled scientists but also effective writers.

Learning Check 3

Reviewing What You've Learned So Far



1. Review questions
    a. What are the main differences between experimental and nonexperimental research methods?
    b. What are confounding variables? What do researchers do to minimize the effects of confounding variables?
    c. What are the main types of nonexperimental research methods?
    d. What are the relative strengths and weaknesses of experimental and nonexperimental research?
    e. What are the main purposes of descriptive and inferential statistics?

For many years, the American Psychological Association (APA), the professional association for psychologists, has provided researchers guidance on how to communicate the results of their research. In 1929, the APA published a seven-page article entitled, “Instructions in Regard to Preparation of Manuscript.” In contrast, the seventh edition of the *Publication Manual of the American Psychological Association* , published in 2020, consists of 428 pages! This increase in length highlights the challenge of communicating research effectively and appropriately.

1.4 Plan of the Book

The primary purpose of this chapter was to introduce statistics and place it within the larger research process. The remainder of this book discusses different statistics used by researchers to examine various questions of interest. Although it's critical for you to learn to correctly calculate statistics, it is equally important to understand the role of statistics within the research process and appreciate issues related to the use (and sometimes misuse) of statistics.

The first half of the book (Chapters 2 through 6) introduces conceptual and mathematical issues that are the foundation of statistical analyses. Chapters 2, 3, and 4 discuss how data may be examined, described, summarized, and presented in numeric, visual, and graphic form. Descriptive statistics are introduced and discussed in these chapters. Chapters 5 and 6 introduce characteristics of the inferential statistics used to test research hypotheses. The chapters in the second half of the book cover a number of different inferential statistics used with different types of variables.

In most cases, the chapters in this book share a uniform format and structure and are centered on the research process, using examples of actual published research studies that address questions with which you may be familiar. For each example, we'll clearly state the research hypothesis, briefly describe how the data in the study were collected, illustrate the calculation and interpretation of descriptive and inferential statistics, discuss the extent to which the statistical analyses support the research hypothesis, and show how to communicate one's research to a broader audience.

1.5 Looking Ahead

We began this chapter by defining statistics and providing reasons why learning about them may be of value to you. Next, we placed statistics within the stages of the larger research process, a process that will be the foundation of this textbook. As we've explained, the research process is centered on a research hypothesis, which is a predicted relationship between variables. Once a hypothesis has been stated, the next step is to collect data to be analyzed. However, before conducting statistical analyses on a set of data, there is preliminary work that must be completed. The next chapter will focus on methods used by researchers to examine data.

1.6 Summary

Statistics may be defined as a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data. As such, statistics not only is concerned about how data are analyzed but also recognizes the importance of understanding how data are collected and how the results of analyses are interpreted and communicated. There are a variety of reasons why students should learn statistics: Statistics are encountered in a wide variety of daily activities, students may be asked or required to read and interpret the results of statistical analyses in their courses, students may use statistics in conducting their own research, and statistics may help one's career.

Researchers conduct research using the **scientific method** of inquiry, a method of investigation that uses the objective and systematic collection and analysis of empirical data to test theories and hypotheses. The research process used within the scientific method of inquiry consists of five main steps: developing a research hypothesis to be tested, collecting data, analyzing the data, drawing a conclusion regarding the research hypothesis, and communicating the findings to the study.

A **research hypothesis** is a statement regarding an expected or predicted relationship between variables. Developing a research hypothesis involves identifying a question or issue to be examined, reviewing and evaluating relevant theories and research, and stating the research hypothesis to be tested in the study. A **theory** is a set of propositions used to describe or explain a phenomenon. A research hypothesis contains **variables** , which are properties or characteristics of some object, event, or person that can take on different values. More specifically, a research hypothesis states the nature and direction of a proposed relationship between an **independent variable** (a variable manipulated or controlled by the researcher) and a **dependent variable** (a variable measured by the researcher that is expected to change or vary as a function of the independent variable).

Collecting data involves drawing a sample from a population, determining how the variables will be measured, and selecting a method to collect the data. A **population** is the total number of possible units or elements that could potentially be included in a study; a **sample** is a subset or portion of a population. Variables can be measured at one of four levels of measurement: **nominal** (values differing in category or type), **ordinal** (values placed in an order relative to the other values), **interval** (values equally spaced along a numeric continuum), or **ratio** (values equally spaced along a numeric continuum with a true zero point).

There are two main types of methods used to collect data: **experimental research methods** , designed to test whether changes in independent variables produce or cause changes in dependent variables, and **nonexperimental (correlational) research methods** , designed to examine the relationship between variables without the ability to infer cause-effect relationships.

In experimental research, researchers are concerned about possible **confounding variables** , which are variables related to independent variables that provide an alternative explanation for the relationship between independent and dependent variables. To minimize the impact of confounding variables, a research study may exercise experimental control over various aspects of the situation, including the use of a **control group** , which is a group of participants not exposed to the independent variable the research is designed to test, or **random assignment** , which involves assigning participants to each category of an independent variable in such a way that each participant has an equal chance of being assigned to each category.

Examples of nonexperimental research methods are **survey research** , in which information is directly obtained from a group of people regarding their opinions, beliefs, or behavior; **observational research** , which involves the systematic and objective observation of naturally occurring events; and **archival research** , which uses archives (records or documents of the activities of individuals, groups, or organizations) to examine research questions or hypotheses. In nonexperimental research, the independent variable and

dependent variable may be referred to as the **predictor variable** and **criterion variable**, respectively.

Once data have been collected, the next step is to analyze them using two main types of statistics: **descriptive statistics**, which summarize and describe a set of data, and **inferential statistics**, which are statistical techniques used to test hypotheses and draw conclusions from data collected during research studies.

Once the statistical analyses have been completed, the results of the analyses are interpreted regarding whether they support or not support the study's research hypothesis. The final step in the research process is to communicate the study to a broader audience.

1.7 Important Terms

archival research

confounding variable

control group

criterion variable

dependent variable

descriptive statistics

experimental research methods

independent variable

inferential statistics

interval level of measurement

level of measurement (nominal, ordinal, interval, ratio)

measurement

nonexperimental research methods

observational research

population

predictor variable

quasi-experimental research

random assignment

research hypothesis

sample

scientific method

statistics

survey research

theory

variable

1.8 Exercises

1. Identify the independent and dependent variable for each of the following research questions and hypotheses from published articles.
    a. “We hypothesize that finance majors will embody narcissistic traits more than other business majors” (Wood et al., 2021, p. 924).
    b. “Men are expected to employ physical characteristics (particularly those that are directly related to sex) more often than women in selecting dating candidates” (Hetsroni, 2000, p. 91).
    c. “It is hypothesized that perceptions of emotional pain and support needs will be higher for the community of the Black victim than those of the officers’” (Kahn et al., 2024).
    d. “Gen Z students are more likely to believe that ‘Group projects bring out the worst in people’ than Millennials” (Schlee et al., 2020, p. 145).
    e. “We hypothesized to find . . . higher stress and anxiety post COVID in the Portuguese group in contrast with the British ” (Penalba-Sanchez et al., 2022, p. 322).
    f. “We hypothesized that consumers would think that a ‘sale price’ presentation would generate a greater monetary savings than an ‘everyday low price’ presentation” (Tom & Ruiz, 1997, p. 403).
2. The following are a number of research questions and hypotheses from actual published articles. For each hypothesis, identify the independent and dependent variable.
    a. “The use of color in a Yellow Pages advertisement will increase the perception of quality of the products for a particular business when compared with noncolor advertisements” (Lohse & Rosen, 2001, p. 75).
    b. “We hypothesized that parents who use more frequent corporal and verbal punishmen. . . . will report more problem behaviors in their children” (Brenner & Fox, 1998, p. 252).
    c. “It was hypothesized that adolescents with anorexia nervosa woul. . . . be more respectful when compared with peers with bulimia nervosa” (Pryor & Wiederman, 1998, p. 292).
    d. “The purpose of the present research was to assess brand name recognition as a function of humor in advertisements. . . . It was predicted that participants would recognize product brand names that had been presented with humorous

advertisements more often than brand names presented with nonhumorous advertisements" (Berg & Lippman, 2001, p. 197).

e. "The purpose of this study was to explore the relation between time spent in daycare and the quality of exploratory behaviors in 9-month-old infant. . . . it was hypothesized tha. . . . infants who spent greater amounts of time in center-based care would demonstrate more advanced exploratory behaviors than infants who did not spend as much time in center-based care" (Schuetze et al., 1999, p. 269).

f. "It was hypothesized that students would score higher on test items in which the narrative contains topics and elements that resonate with their daily experiences versus test items comprised of material that is unfamiliar" (Erdodi, 2012, p. 172).

3. Identify the independent and dependent variable for each of the following research questions and hypotheses from published articles.
    a. "It was hypothesized that the number of selfies posted . . . would be positively correlated with dimensions of narcissism" (Barry et al., 2017, p. 51).
    b. "We hypothesized that preschoolers' self-reported preferences for aggressive media . . . would be positively related to parent reports of children's aggressive behaviors" (Jamnik & DiLalla, 2018, p. 159).
    c. "We hypothesized that helicopter parenting would lead to higher levels of depression" (Darlow et al., 2017, p. 3).
    d. "We predicted that weight-related shame would be associated with greater binge eating" (Craven & Fekete, 2019, p. 45).
    e. "We hypothesize that men with strong gender identities will be more sensitive to manipulated threats to their masculinity compared with men with gender identities that are less central to their identity" (Harrison & Michelson, 2019, p. 65).
    f. "After exposure to fake news story with user comments identifying the news as fake, people are less likely to share the fake news than after exposure to the fake news story with user comments supportive of the content" (Colliander, 2019, p. 204).
4. Name the scale of measurement (nominal, ordinal, interval, ratio) for each of the following variables:
    a. The amount of time needed to react to a sound
    b. Gender
    c. Score on the Scholastic Aptitude Test (SAT)
    d. Political orientation (not at all conservative, conservative, very conservative)
    e. Political affiliation (Democrat, Republican, Independent)
5. Name the scale of measurement (nominal, ordinal, interval, ratio) for each of the following variables:
    a. One's age (in years)
    b. Size of soft drink (small, medium, large, extra large)
    c. Voting behavior (in favor vs. against)
    d. IQ score
    e. Parent (mother vs. father)
6. A faculty member wishes to assess the relationship between students' scores on the SAT and their performance in college.
    a. What is a possible research hypothesis in this situation?
    b. What are the independent and dependent variables?
    c. How could you measure the variable "performance in college" at each of the four levels of measurement?

7. A researcher hypothesizes that drivers who use cellular phones will get into a greater number of traffic accidents than drivers who do not use these phones. For this example,
    a. What are the independent and dependent variables?
    b. How could you measure the dependent variable?
    c. How could you conduct this study using an experimental research method? How could you conduct this study using a nonexperimental method?

Answers to Learning Checks

Learning Check 1

1.

    a. IV: Instructor's dress; DV: Expertise

    b. IV: Amount of television viewing; DV: Posttraumatic stress symptoms

    c. IV: Student group; DV: Financial distress

    d. IV: Activity while driving; DV: Driving performance

Learning Check 2

2.

    a. Nominal

    b. Ratio

    c. Ordinal

    d. Ratio

Answers to Odd-Numbered Exercises

1.

    a. IV: Type of business major; DV: Narcissistic traits

    b. IV: Biological sex; DV: Employment of physical characteristics

    c. IV: Community; DV: Perceptions of emotional pain and support needs

    d. IV: Generation; DV: Beliefs about group projects

e. IV: Country; DV: Stress and anxiety

f. IV: Price presentation; DV: Perception of savings

3.

a. IV: Number of selfies posted; DV: Narcissism

b. IV: Preference for aggressive media; DV: Parent reports of children's aggressive behaviors

c. IV: Helicopter parenting; DV: Depression

d. IV: Weight-related shame; DV: Binge eating

e. IV: Strength of gender identity; DV: Sensitivity to threat to masculinity

f. IV: Type of user comments to news story; DV: Sharing of news story

5.

a. Ratio

b. Ordinal

c. Nominal

d. Interval

e. Nominal

7.

a. IV: Cellular phone use; DV: Number of traffic accidents

b. Example: Determine the number of traffic accidents each person has experienced in the past year.

c. Example of experimental: Use a driving simulation with the "driver" talking on a phone and measure the number of potential driving mistakes or accidents.

d. Example of nonexperimental: Take a survey of the number of accidents people have been in and whether or not they were using a phone during the accident.

CHAPTER 2 EXAMINING DATA: TABLES AND FIGURES

Learning Objectives

2.1 Describe the main purpose and the research hypothesis of the “lottery” study.

2.2 Describe different reasons for examining data.

2.3 Describe the steps used to create a frequency distribution table.

2.4 Describe the steps used to create a grouped frequency distribution table.

2.5 Explain the differences between bar charts, pie charts, histograms, and frequency polygons.

2.6 Describe the three aspects and types of distributions of data.

Chapter 1 described a process by which research may be conducted, which begins by stating a research hypothesis. The next step involves collecting data that will ultimately be used in statistical analyses designed to test the research hypothesis. However, researchers don’t conduct analyses until they’ve first examined the data to make sure appropriate conclusions can be drawn from the analyses. This chapter describes *why* researchers examine data and *how* data may be examined using tables, charts, and graphs. As with many of the chapters in this book, this discussion will be guided by the findings from a published research study.

2.1 An Example From the Research: Winning the Lottery

A friend of yours flips a coin and asks you to guess whether it’s “heads” or “tails.” You guess “tails,” but the answer is “heads.” Your friend flips the coin a second time, and again you predict “tails.” But again the answer is “heads.” On the third coin flip, you guess “tails” once more—but, once again, you’re wrong. As your friend prepares to flip the coin a fourth time, you tell yourself there’s little chance the coin could be “heads” again. But the reality is that the fourth coin flip is equally likely to be “heads” or “tails” because the outcome of the first three coin flips has absolutely no impact on the fourth flip. In other words, the outcome of a coin flip is a random event that can’t be controlled or predetermined. But do you really believe this?

The impact of random events can take on greater significance than the outcome of a coin flip. For example, government-run lotteries provide the opportunity to win huge amounts of money through the random selection of numbers. In 2001, Dr. Karen Hardoon, a researcher at McGill University in Canada, studied the thought processes people use when purchasing lottery tickets (Hardoon et al., 2001). As previous research had found gamblers believe they can control random events such as the rolling of dice or the spinning of a roulette wheel, Dr. Hardoon and her colleagues were interested in

determining “whether individuals with gambling problems perceive the purchase of lottery tickets in a similar manner as non-problem gamblers” (p. 752). On the basis of earlier findings, Dr. Hardoon hypothesized that problem gamblers are less likely to believe lottery outcomes are random than are nonproblem gamblers.

To test their research hypothesis, Dr. Hardoon and her associates located a group of gamblers and asked whether they had ever experienced problems as a result of their gambling. Each participant was classified into one of two groups: problem gamblers and nonproblem gamblers.

The participants in the study, which we’ll refer to as the lottery study, were shown four lottery tickets and asked, “If you were to buy a ticket to play in the lottery, which one would you select?” Each ticket contained six numbers selected from among the numbers 1 through 49; the four tickets were given a particular label. The *sequence* ticket had numbers in consecutive order (30-31-32-33-34-35); numbers on the *pattern* ticket increased by 5s (5-10-15-20-25-30); the *nonequilibrated* numbers (35-37-40-43-44-49) were clustered at the upper end of the 49 numbers, and the *random* ticket (7-8-23-34-36-42) shared none of the characteristics of the other three tickets. We’ll call the variable in this study “lottery ticket,” a variable with four categories: sequence, pattern, nonequilibrated, and random.

The researchers recorded which of the four tickets each participant chose. Table 2.1 lists the data for the lottery ticket variable for the nonproblem gamblers (the data for the problem gamblers are provided in the exercises at the end of this chapter). Looking at this table, we see that the nonproblem gamblers differed in their lottery ticket choices; however, it’s difficult to describe these differences given the unorganized nature of the collected data. Consequently, before conducting statistical analyses on their data, researchers typically organize and examine them. In the following sections, we will discuss why and how data are organized and examined.

2.2 Why Examine Data?

There are a variety of reasons why researchers examine data they’ve collected before analyzing the data. These reasons include the following:

- To gain an initial sense of the data
- To identify outliers
- To evaluate research methodology
- To determine whether data meet statistical criteria and assumptions

Each of these reasons is discussed in the following sections.

Table 2.1 The Lottery Ticket Choices of 22 Nonproblem Gamblers

Participant	Lottery Ticket
1	**Random**
2	**Random**
3	**Nonequilibrated**

Participant	Lottery Ticket
4	**Pattern**
5	**Random**
6	**Sequence**
7	**Random**
8	**Random**
9	**Pattern**
10	**Pattern**
11	**Random**
12	**Pattern**
13	**Random**
14	**Nonequilibrated**
15	**Random**
16	**Pattern**
17	**Nonequilibrated**
18	**Random**
19	**Random**
20	**Pattern**
21	**Random**
22	**Random**

Gaining an Initial Sense of the Data

Because researchers spend a great amount of time and energy designing their research studies, they're eager to analyze the data they've collected. Students taking courses in statistics, on the other hand, face a different challenge: They're given sets of data with which they have no prior experience. Although both researchers and students may be motivated to analyze their data as quickly as possible, it's important to first examine the data in order to gain an initial sense of them.

As an example, imagine you conduct a study regarding the fuel efficiency of automobiles and collect data on the miles per gallon (MPG) of different types of cars. You might expect the scores in your data to range from 10 to 50 MPG, with the majority between 25 and 35 MPG. Examining data helps researchers gain an initial sense of whether the data are in line with their expectations. Examining data also provides a way of confirming the results of later statistical analyses. We've found that students' errors in their calculations could have been avoided had they first looked at the set of data. For example, examining data will help you avoid making statements such as, "The average MPG in my sample was 127 MPG."

Identifying Outliers

Researchers also examine their data to identify **outliers** , defined as rare, extreme scores that lie outside of the range of the majority of scores in a set of data. Imagine, for example, that a group of 15 students is asked to take a 20-item test of mathematical calculations; each student's score is the number of test items answered correctly. In scoring the tests, a researcher finds that 14 students

correctly answered between 8 and 17 questions. However, one student provided the wrong answer to every question for a score of 0. In this case, the score of 0 would be considered an outlier.

If not identified, outliers may distort conclusions researchers draw about their sample and their data, particularly when the sample is relatively small. In the previous example, the single score of 0 might lead an observer to conclude the class as a whole performed more poorly on the test than they actually did.

Evaluating Research Methodology

Examining data is also useful in assessing the effectiveness of the methods used to collect the data. For the 20-item math test, for example, it might be reasonable to expect the majority of scores to fall in the middle of the 0 to 20 range. But what if every student answered at least 17 of the 20 questions correctly? You might conclude the test was too easy and that more difficult questions are needed to accurately assess students' math skills. By comparing the collected data for a variable with the range of possible values, researchers are able to evaluate and modify their measurement tools.

Determining Whether Data Meet Statistical Criteria and Assumptions

Researchers also examine their data to see whether they meet certain statistical criteria and assumptions. Many of the statistics in this book are based on assumptions regarding the shape and nature of a set of data that has been collected. For example, these procedures assume that in most data sets, the majority of scores will fall in the middle of the range of possible values, with a small amount of scores at the highest and lowest possible values (you may have heard this referred to as "the bell curve"). In the example of the 20-item math test, it might be assumed that most scores will be between 7 and 13, with smaller number of scores less than 7 or greater than 13. When assumptions such as this are not met, it's possible to draw inappropriate conclusions from statistical analyses that are conducted.

2.3 Examining Data Using Tables

One part of examining data that have been collected for a variable involves answering a series of questions:

> What are the values of the variable?
>
> How many participants in the sample have each value of the variable?
>
> What percentage of the sample has each value of the variable?

As it may be difficult to answer questions such as these just by looking at the data in Table 2.1 , researchers typically organize their data by creating a table known as a **frequency distribution table** , a table that summarizes the number and percentage of participants for the different values of a variable. This section discusses the steps involved in creating frequency distribution tables using these questions.

What Are the Values of the Variable?

The first step in organizing the data for a variable is to identify all of the possible values for the variable. Table 2.2 begins the construction of a frequency distribution table by identifying the four values for the lottery ticket variable: sequence, pattern, nonequilibrated, and random.

How Many Participants in the Sample Have Each Value of the Variable?

Once the values of the variable have been identified, the next step in organizing a set of data is to determine the **frequency** of each value, which is the number of participants in a sample corresponding to a value of a variable. These frequencies help indicate the nature and shape of the distribution of scores for a variable in a sample.

Table 2.2 Identifying the Values of the Lottery Ticket Variable

Number	Lottery Ticket
1	Sequence
2	Pattern
3	Nonequilibrated
4	Random

Table 2.3 Using the Tick Mark Method to Determine the Frequencies of the Lottery Ticket Variable

Lottery Ticket	Tick Marks	*f*
Sequence	|	1
Pattern	||||— |	6
Nonequilibrated	|||	3
Random	||||— ||||— ||	12
Total	||||— ||||— ||||—||||—||	22

One way to determine the frequencies for a variable is to sort the data into the different values and then count the number of participants having each value using a “tick mark method,” where a tick mark is placed next to a value of the variable each time that value appears in the sample. As shown in Table 2.3 , the tick marks are then counted to determine the frequency for each value. For example, the number of tick marks and therefore the frequency (f) for the sequence ticket is 1 because it was picked by one participant (Participant 6 in Table 2.1). Similarly, the tick marks for each of the other three types of tickets can be recorded and counted until all of the data in the sample are accounted for.

What Percentage of the Sample Has Each Value of the Variable?

Although it’s important to determine the frequency for each value of a variable, these frequencies by themselves are not always particularly informative. For example, how you interpret “There are 35 Democrats in my sample” is different when the sample consists of 40 voters rather than 400 voters.

Because the interpretation of frequencies depends on the size of the sample, it's useful to calculate the *percentage* of the sample having each value of a variable. These percentages can be calculated using the following formula:

$$\% = \frac{f}{total\,number\,of\,scores} * 100$$

where *f* is the frequency for a value of a variable.

Looking at Table 2.3 , 1 of the 22 nonproblem gamblers chose the sequence ticket. Therefore, the percentage of the sample choosing this ticket is

$$\begin{aligned} \% &= \frac{f}{total\,number\,of\,scores} * 100 \\ &= \frac{1}{22} * 100 = .05 * 100 \\ &= 5\% \end{aligned}$$

After calculating the percentage for each of the four tickets, Table 2.4 provides the frequency distribution table for the lottery ticket variable.

Once a frequency distribution table has been constructed for a variable, researchers can begin to draw conclusions about their sample. For the lottery ticket variable, the random ticket had the highest frequency and percentage (*f* = 12, 54%) compared to the pattern (*f* = 6, 27%), nonequilibrated (*f* = 3, 14%), and sequence (*f* = 1, 5%) tickets. As a result, the researchers in the lottery study made the following observation:

"The results of the present study indicated that, for the entire sample . . . the greatest percentage of tickets selected by the entire sampl. . . . were random tickets." (Hardoon et al., 2001, p. 760)

Table 2.4 Frequency Distribution Table for the Lottery Ticket Variable

Lottery Ticket	***f***	**%**
Sequence	1	**5%**
Pattern	6	**27%**
Nonequilibrated	3	**14%**
Random	12	**54%**
Total	22	**100%**

Learning Check 1

Reviewing What You've Learned So Far



1. Review questions
    a. What are some reasons for examining data before conducting statistical analyses?
    b. Why are outliers of concern to researchers?
    c. What questions do you address about a set of data for a variable in constructing frequency distribution tables?
    d. Why is it useful to calculate percentages in addition to frequencies for each value of a variable?
2. For each of the following situations, create a frequency distribution table.
    a. A college counselor asks a group of 116 seniors what their plans are after graduating from college; 71 say they are going to “work,” 34 are planning on going to “graduate school,” and 11 say they are “not sure” what their plans are.
    b. A high school instructor teaching a course finds that of the students she teaches, 69 are Freshmen, 18 are Sophomores, 12 are Juniors, and 9 are Seniors.
    c. A pollster stops 12 people and asks if they are either “in favor” or “against” a local proposition in Table 2.5 :

Table 2.5 Example of People’s Position

Person	Position
1	**In favor**
2	**In favor**
3	**Against**
4	**In favor**
5	**Against**
6	**Against**
7	**In favor**
8	**Against**
9	**In favor**
10	**Against**
11	**In favor**
12	**In favor**

    d. The company that makes M&Ms candy (www.mms.com) conducted a survey asking people which new color they would like to have added: purple, aqua, or pink. Table 2.6 are the votes of a hypothetical sample:

Table 2.6
Hypothetical Sample of New M&M Color

Person	Color
1	**Pink**
2	**Purple**
3	**Pink**
4	**Aqua**
5	**Purple**
6	**Purple**
7	**Purple**
8	**Aqua**
9	**Purple**
10	**Pink**
11	**Aqua**
12	**Pink**
13	**Pink**
14	**Purple**
15	**Aqua**
16	**Pink**
17	**Purple**
18	**Purple**
19	**Aqua**
20	**Purple**
21	**Purple**
22	**Pink**
23	**Pink**
24	**Purple**

2.4 Grouped Frequency Distribution Tables

The frequency distribution table in Table 2.4 was sufficient for presenting the frequencies for the lottery ticket variable. However, other situations may involve variables with a large number of possible values. Consider, for example, the variable grade point average (GPA). The values for GPA typically range from 0.00 to 4.00, with hundreds of possible values in between. Creating a frequency distribution table for GPA could result in an extremely large table, with many values having frequencies of zero.

To examine a variable with a large number of values, it is useful to construct a **grouped frequency distribution table** , a table that groups the values of a variable measured at the interval or ratio level

of measurement into a small number of intervals and then provides the frequency and percentage for each interval. To illustrate how grouped frequency distribution tables are constructed, Table 2.7 (a) provides a sample of 42 hypothetical GPAs. In Table 2.7 (b), each of these 42 GPAs has been placed into one of eight intervals that comprise a grouped frequency distribution table.

To illustrate how to read and interpret a grouped frequency distribution table, the lowest interval in Table 2.7 (b) consists of students with GPAs less than (<) 2.00. The frequency in this interval (*f* = 1) consists of one student: Student 29 (GPA = 1.96). The next interval consists of students with GPAs ranging from 2.00 to 2.29 and contains three students (4, 22, and 38). This grouping continues until all of the GPAs in the sample have been accounted for.

To create a grouped frequency distribution table, the real limits of each interval should be identified. The **real limits** of an interval are the values of the variable that are halfway between the top of one interval and the bottom of the next interval. For example, because of how values of GPA are rounded, the lowest value of GPA that would lead someone to be placed in the interval 2.30–2.59 is 2.295, which is halfway between the top of the 2.00–2.29 interval and the bottom of the 2.30–2.59 interval. The **real lower limit** is the smallest value of a variable that would be grouped into a particular interval. In this case, 2.295 represents the real lower limit for the 2.30–2.59 interval. On the other hand, the **real upper limit** is the largest value of a variable that would be grouped into a particular interval. For the 2.30–2.59 interval, the real upper limit is 2.594.

Once the grouped frequency table for the GPA example has been created (Table 2.7(b)), what preliminary conclusions might be drawn about these students? First, because the 2.90–3.19 interval has the greatest frequency (*f* = 10), you might conclude that the typical student in this sample has a GPA close to 3.00. Second, moving in both directions away from the 2.90–3.19 interval, the frequencies of the different intervals grow progressively smaller, with fewer and fewer students having either very high or very low GPAs. Table 2.8 provides a list of guidelines for creating grouped frequency distribution tables.

Cumulative Percentages

In addition to determining the percentage of the sample that has each individual value or interval for a variable, it is sometimes useful to combine these percentages. For the GPA example, you may be interested in knowing, “What percentage of the sample had GPAs less than 2.60?” To answer this question, the percentages in the “%” column of the frequency distribution table must be combined. This type of percentage is referred to as a **cumulative percentage** , defined as the percentage of a sample at or below a particular value of a variable.

Table 2.7a Grade Point Average (GPA), 42 Students: Individual GPAs

Student	GPA
1	**3.05**
2	**2.83**
3	**3.26**
4	**2.19**

Student	GPA
5	**3.52**
6	**3.34**
7	**3.02**
8	**2.97**
9	**3.71**
10	**2.50**
11	**2.77**
12	**3.10**
13	**3.06**
14	**3.83**
15	**2.39**
16	**3.16**
17	**3.92**
18	**2.37**
19	**3.65**
20	**3.70**
21	**3.00**
22	**2.26**
23	**2.65**
24	**3.40**
25	**2.87**
26	**2.92**
27	**2.89**
28	**3.41**
29	**1.96**
30	**3.77**
31	**3.20**
32	**2.86**
33	**2.94**
34	**2.71**
35	**3.43**
36	**3.95**
37	**3.32**
38	**2.28**
39	**3.08**
40	**3.25**
41	**2.40**
42	**3.22**

Table 2.7b Grade Point Average (GPA), 42 Students: Grouped Frequency Distribution Table

GPA	*f*	%
3.80+	3	7%
3.50−3.79	5	12%
3.20−3.49	9	21%
2.90−3.19	10	24%
2.60−2.89	7	17%
2.30−2.59	4	10%
2.00−2.29	3	7%
<2.00	1	2%
Total	42	100%

Table 2.8 Guidelines for Creating Grouped Frequency Distribution Tables

Guideline	Guideline Description
1. *Variables are grouped into approximately 10 intervals.*	The exact number of intervals depends on the data in each sample. For example, data that fall within a small range of values may be accurately represented with a relatively small number of intervals, whereas data extending across a wide range may require a larger number of intervals.
2. *The number of intervals should accurately represent the data.*	The intervals should represent the nature of the data as accurately as possible. For example, there should not be so large a number of intervals that many intervals have frequencies of zero (e.g., 3.00−3.05) or so few intervals that a large majority of the sample falls into only one or two intervals (e.g., 3.00−3.99).
3. *Intervals should be of equal size.*	To make the intervals comparable, they should be of equal size or width. For example, you would not want one interval of 2.01−2.50 (a width of .50) and another of 2.51−2.70 (a width of .20). One exception to this guideline is the lowest or highest interval (e.g., less than 2.00 or greater than 3.70), which is typically wider than the other intervals because it contains a relatively small frequency.
4. *Intervals should not overlap.*	The intervals should be distinct from each other, such that each score may be placed in only one interval. For example, intervals of 2.30− *2.60* and *2.60*− 2.90 are a problem because the GPA 2.60 is in both intervals leading to confusion regarding within which interval to place someone with a GPA of 2.60.

Table 2.9 includes the cumulative percentages for the GPA example. As an example of a cumulative percentage, the cumulative percentage for the 2.30–2.59 interval (19%) has been calculated by combining the percentages for three intervals: < 2.00, 2.00–2.29, and 2.30–2.59: 2% + 7% + 10% = 19%. As a result, we could conclude that 19% of this sample had GPAs less than 2.60. The accumulating of percentages continues until all (100%) of the scores in the sample are combined. As

shown in Table 2.9 , cumulative percentages are typically located in the final column of a frequency distribution table and labeled “Cum %.”

Table 2.9 Frequency Distribution Table With Cumulative Percentages for the GPA Variable

GPA	*f*	%	Cum %
3.80+	3	7%	**100%**
3.50−3.79	5	12%	**93%**
3.20−3.49	9	21%	**81%**
2.90−3.19	10	24%	**60%**
2.60−2.89	7	17%	**36%**
2.30−2.59	4	10%	**19%**
2.00−2.29	3	7%	**9%**
< 2.00	1	2%	**2%**
Total	42	100%	

2.5 Examining Data Using Figures

In addition to organizing data into frequency distribution tables, researchers also examine data visually using figures such as charts or graphs. There are four types of figures often used to visually portray data for variables: bar charts, pie charts, histograms, and frequency polygons. Which type of figure a researcher may use depends on the variable’s level of measurement (nominal, ordinal, interval, ratio). Data for variables measured at the nominal or ordinal level of measurement are typically displayed using bar charts or pie charts. The values of interval and ratio variables are graphed using histograms and frequency polygons. Each of these types of visual illustration is discussed in the following sections.

Displaying Nominal and Ordinal Variables: Bar Charts and Pie Charts

The values of variables measured at the *nominal* level of measurement differ in category or type; one example of a nominal variable is the lottery ticket variable discussed earlier, which consisted of four categories: sequence, pattern, nonequilibrated, and random. The values of variables measured at the *ordinal* level of measurement can be placed in an order relative to the other values; age (young, middle age, elderly) is one example of an ordinal variable.

One way to display nominal and ordinal variables is to use a bar chart. A **bar chart** is a figure that uses bars to represent the frequency or percentage of a sample corresponding to each value of a variable, with the different bars not touching each other. Figure 2.1 shows a bar chart for the data from the lottery ticket variable. In this figure, each bar corresponds to one of the four types of lottery tickets; the height of each bar corresponds to the frequency (*f*) of nonproblem gamblers who selected that type of ticket.

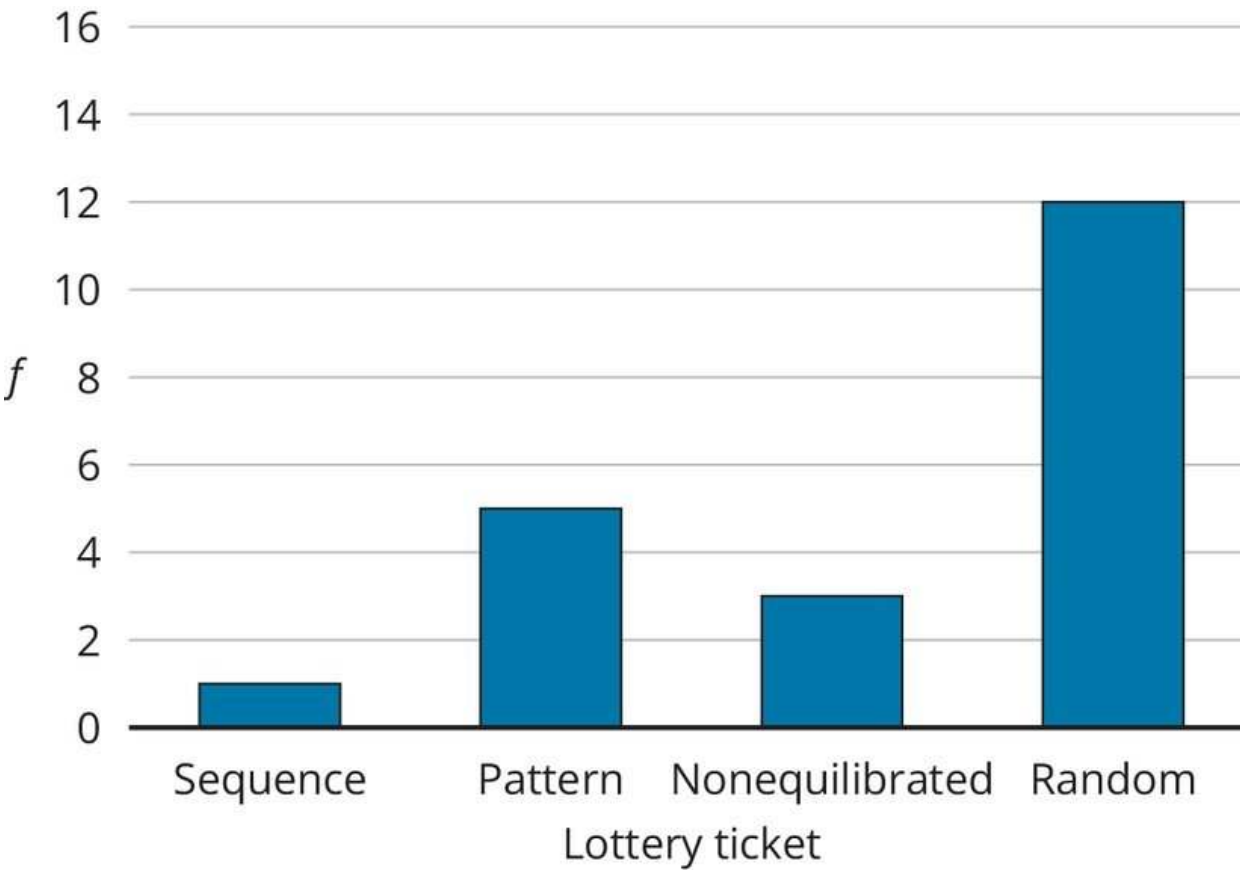


Figure 2.1 Bar Chart Showing the Frequencies of the Lottery Ticket Variable for the Nonproblem Gamblers

When creating a bar chart, the horizontal (X) axis contains the name of the variable and the values of the variable; the vertical (Y) axis contains the frequency (*f*) or percentage (%) for each value. Notice that the bars in the bar chart in Figure 2.1 do not touch each other; the gap between the bars indicates that the values of the variable represent distinct groups or categories that can't be placed along a numeric continuum.

A second type of figure used to display nominal or ordinal variables is a pie chart. A **pie chart** is a figure that uses a circle divided into proportions (slices) that represent the percentage of the sample for each value of a variable. Pie charts are created by dividing the 360 degrees of a circle into proportions corresponding to the percentage of each value.

Figure 2.2 shows a pie chart for the data for the lottery ticket variable. To illustrate how pie charts are constructed, the frequency distribution table in Table 2.4 shows that 5% of nonproblem gamblers chose the sequence ticket. Given that a pie chart consists of 360 degrees, 5% of the 360 degrees are needed to represent the sequence ticket. As 5% * 360 = 18, 18 degrees of the pie is assigned to the sequence ticket. Similarly, 97 degrees of the pie chart (27% of 360) are used to represent the pattern ticket. This process continues until all 360 degrees of the pie chart are accounted for.

The decision whether to use a bar chart or pie chart to illustrate a variable may depend on the number of values for the variable. For example, a pie chart for a variable that has 12 values may

result in many of the 12 “slices” being very thin and small, making them hard to see, meaning that a bar chart may be a better choice to illustrate this variable.

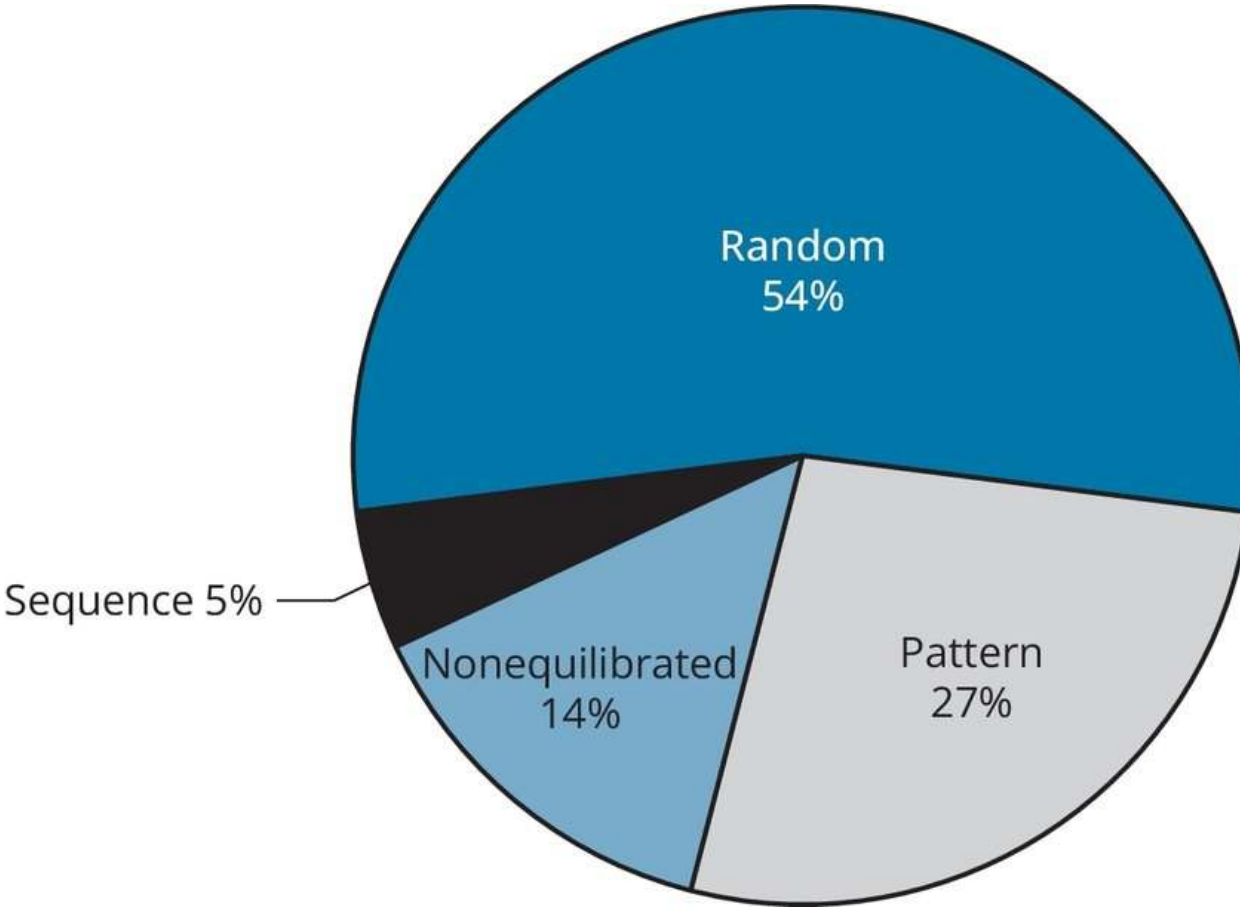


Figure 2.2 Pie Chart Showing the Percentages of the Lottery Ticket Variable for the Non−Problem Gamblers

Displaying Interval and Ratio Variables: Histograms and Frequency Polygons

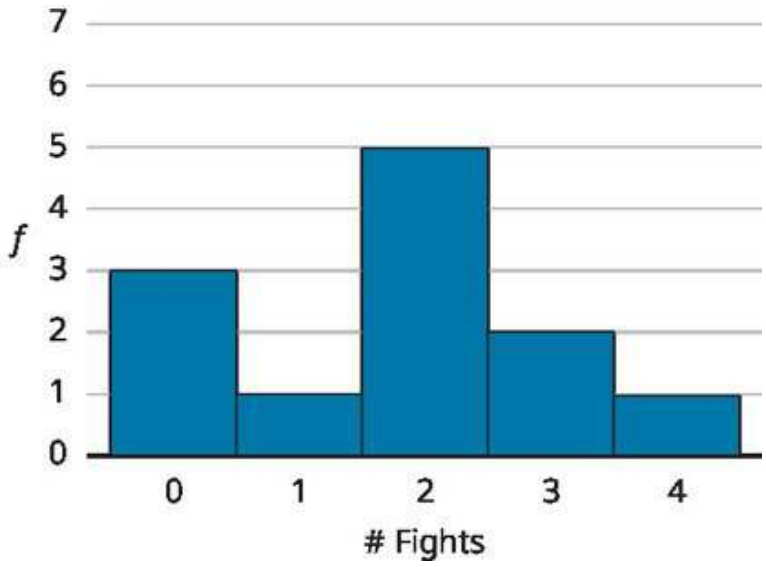
Interval and ratio variables have values equally spaced along a numeric continuum and are identical to each other with one exception: Ratio scales possess a true zero point, for which the value of zero (0) represents the complete absence of the variable. The values of interval and ratio variables may be illustrated using histograms and frequency polygons, the choice of which depends on the number of values for the variable.

A **histogram** is a figure in which bars are used to represent the frequency of each value of a variable, with the different bars touching each other. The name and values of the variable are located on the X-axis, and the frequency or percentage for each value is on the Y-axis. Histograms are similar to bar charts, with one important difference: The bars in a histogram touch each other, indicating that the values of the variable are connected to each other along a numeric continuum.

As an example, imagine a teacher is interested in studying aggressive behavior in children and measures this variable by counting the number of fights each of her students is involved in over a 3-day period. “Number of fights” is a ratio variable rather than an interval variable because the value of 0 represents the complete absence of fights. Figure 2.3 provides a frequency distribution table and histogram for this variable.

A histogram is used when an interval or ratio variable consists of a relatively small number of values. But what if, for example, the number of fights could be anywhere from 0 to 20 rather than 0 to 5? In order to have the histogram fit on one page, the 21 bars would need to be very thin and crowded together, making the histogram confusing and difficult to read. In situations such as this, the values of the variable may be more clearly illustrated using a figure known as a **frequency polygon**, which is a line graph that uses data points to represent the frequency of each value of a variable, with straight lines connecting the data points. In a frequency polygon, the X-axis represents the name and values of the variable, while the Y-axis represents the frequency for each value.

# Fights	*f*
4	1
3	2
2	5
1	1
0	3
Total	12



▶ Long Description: Figure 2.3 Frequency Distribution Table and Histogram for the Number of Fights Variable

Figure 2.3 Frequency Distribution Table and Histogram for the Number of Fights Variable

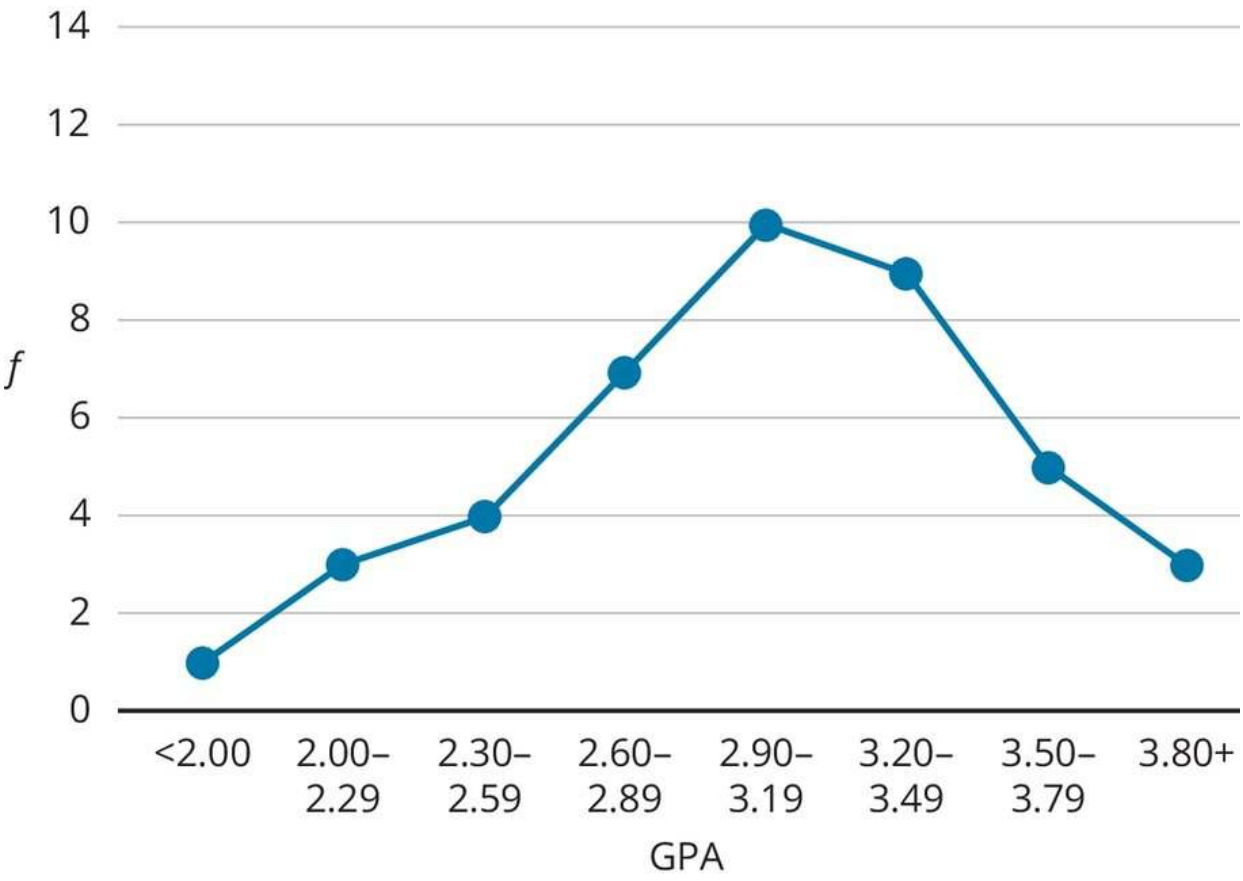


Figure 2.4 Frequency Polygon Showing the Frequency of Students in Each Interval of the Grade Point Average (GPA) Variable

Figure 2.4 shows a frequency polygon based on the grouped frequency distribution table of the earlier example of students' GPA (Table 2.7(b)). In this frequency polygon, the data points representing the frequencies of the eight GPA intervals are connected using straight, sharp lines. These lines indicate that the values of the variable can be connected together along a numeric continuum.

Although a frequency polygon like the one in Figure 2.4 may look jagged as it moves from one data point to the next, it provides a general sense of the shape of the distribution of scores for a variable. For example, looking at Figure 2.4 , you might conclude that the majority of GPAs in this sample are in the middle of the 1.00 to 4.00 range, with relatively low frequencies of students with either low or high GPAs.

Learning Check 2

Reviewing What You've Learned So Far

1. Review questions
    a. What is the difference between a frequency distribution table and a grouped frequency distribution table?
    b. For what levels of measurement would you create a pie chart, bar chart, histogram, or a frequency polygon?
    c. What is the difference between a bar chart and a histogram?
    d. Why would it be inappropriate to create a histogram or frequency polygon for a nominal or ordinal variable?
    e. Under what circumstances might you use a frequency polygon rather than a histogram to graph an interval or ratio variable?
2. For each of the following variables, identify whether you could use a bar chart, pie chart, histogram, and/or a frequency polygon to visually display the data.
    a. Miles per gallon (MPG)
    b. Living situation (on-campus, off-campus)
    c. Number of brothers and sisters
    d. Marital status (single, married, divorced)

Drawing Inappropriate Conclusions From Figures

A well-constructed figure can greatly facilitate the understanding of complex information. However, a poorly designed figure may not only be confusing, it can lead to inappropriate or misleading conclusions. Box 2.1 provides guidelines for constructing figures based on recommendations from the American Psychological Association (APA).

A simple example may demonstrate how altering the features of a figure can lead to different interpretations of the same set of data. Imagine a pollster assesses how voters with different political affiliations plan to vote on a ballot measure. She asks a sample of Democrats, Republicans, and Independents the question, “Do you plan on voting in favor of Measure A?” In examining her data, the pollster finds 56% of Democrats, 47% of Republicans, and 52% of Independents intend to vote “yes” on the measure.

Figure 2.5 graphs these three percentages using two bar charts; bar charts are used to display the data because the variable of interest (political affiliation) is a nominal variable (Democrat, Republican, or Independent). The bars don’t touch each other because the three political affiliations are distinct from each other and can’t be placed along a numeric continuum.

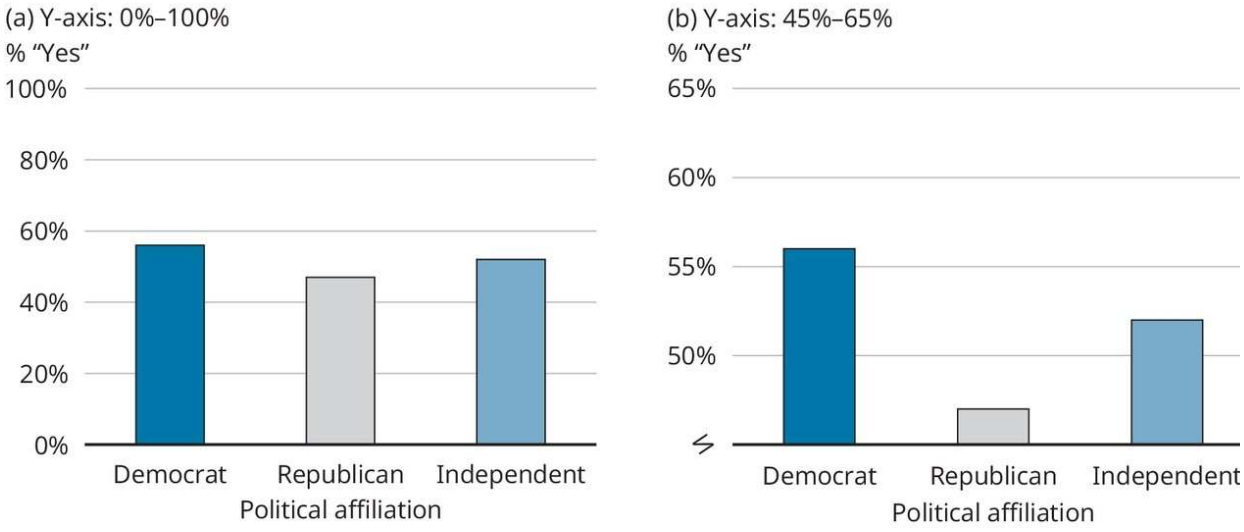
Although the two bar charts are similar, there is one important difference in how they are constructed. Following the guidelines in Box 2.1, the bar chart in Figure 2.5 (a) starts the Y-axis with the value 0% and ends with 100%. However, the bar chart in Figure 2.5 (b) starts with the value 45% and ends with 65%. By limiting the range of values of the percentages, the bar chart in Figure 2.5 (b) exaggerates differences in the heights of the three bars. As a result, someone looking at Figure 2.5 (b) may interpret the differences between the three political affiliations as being larger than someone looking at Figure 2.5 (a).

The saying “a picture is worth a thousand words” is particularly relevant for the presentation of statistical data. Because figures can provide a vivid representation of data, it’s the responsibility of researchers to portray their data in an accurate manner and to avoid creating images that may mislead or confuse.

Box 2.1 Some APA Guidelines for Creating Figures

1. The figure should be formatted in a portrait rather than landscape page orientation (the reader should not turn the page sideways to read the table).
2. The figure is numbered using Arabic numbers (Figure 1, not Figure *I* or Figure *A*).
3. The figure includes a title that describes the contents of the figure. For example, it is not enough to name a figure simply " Figure 2.1 " or " Figure 2.1 . Bar chart."
4. Figures have a rectangular shape, with the height of the vertical (Y) axis two thirds to three fourths the length of the horizontal (X) axis.
5. The horizontal (X) axis contains the values of the variable; the vertical (Y) axis contains the frequency (*f*) or percentage (%) of the sample having each value of the variable.
6. The *X* - and *Y* -axes include descriptive labels written in both upper- and lowercase letters (as opposed to all capital letters).
7. The labels for the *X* - and *Y* -axes are short enough to fit on one line along the axes.
8. The label of the Y-axis is often placed parallel to the Y-axis (the letters face sideways rather than vertically) but can sometimes be seen on top of the y-axis for accessibility.
9. The Y-axis starts with the value zero (0) and is divided into equally spaced values or intervals.
10. The Y-axis is divided into 6 to 10 values (just enough values to accommodate bar length). The bars should neither be flattened along the bottom of the Y-axis nor bumped up against the top of the Y-axis.

Political affiliation	% "Yes"
Democrat	56%
Republican	47%
Independent	52%



▶ Long Description: Figure 2.5 Different Displays of the Same Data

Figure 2.5 Different Displays of the Same Data

2.6 Examining Data: Describing Distributions

We've discussed how researchers construct tables and figures to organize their data. Looking at a table or figure enables researchers to describe the distribution of values or scores for a variable. Describing a distribution involves addressing questions such as, "Which values of the variable occur most often?," "How do the frequencies of the values of the variable change in relation to the most common values?," and "What is the amount of differences among the different values?"

Questions such as these pertain to three aspects of distributions: modality, symmetry, and variability. These three aspects of distributions are introduced in the following sections; they'll also be discussed in greater detail in the next two chapters of this book.

Modality

The first feature of a distribution of data for a variable is its modality. The **modality** of a distribution refers to the value or values of the variable that have the highest frequency or occur most often in a set of data. Using the lottery study as an example, identifying the modality of the lottery ticket variable involves answering the question, "Which of the four lottery tickets was chosen the most often?"

The modality of a distribution can be identified by looking at a frequency distribution table; the modality is the value(s) of the variable with the highest frequency (*f*) or percentage (%). The modality may also be determined by examining a figure and locating the tallest bar in a bar chart or histogram, the highest point in a frequency polygon, or the biggest "slice" in a pie chart. For the lottery ticket variable, the modality is the random ticket.

In identifying the modality of a distribution, it may be that one value or a limited number of values adjacent to one another clearly have the highest frequency. In such cases, the distribution is referred to as a unimodal distribution. A **unimodal distribution** is a distribution where one value of a variable occurs with the greatest frequency. Looking at the frequency distribution table (Table 2.4) or figure (Figure 2.1) for the lottery ticket variable, this distribution would be considered unimodal as one value, the random ticket, clearly occurred more often than the other three ticket choices.

In contrast to a unimodal distribution, it's possible that two distinct values may both have relatively high frequencies; if so, the distribution is referred to as a **bimodal distribution** , which is a distribution where two values clearly distinct from each other have the greatest frequency. A bimodal distribution sometimes resembles the two humps on a camel's back. As an example of a bimodal distribution, in April 2024 the Pew Research Center (www.pewresearch.org) asked a sample of 7,000 voters whether they felt Donald Trump had the mental fitness needed to be president, with five responses ranging from "extremely" to "not at all." The majority of voters chose one of the two extreme choices, either "extremely" (23%) or "not at all" (35%), with only 14% choosing the middle response of "somewhat."

And, in addition to unimodal and bimodal distributions, it's possible for a distribution to have more than two values with relatively equally high frequencies; this type of distribution is known as a **multimodal distribution** —a distribution where more than two values have the greatest frequency. In examining people's attitudes toward gun control, in May 2023, the news network CNN asked 1,227 people the question, "Do you think allowing gun owners to carry their guns in public places makes those places safer, less safe, or makes no difference in how safe a public place is?" They found people differed greatly in their responses, with 32% saying "makes public places safer," 36% saying

“makes public places less safe,” and 32% saying “makes no difference,” Figure 2.6 presents an example of a unimodal, bimodal, and multimodal distribution.

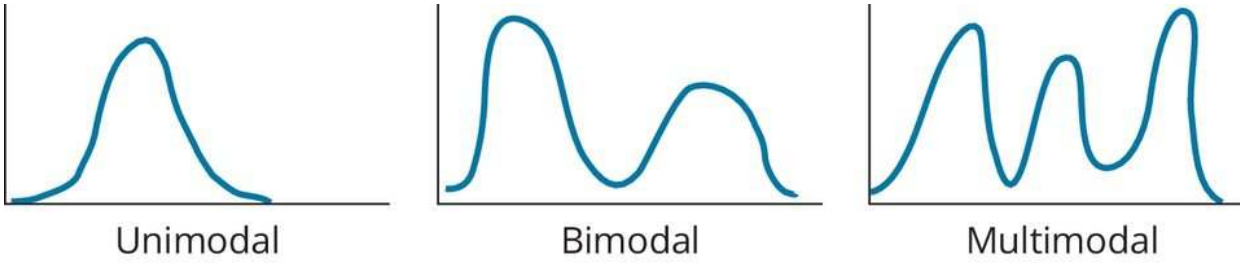


Figure 2.6 Distributions With Different Modalities (Unimodal, Bimodal, and Multimodal)

Symmetry

The second feature of a distribution of data is its symmetry. The **symmetry** of a distribution refers to how the frequencies of values of a variable change in relation to the most common or frequently occurring values. Describing the symmetry of a distribution means answering the question, “What is the shape of the distribution?”

The symmetry of a distribution can be determined by looking at the frequencies in a frequency distribution table or the bars or points in a figure. By doing so, a distribution may be described as being either symmetric or asymmetric. A **symmetric distribution** is one in which the frequencies change in a similar manner moving away in both directions from the most frequently occurring values. An example of a symmetric distribution is illustrated in Figure 2.7 . One way to determine whether a distribution is symmetric is to draw the distribution on a piece of paper and then fold the distribution vertically down the middle so that the two halves are on top of each other. A distribution is symmetric when the two halves are similar in shape such that they resemble mirror images of each other.

However, a distribution may also be asymmetric. An **asymmetric distribution** (also known as a *skewed* distribution) is one in which the frequencies change in a different manner moving away in both directions from the most frequently occurring values. In asymmetric distributions, the highest frequencies are located at one end of the distribution rather than in the middle, such that if you fold an asymmetric distribution down the middle, the shapes of the two halves do not look the same.

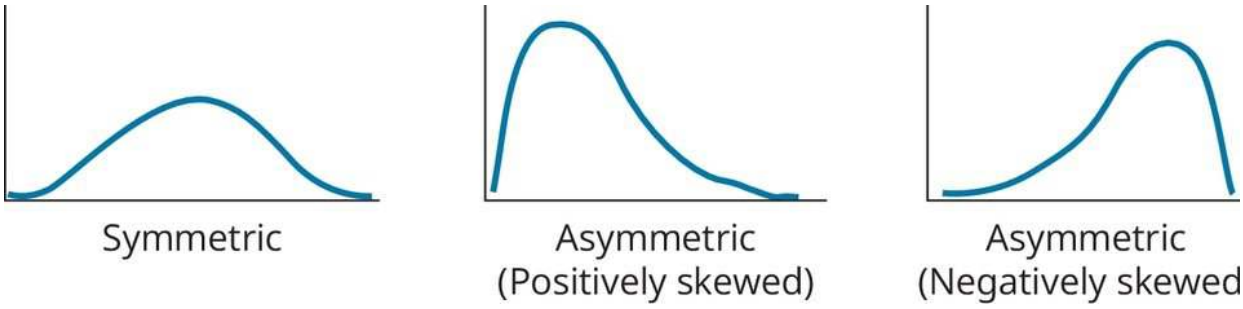


Figure 2.7 Distributions With Different Symmetries (Symmetric, Positively Skewed, and Negatively Skewed)

In an asymmetric distribution, the frequencies take on the shape of a long tail as the values move toward the end of the distribution; scores in the tail of the distribution are often considered to be outliers. For example, in the 2022 Current Population Survey, the U.S. Census Bureau (www.census.gov) estimated that most Americans (63%) had annual incomes less than $50,000. However, a very small percentage (3%) were greater than $200,000; these higher incomes would be considered outliers.

Depending on the location of its tail, asymmetric distributions are referred to either as positively skewed or negatively skewed. A **positively skewed distribution** is one in which the higher frequencies are at the lower end of the distribution, with the tail on the upper (right) end of the distribution; a **negatively skewed distribution** is one in which the higher frequencies are at the upper end of the distribution, with the tail on the lower (left) end of the distribution. Figure 2.7 provides an illustration of a symmetric distribution and two asymmetric distributions (one positively skewed and one negatively skewed). Annual income, illustrated by the U.S. Census Bureau example mentioned earlier, is an example of a positively skewed distribution.

Variability

The third way in which distributions may be described is in terms of the amount or nature of differences among the different values for a variable. The **variability** of a distribution refers to the amount of differences in a distribution of data for a variable. The variability in a distribution is described by answering the question, “To what extent are the scores in the distribution different from each other?”

The amount of variability in a distribution can be estimated by looking at a frequency distribution table or figure. Imagine, for example, an instructor gives an 11-item quiz to her class and finds almost all of the students answer six, seven, or eight of the items correctly. In this situation, the distribution would resemble the shape of a tall mountain: thin and tall, with many of the scores residing in a few values of the variable. A distribution with a small amount of variability such as this one is referred to as a **peaked distribution** , defined as a distribution where much of the data are in a small number of values of a variable.

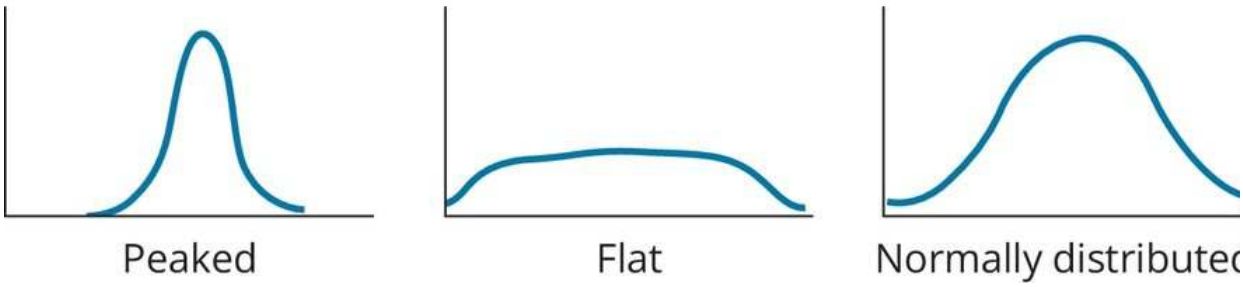


Figure 2.8 Distributions With Different Variability (Peaked, Flat, and Normally Distributed)

On the other hand, what would the distribution look like if she had found her students' scores were distributed relatively evenly across the entire range of possible scores? In this situation, the distribution would not be tall and thin like a mountain but rather would be low and wide like a plateau. This type of distribution is referred to as a **flat distribution** , which is a distribution where the data are spread evenly across the values of a variable. The CNN gun control example described earlier is an example of a flat distribution.

In discussing modality, symmetry, and variability, distributions may have more than one mode, may be asymmetric (skewed), and may be peaked or flat. But what if a distribution has none of these qualities? A distribution that is unimodal, symmetric, and neither peaked nor flat may be described as "bell shaped." If you look at a bell, you'll notice that it is the highest in the middle, and as you move away from the top of a bell in both the left and right directions, the shape of the bell slopes downward in the same manner. A bell-shaped distribution for a variable is given a special name: a normally distributed variable.

A **normally distributed variable** is a distribution for a variable that is considered unimodal, symmetric, and neither peaked nor flat. Note that the word *normal* is a statistical concept and not a value judgment. In other words, asymmetric, peaked, or flat distributions are not considered "abnormal." The concept of normal distributions will be discussed in greater detail in Chapter 5. Figure 2.8 provides an illustration of three distributions with different amounts of variability: peaked, flat, and normally distributed.

Learning Check 3

Reviewing What You've Learned So Far



1. Review questions
    a. What are the three aspects used to describe distributions of variables?
    b. What is the difference between unimodal, bimodal, and multimodal distributions?
    c. What is the difference between symmetric and asymmetric (skewed) distributions?
    d. What is the difference between positively and negatively skewed distributions?
    e. What is the difference between peaked and flat distributions?
    f. What are the characteristics of a normally distributed variable?
2. For each of the following sets of data, determine whether the distribution of scores is unimodal, bimodal, or multimodal.
    a. 4, 2, 11, 6, 10, 2, 6, 13, 2, 10, 6, 10, 1
    b. 15, 17, 13, 18, 17, 14, 20, 18, 17, 14, 17
    c. 20, 10, 50, 10, 80, 30, 70, 10, 60, 10, 80, 0, 30, 80, 10, 100, 80
3. For each of the following sets of data, determine whether the distribution of scores is symmetric or asymmetric (skewed).
    a. 4, 5, 1, 4, 2, 4, 3, 5, 3, 4, 4, 3
    b. 220, 230, 250, 230, 200, 240, 210, 230, 270, 220, 240, 230, 240, 230
    c. 3, 1, 9, 2, 3, 6, 2, 7, 3, 5, 2, 6, 3, 2, 2
4. For each of the following sets of data, determine whether the distribution of scores is peaked or flat.

a. 16, 10, 19, 12, 13, 15, 14, 17, 15, 20, 11, 13, 19, 16, 15
b. 12, 15, 12, 11, 13, 11, 12, 12, 14, 10, 12, 11, 16, 12, 13, 12, 11
c. 14, 12, 16, 14, 13, 15, 12, 14, 14, 15, 14, 14, 13, 14

2.7 Looking Ahead

This chapter has discussed why and how researchers examine data. Examining data by creating tables and figures requires a number of different skills, including organizational skills (to prepare a set of data for statistical analysis), communication skills (to present data in tables or figures), and conceptual skills (to describe a distribution). The next two chapters focus on mathematical skills related to the statistical analysis of data, describing how researchers describe distributions numerically rather than visually. Given that this is a statistics textbook, we will focus on calculating statistics to describe and analyze data. However, and we cannot say this strongly enough, it is *extremely* important to examine a set of data before conducting statistical analyses as this will prevent errors both in your calculations of statistics as well as your interpretation of the results of these calculations.

2.8 Summary

Before conducting statistical analyses, researchers typically examine their data to gain an initial sense of the data, detect data coding or data entry errors, identify ***outliers*** (rare, extreme scores that lie outside of the range of the majority of scores in a set of data), evaluate research methodology, and determine whether the data meet statistical criteria and assumptions.

One common way of examining variables involves creating tables known as ***frequency distribution tables*** (tables that summarize the number and percentage of participants that express the different values for a variable) and ***grouped frequency distribution tables*** (tables that group the values of a variable measured at the interval or ratio level of measurement into a small number of intervals and then provide the frequency and percentage within each interval). To create a grouped frequency table, the ***real limits*** of each interval (the values of the variable that fall halfway between the top of one interval and the bottom of the next interval) need to be identified.

Data for a variable may also be examined by creating figures such as bar charts, pie charts, histograms, and frequency polygons. ***Bar charts*** and ***pie charts*** are used for variables measured at the nominal or ordinal level of measurement; a bar chart uses bars to represent the frequency or percentage of a sample corresponding to each value of a variable, with the different bars not touching each other; a pie chart is a circle divided into proportions that represent the percentage of the sample corresponding to each value of a variable. ***Histograms*** and ***frequency polygons*** are used for variables measured at the interval or ratio level of measurement; a histogram is a figure in which bars are used to represent the frequency of each value of a variable, with the different bars touching each other; a frequency polygon is a line graph that uses data points to represent the frequency of each value of a variable, with lines connecting the data points.

From examining their data, researchers can describe the distribution of values or scores for a variable. There are three main aspects of distributions: modality, symmetry, and variability.

The ***modality*** of a distribution is the value(s) of the variable that have the highest frequencies or occur most often in a set of data. ***Unimodal distributions*** are distributions where one value occurs with the greatest frequency, ***bimodal distributions*** have two values with the greatest frequency, and ***multimodal distributions*** have more than two values with the greatest frequencies.

The ***symmetry*** of a distribution refers to how the frequencies of values of a variable change in relation to the most common or frequently occurring values of the variable. In ***symmetric distributions*** , the frequencies change in a similar manner, moving away in both directions from the most common values of the variable. In ***asymmetric (skewed) distributions*** , the frequencies change in a different manner, moving away in both directions from the most common values of the variable.

The ***variability*** of a distribution refers to the amount of differences in a distribution of data. In ***peaked distributions*** , much of the data are in a few values of a variable; in ***flat distributions*** , the data are spread evenly across the values of a variable. A ***normally distributed variable*** has a symmetric distribution that is neither peaked nor flat.

2.9 Important Terms

asymmetric (skewed) distribution

bar chart

bimodal distribution

cumulative percentage

flat distribution

frequency (*f*)

frequency distribution table

frequency polygon

grouped frequency distribution table

histogram

modality

multimodal distribution

negatively skewed distribution

normally distributed variable

outlier

peaked distribution

pie chart

positively skewed distribution

real limits

real lower limit

real upper limit

symmetric distribution

symmetry

unimodal distribution

variability

2.10 Formula Introduced in This Chapter

Percentage (%)

$$\% = \frac{f}{total\ number\ of\ scores} * 100$$

2.11 Using IBM SPSS Software *

Creating Data Files: Defining Variables and Entering Data

* IBM SPSS software. SPSS is a registered trademark of International Business Machines Corporation.

1. Open a new data file (See Figure 2.9).

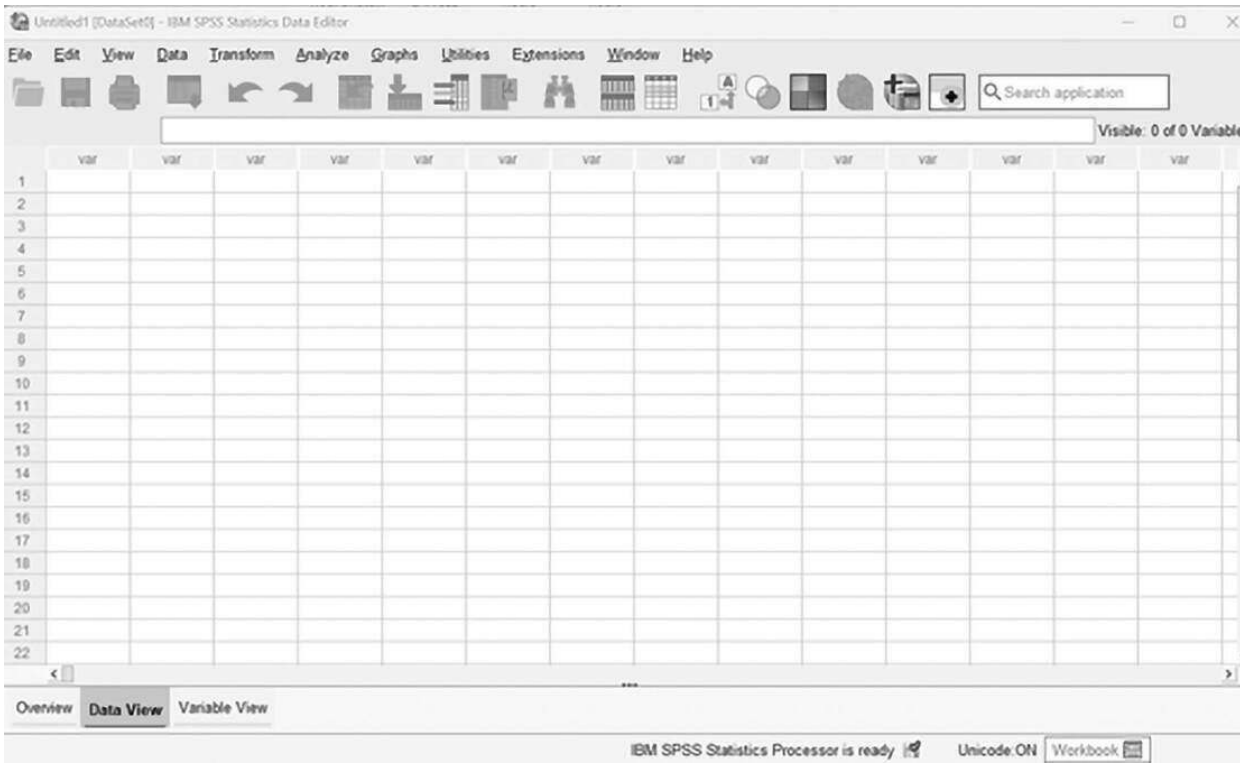


Figure 2.9 Data View

2. Define and name variable(s) (See Figure 2.10).

How? (1) Click **Variable View** tab, (2) click below **Name** , and (3) type name of variable in the cell.

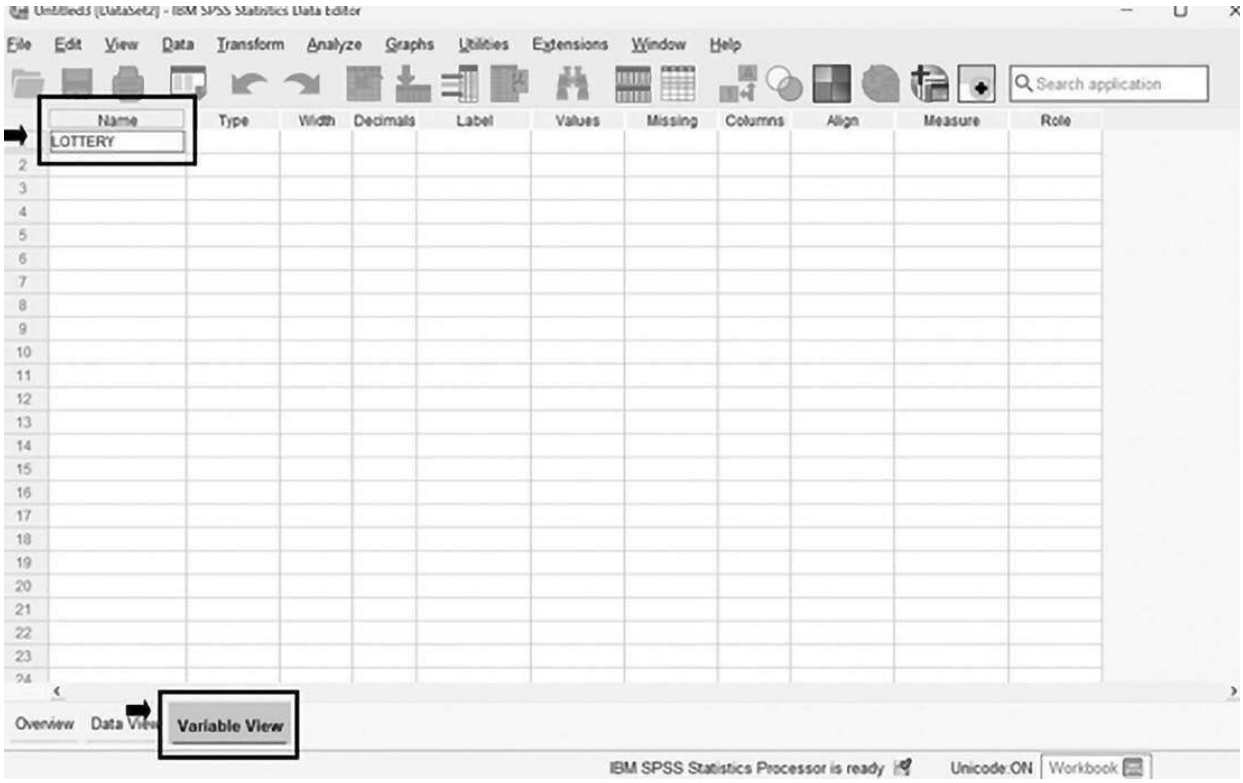


Figure 2.10 Variable View

3. Determine number of decimals for your variable(s) (See Figure 2.11).

How? Click desired number of decimal places in the **Decimals** box.

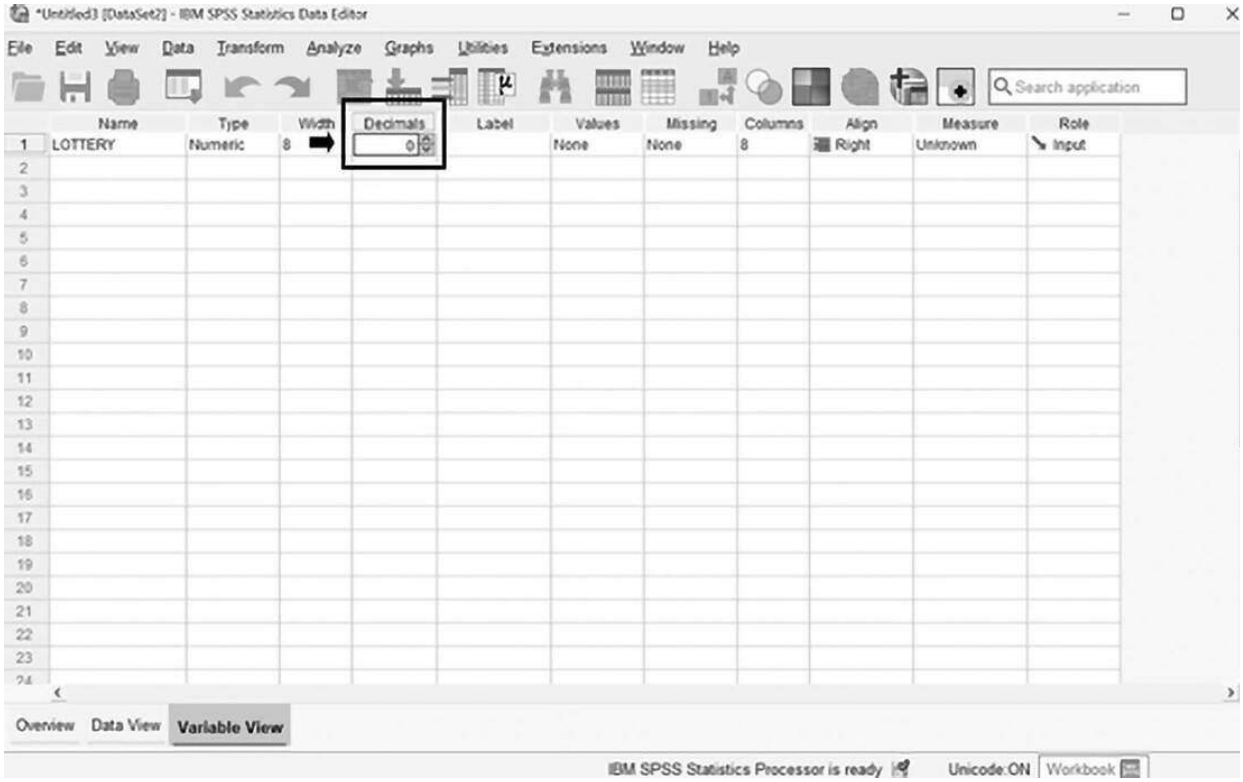


Figure 2.11 Decimals Box

4. Provide labels for your variable(s) (See Figure 2.12).

How? (1) Click below **Label** , and (2) type the label for the variable in the cell.

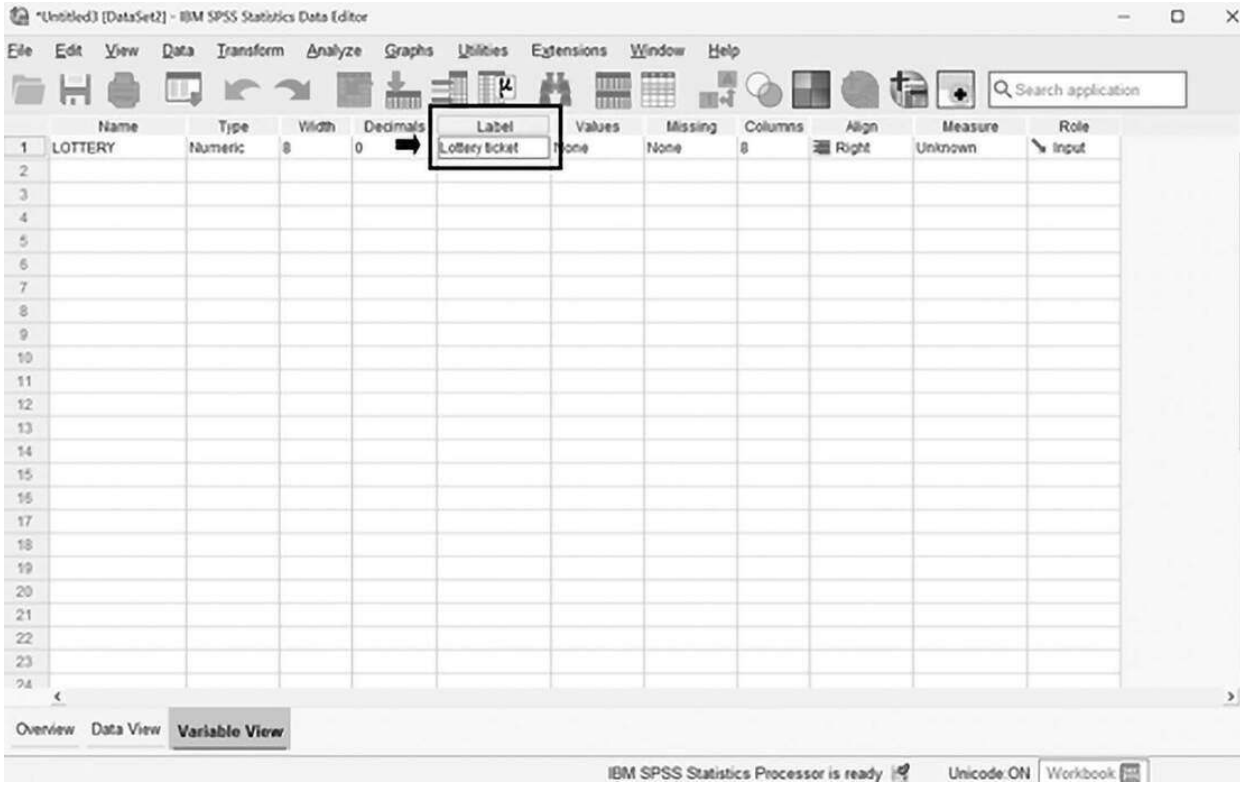


Figure 2.12 Label Box

5. Provide labels for values of nominal (categorical) variable(s) (See Figure 2.13).

How? (1) Click below **Values** , and (2) type labels for coded values of the variable.

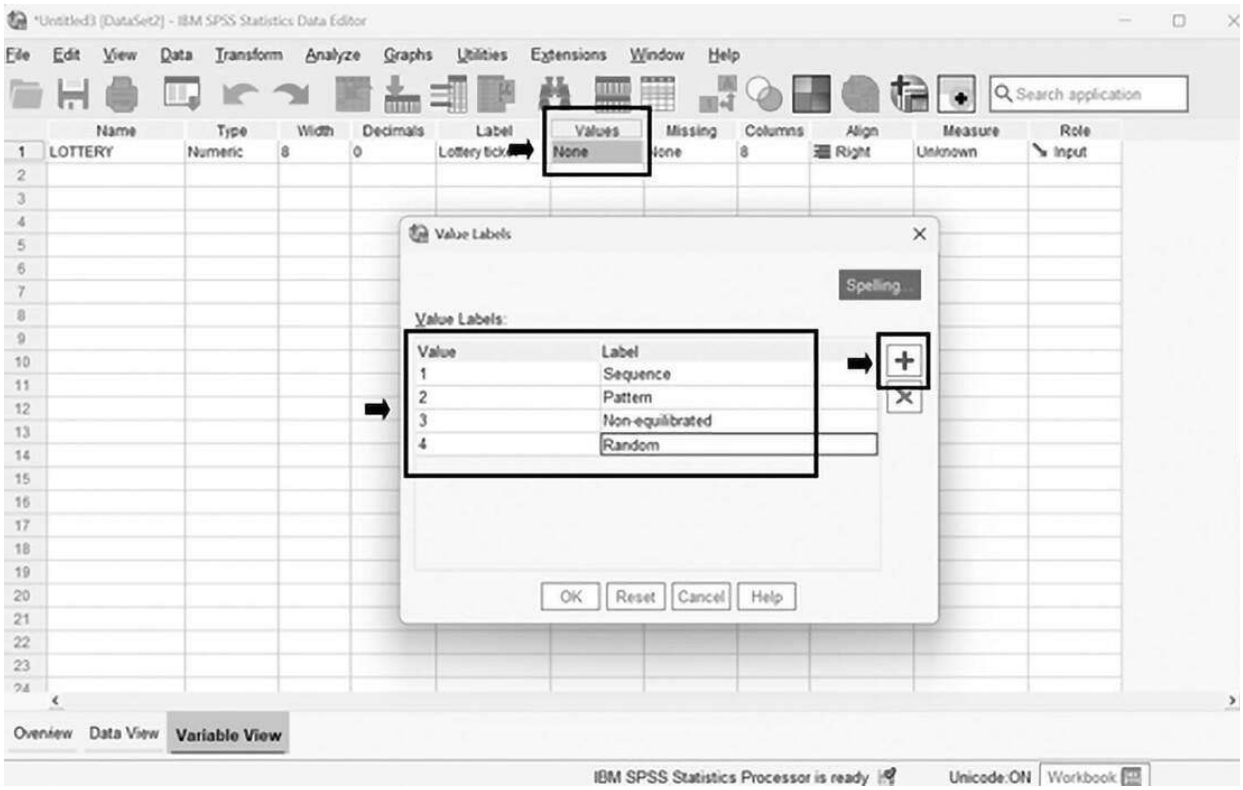


Figure 2.13 Nominal Variables

6. Enter data (See Figure 2.14).

How? (1) Click **Data View** tab and (2) enter data for the variable in the cells.

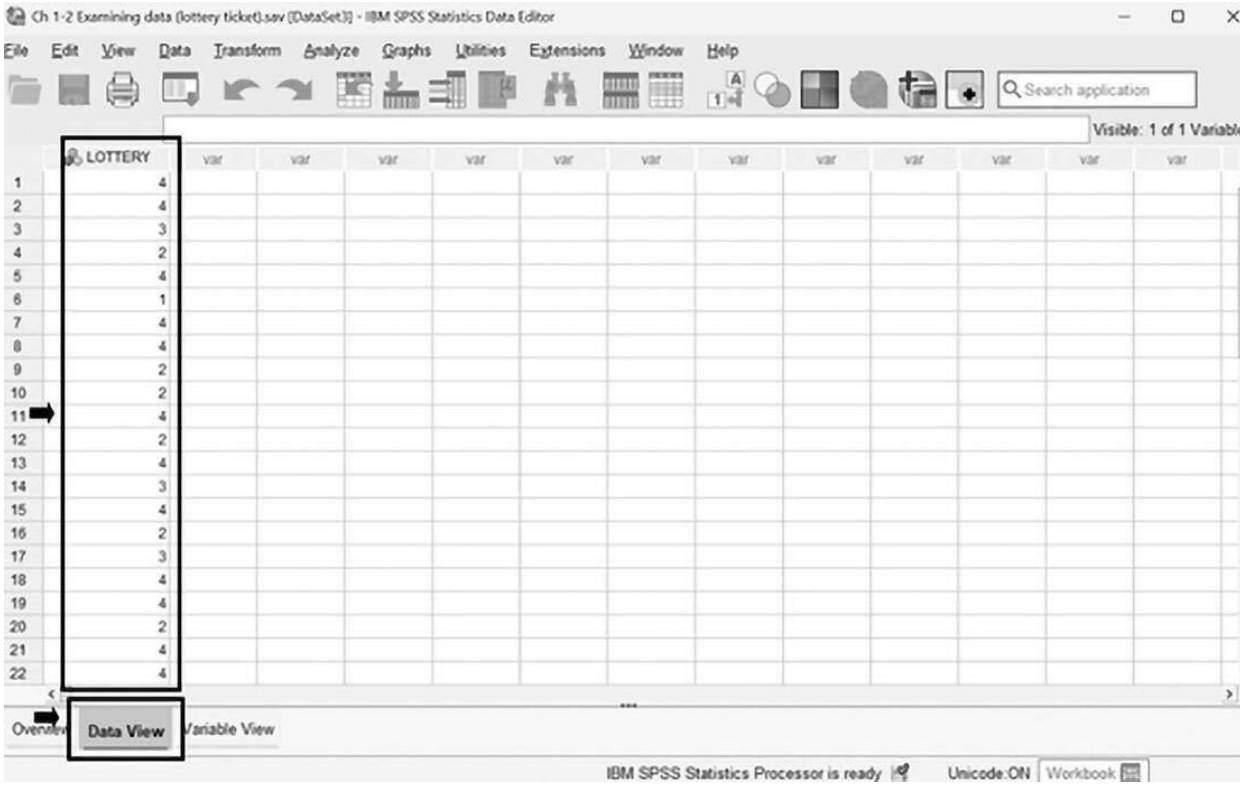


Figure 2.14 Enter Data View

Examining Data: Frequency Distribution Tables and Figures: The Lottery Study (2.1)

1. Select the frequency distribution procedure within SPSS (See Figure 2.15).

   How? (1) Click **Analyze** menu, (2) click **Descriptive Statistics** , and (3) click **Frequencies** .

join this discord server to get full pdf: discord.gg/EVabegZqJH

To Jennifer , my kids Meagan and Eika , and my parents, Katsumi and Grayce Tokunaga.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

As a global academic publisher, Sage is driven by the belief that research and education are critical in shaping society. Our mission is building bridges to knowledge—supporting the development of ideas into scholarship that is certified, taught, and applied in the real world.

Sage's founder, Sara Miller McCune, transferred control of the company to an independent trust, which guarantees our independence indefinitely. This enables us to support an equitable academic future over the long term by building lasting relationships, championing diverse perspectives, and co-creating social and behavioral science resources that transform teaching and learning.

▶ Long Description:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

FUNDAMENTAL STATISTICS FOR THE SOCIAL AND BEHAVIORAL SCIENCES

Third Edition

Howard T. Tokunaga

San Jose State University



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

eISBN: 978-1-0718-6160-8

Copyright © 2027 by Sage.

All rights reserved. Except as permitted by U.S. copyright law, no part of this work may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without permission in writing from the publisher.

All third party trademarks referenced or depicted herein are included solely for the purpose of illustration and are the property of their respective owners. Reference to these trademarks in no way indicates any relationship with, or endorsement by, the trademark owner.



FOR INFORMATION:

2455 Teller Road, Thousand Oaks, California 91320, E-mail: orders@sagepub.com

3rd floor, HYLO, 103-105 Bunhill Row, London, EC1Y 8LZ, United Kingdom

10th Floor, Emaar Capital Tower 2, MG Road Sikanderpur, Sector 26, Gurugram, Haryana – 122002, India

18 Cross Street #10-10/11/12, China Square Central, Singapore 048423

Library of Congress Control Number: 2026943277

ISBN: 978-1-0718-6144-8

Printed in the United States of America

This book is printed on acid-free paper.

26 27 28 29 30 10 9 8 7 6 5 4 3 2 1

Acquisitions Editor: Anna Villarruel

Content Development Editor: Kenzie Offley

Editorial Assistant: Daria Terry

Production Editor: Syeda Aina Rahat Ali

Copy Editor: Michelle Ponce

Typesetter: Lumina Datamatics

Cover Designer: Candice Harman

Marketing Manager: Victoria Velasquez

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CONTENTS

Preface
Acknowledgments
About the Author
Chapter 1 Introduction to Statistics
Chapter 2 Examining Data: Tables and Figures
Chapter 3 Measures of Central Tendency
Chapter 4 Measures of Variability
Chapter 5 Normal Distributions
Chapter 6 Probability and Introduction to Hypothesis Testing
Chapter 7 Testing One Sample Mean
Chapter 8 Estimating the Mean of a Population
Chapter 9 Testing the Difference Between Two Means
Chapter 10 Errors in Hypothesis Testing, Statistical Power, and Effect Size
Chapter 11 One-Way Analysis of Variance (ANOVA)
Chapter 12 Two-Way Analysis of Variance (ANOVA)
Chapter 13 Correlation
Chapter 14 Linear Regression and Multiple Correlation
Chapter 15 Chi-Square
Appendix A: Tables
Appendix B: Review of Basic Mathematics
Glossary
References
Index

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

DETAILED CONTENTS

Preface

Acknowledgments

About the Author

Chapter 1 Introduction to Statistics

- 1.1 What Is Statistics?
- 1.2 Why Learn Statistics?
- 1.3 Introduction to the Stages of the Research Process
  - Developing a Research Hypothesis to Be Tested
    - Identifying a Question or Issue to Be Examined
    - Reviewing and Evaluating Relevant Theories and Research
    - Stating a Research Hypothesis: Independent and Dependent Variables
  - Collecting Data
    - Drawing a Sample From a Population
    - Determining How Variables Will Be Measured: Levels of Measurement
    - Selecting a Method to Collect the Data: Experimental and Nonexperimental Research Methods
  - Analyzing the Data
    - Calculating Descriptive Statistics
    - Calculating Inferential Statistics
  - Drawing a Conclusion Regarding the Research Hypothesis
  - Communicating the Findings of the Study
- 1.4 Plan of the Book
- 1.5 Looking Ahead
- 1.6 Summary
- 1.7 Important Terms
- 1.8 Exercises

Answers to Learning Checks

Learning Check 1

Learning Check 2

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Odd-Numbered Exercises

Chapter 2 Examining Data: Tables and Figures

2.1 An Example From the Research: Winning the Lottery

2.2 Why Examine Data?

Gaining an Initial Sense of the Data

Identifying Outliers

Evaluating Research Methodology

Determining Whether Data Meet Statistical Criteria and Assumptions

2.3 Examining Data Using Tables

What Are the Values of the Variable?

How Many Participants in the Sample Have Each Value of the Variable?

What Percentage of the Sample Has Each Value of the Variable?

2.4 Grouped Frequency Distribution Tables

Cumulative Percentages

2.5 Examining Data Using Figures

Displaying Nominal and Ordinal Variables: Bar Charts and Pie Charts

Displaying Interval and Ratio Variables: Histograms and Frequency Polygons

Drawing Inappropriate Conclusions From Figures

2.6 Examining Data: Describing Distributions

Modality

Symmetry

Variability

2.7 Looking Ahead

2.8 Summary

2.9 Important Terms

2.10 Formula Introduced in This Chapter

Percentage (%)

2.11 Using IBM SPSS Software *

Creating Data Files: Defining Variables and Entering Data

Examining Data: Frequency Distribution Tables and Figures: The Lottery Study (2.1)

2.12 Exercises

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 3 Measures of Central Tendency
3.1 An Example From the Research: The 10% Myth
3.2 Understanding Central Tendency
3.3 The Mode
Determining the Mode
3.4 The Median
Calculating the Median: Odd Number of Scores
Calculating the Median: Even Number of Scores
Calculating the Median When Multiple Scores Have the Median Value
3.5 The Mean
Calculating the Mean
Reporting Calculated Numbers
Rounding Calculated Numbers
Calculating the Mean From Frequency Distribution Tables
The Mean as a Balancing Point in a Distribution
The Sample Mean vs. the Population Mean: Statistics vs. Parameters
3.6 Comparison of the Mode, Median, and Mean
Strengths and Weaknesses of the Mode
Strengths and Weaknesses of the Median
Strengths and Weaknesses of the Mean
Choosing a Particular Measure of Central Tendency
3.7 Measures of Central Tendency: Drawing Conclusions
3.8 Looking Ahead
3.9 Summary
3.10 Important Terms
3.11 Formulas Introduced in This Chapter
Median (Odd Number of Scores)
Median (Even Number of Scores)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Mean
Mean (Frequency Distribution Table)
Mean (Balancing Point in a Distribution)
3.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 4 Measures of Variability
4.1 An Example From the Research: How Many Sometimes" in an Always"?
4.2 Understanding Variability
4.3 The Range
Strengths and Weaknesses of the Range
4.4 The Interquartile Range
Strengths and Weaknesses of the Interquartile Range
4.5 The Variance (s^2)
Definitional Formula for the Variance
Computational Formula for the Variance
Why Not Use the Absolute Value of the Deviation in Calculating the Variance?
Why Divide by N − 1 Rather Than N in Calculating the Variance?
4.6 The Standard Deviation (s)
Definitional Formula for the Standard Deviation
Computational Formula for the Standard Deviation
Eyeball Estimating" the Mean and Standard Deviation
4.7 Measures of Variability for Populations
The Population Variance (σ^2)
The Population Standard Deviation (σ)
Measures of Variability for Samples vs. Populations
4.8 Measures of Variability for Nominal Variables
The Index of Qualitative Variation
4.9 Measures of Variability: Drawing Conclusions

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

4.10 Looking Ahead
4.11 Summary
4.12 Important Terms
4.13 Formulas Introduced in This Chapter
Range
Interquartile Range
Variance (s^2) (Definitional Formula)
Variance (s^2) (Computational Formula)
Standard Deviation (s) (Definitional Formula)
Standard Deviation (s) (Computational Formula)
Population Variance (σ^2) (Definitional Formula)
Population Standard Deviation (σ) (Definitional Formula)
Index of Qualitative Variation (IQV)
4.14 Using SPSS
Calculating Measures of Central Tendency and Variability: The Frequency Rating Study (4.1)
4.15 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 5 Normal Distributions
5.1 Normal Distributions
What Are Normal Distributions?
Why Are Normal Distributions Important?
Is There Only One Normal Distribution?
5.2 The Standard Normal Distribution
What Is the Standard Normal Distribution?
Interpreting z-Scores: Distance From the Mean
Interpreting z-Scores: Relative Position Within the Distribution
Finding the Area Between the Mean and a z-Score

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Finding the Area Less or Greater Than a z-Score
Finding the Area Between Two z-Scores
5.3 Applying z-Scores to Normal Distributions
Transforming Normal Distributions Into the Standard Normal Distribution
Understanding the z-Score Formula
Example: Transforming SAT Scores Into z-Scores
Interpreting z-Scores: Distance From the Mean
Interpreting z-Scores: Relative Position Within Distribution
Finding the Area Between the Mean and a Score
Finding the Area Less or Greater Than a Score
Finding the Area Between Two Scores
Transforming Scores Into z-Scores: Drawing Conclusions
5.4 Standardizing Frequency Distributions
Example: Measuring Politicians' Behavior
Standardizing a Frequency Distribution
Interpreting Standardized Scores
Using Standardized Scores to Compare or Combine Variables
Characteristics of Standardized Distributions
When Can You Apply the Normal Curve Table to a Standardized Distribution?
5.5 Looking Ahead
5.6 Summary
5.7 Important Terms
5.8 Formulas Introduced in This Chapter
z-Score
Standardized Score
5.9 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 6 Probability and Introduction to Hypothesis Testing

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

6.1 A Brief Introduction to Probability
- What Is Probability?
- Why Is Probability Important to Researchers?
- Applying Probability to Normal Distributions
- Applying Probability to Binomial Distributions

6.2 Example: Making Heads or Tails of the Super Bowl

6.3 Introduction to Hypothesis Testing
- State the Null and Alternative Hypotheses (H_0 and H_1)
  - The Null Hypothesis (H_0)
  - The Alternative Hypothesis (H_1)
- Make a Decision About the Null Hypothesis
  - Defining a Low" Probability of a Statistic: Setting Alpha (α)
  - Identifying the Values of the Statistic With a Low Probability
  - Stating a Decision Rule
  - Calculating a Value of a Statistic
  - Making the Decision Whether to Reject the Null Hypothesis
- Draw a Conclusion From the Analysis
- Relate the Result of the Analysis to the Research Hypothesis
- Summary of Steps in Hypothesis Testing

6.4 Issues Related to Hypothesis Testing: An Introduction
- The Issue of Proof" in Hypothesis Testing
- Errors in Decision Making
- Factors Influencing the Decision About the Null Hypothesis
  - Sample Size
  - Alpha (α)
  - Directionality of the Alternative Hypothesis

6.5 Looking Ahead

6.6 Summary

6.7 Important Terms

6.8 Formula Introduced in This Chapter

Probability

6.9 Exercises

Answers to Learning Checks

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 7 Testing One Sample Mean
7.1 An Example From the Research: Do You Read Me?
7.2 The Sampling Distribution of the Mean
Characteristics of the Sampling Distribution of the Mean
7.3 Inferential Statistics: Testing One Sample Mean (σ Known)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: z-Test for One Mean
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the z-Test for One Mean
Summary
7.4 A Second Example From the Research: Unique Invulnerability
7.5 Introduction to the t-Distribution
Characteristics of the Student t-Distribution
7.6 Inferential Statistics: Testing One Sample Mean (σ Not Known)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: t-Test for One Mean
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the t-Test for One Mean

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Summary
7.7 Factors Affecting the Decision About the Null Hypothesis
Sample Size
Alpha (α)
Directionality of the Alternative Hypothesis
7.8 Looking Ahead
7.9 Summary
7.10 Important Terms
7.11 Formulas Introduced in This Chapter
z-Test for One Mean
Population Standard Error of the Mean $\left(\mathbf{i}t\sigma_{\mathbf{i}t\overline{X}}\right)$
Degrees of Freedom (df) for One Sample
t-Test for One Mean
Standard Error of the Mean $\left(\mathbf{i}ts_{\mathbf{i}t\overline{X}}\right)$
7.12 Using SPSS
Testing One Sample Mean (σ Not Known): The Unique Invulnerability Study (7.6)
7.13 Exercises
Answers to Learning Checks
Learning Check 2
Learning Check 3
Learning Check 5
Learning Check 6
Answers to Odd-Numbered Exercises
Chapter 8 Estimating the Mean of a Population
8.1 An Example From the Research: Salary Survey
8.2 Introduction to the Confidence Interval for the Mean
Using the Sampling Distribution of the Mean to Create the Confidence Interval for the Mean
8.3 The Confidence Interval for the Mean (σ Not Known)
State the Desired Level of Confidence
Calculate the Confidence Interval and Confidence Limits
Calculate the Standard Error of the Mean $\left(s_{\overline{X}}\right)$

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Calculate the Degrees of Freedom (df) and Identify the Critical Value of t

Calculate the Confidence Interval and Confidence Limits

Draw a Conclusion About the Confidence Interval

Probability of an Interval or the Probability of a Population Mean?

Summary

8.4 The Confidence Interval for the Mean (σ Known)

State the Desired Level of Confidence

Calculate the Confidence Interval and Confidence Limits

Calculate the Population Standard Error of the Mean $\left(\sigma_{\overline{X}}\right)$

Identify the Critical Value of z

Calculate the Confidence Interval and Confidence Limits

Draw a Conclusion About the Confidence Interval

Summary

8.5 Factors Affecting the Width of the Confidence Interval for the Mean

Sample Size

Estimating the Sample Size Needed for a Desired Confidence Interval

Level of Confidence

Choosing Between Different Levels of Confidence: Probability Versus Precision

8.6 Interval Estimation and Hypothesis Testing

Concerns About Hypothesis Testing

The Interpretation of Statistical Significance and Nonsignificance

The Influence of Sample Size on the Decision About the Null Hypothesis

Interval Estimation as an Alternative to Hypothesis Testing

8.7 Looking Ahead

8.8 Summary

8.9 Important Terms

8.10 Formulas Introduced in This Chapter

Confidence Interval for the Mean (σ Not Known)

Confidence Interval for the Mean (σ Known)

Estimated Sample Size for the Confidence Interval for the Mean

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

8.11 Using SPSS
Calculating the confidence interval for the mean (The salary survey study (8.1))
8.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 9 Testing the Difference Between Two Means
9.1 An Example From the Research: You Can Just Wait
9.2 The Sampling Distribution of the Difference
Characteristics of the Sampling Distribution of the Difference
9.3 Inferential Statistics: Testing the Difference Between Two Sample Means
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: t-Test for Independent Means
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the t-Test for Independent Means
Summary
9.4 Inferential Statistics: Testing the Difference Between Two Sample Means (Unequal Sample Sizes)
An Example From the Research: Kids' Motor Skills and Fitness
Inferential Statistics: Testing the Difference Between Two Sample Means (Unequal Sample Sizes)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Relate the Result of the Analysis to the Research Hypothesis

Assumptions of the t-Test for Independent Means (Unequal Sample Sizes)

Summary

9.5 Inferential Statistics: Testing the Difference Between Paired Means

Research Situations Appropriate for Within-Subjects Research Designs

An Example From the Research: Web-Based Family Interventions

Inferential Statistics: Testing the Difference Between Paired Means

Calculate the Difference Between the Paired Scores

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Draw a Conclusion From the Analysis

Relate the Result of the Analysis to the Research Hypothesis

Assumptions of the t-Test for Dependent Means

Summary

9.6 Looking Ahead

9.7 Summary

9.8 Important Terms

9.9 Formulas Introduced in This Chapter

Degrees of Freedom (df), Difference Between Two Sample Means

t-Test for Independent Means

Standard Error of the Difference $\left(s_{\overline{X}_1-\overline{X}_2}\right)$

Standard Error of the Difference (Unequal Sample Sizes)

Degrees of Freedom (df), Difference Between Paired Means

t-Test for Dependent Means

Standard Error of Difference Scores (s D)

9.10 Using SPSS

Testing the Difference Between Two Sample Means: The Parking Lot Study (9.1)

Testing the Difference Between Paired Means: The Web-Based Intervention Study (9.5)

9.11 Exercises

Answers to Learning Checks

Learning Check 1

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Answers to Odd-Numbered Exercises
Chapter 10 Errors in Hypothesis Testing, Statistical Power, and Effect Size
10.1 Hypothesis Testing vs. Criminal Trials
10.2 An Example From the Research: Truth or Consequences
Correct Decisions and Mistakes in Criminal Trials
10.3 Two Errors in Hypothesis Testing: Type I and Type II Error
Type I Error: The Risk in Rejecting the Null Hypothesis
What Is the Probability of Making a Type I Error?
What Is the Probability of Not Making a Type I Error?
Why Is Type I Error a Concern?
Type II Error: The Risk in Not Rejecting the Null Hypothesis
What Is the Probability of Making a Type II Error?
What Is the Probability of Not Making a Type II Error?
Why Is Type II Error a Concern?
Why Does Type II Error Occur?
Type I and Type II Error: Summary
10.4 Controlling Type I and Type II Error
Controlling Type I Error
Concerns About Controlling Type I Error
Controlling Type II Error
Increasing Sample Size
Raising Alpha (α)
Using a Directional Alternative Hypothesis (H1)
Increasing Between-Group Variability
Decreasing Within-Group Variability
Controlling Type I and Type II Error: Summary
10.5 Measures of Effect Size
Interpreting Inferential Statistics: The Influence of Sample Size
The Inappropriateness of “Highly Significant”

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Effect Size as Variance Accounted for

Measure of Effect Size for the Difference Between Two Sample Means: r^2

Interpreting Measures of Effect Size

Presenting Measures of Effect Size

Reasons for Calculating Measures of Effect Size

A Second Measure of Effect Size for the Difference Between Two Sample Means: Cohen's d

10.6 Looking Ahead

10.7 Summary

10.8 Important Terms

10.9 Formulas Introduced in This Chapter

Probability of Type I Error

Probability of Type II error

r Squared (r^2)

Cohen's d

10.10 Exercises

Answers to Learning Checks

Learning Check 1

Learning Check 2

Learning Check 3

Learning Check 4

Answers to Odd-Numbered Exercises

Chapter 11 One-Way Analysis of Variance (ANOVA)

11.1 An Example From the Research: It's Your Move

11.2 Introduction to Analysis of Variance (ANOVA)

Review: Testing the Difference Between the Means of Two Samples

Understanding Between-Group and Within-Group Variability

Relating Between-Group and Within-Group Variability

Introduction to Analysis of Variance (ANOVA) and the F-Ratio

Characteristics of the F-Ratio Distribution

11.3 Inferential Statistics: One-Way Analysis of Variance (ANOVA)

State the Null and Alternative Hypotheses (H_0 and H_1)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Value, and State a Decision Rule
Calculate a Statistic: F-Ratio for the One-Way ANOVA
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (R^2)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Summary
11.4 A Second Example: The Parking Lot Study Revisited
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Value, and State a Decision Rule
Calculate a Statistic: F-Ratio for the One-Way ANOVA
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (R^2)
Draw a Conclusion From the Analysis
The Relationship Between the t-Test and the F-Ratio
11.5 Analytical Comparisons Within the One-Way ANOVA
Planned Versus Unplanned Comparisons
Between-Group and Within-Group Variance in Analytical Comparisons
Conducting Planned Comparisons
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Conducting Unplanned Comparisons
Concerns Regarding Unplanned Comparisons
Methods of Controlling Familywise Error

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Controlling Familywise Error: The Tukey Test

11.6 Looking Ahead

11.7 Summary

11.8 Important Terms

11.9 Formulas Introduced in This Chapter

Degrees of Freedom for Between-Group Variance, One-Way ANOVA (df_{BG})

Degrees of Freedom for Within-Group Variance, One-Way ANOVA (df_{WG})

Between-Group Variance, One-Way ANOVA (MS_{BG})

Total Sample Mean, One-Way ANOVA $\left(\overline{X}_T\right)$

Within-Group Variance, One-Way ANOVA (MS_{WG})

F-Ratio, One-Way ANOVA (F)

Measure of Effect Size, One-Way ANOVA (R^2)

Relationship Between t-Test and F-Ratio for the One-Way ANOVA

F-Ratio, Analytical Comparison (F_{comp})

Between-Group Variance, Analytical Comparison $\left(MS_{\mathrm{BG}_{\mathrm{comp}}}\right)$

Measure of Effect Size, Analytical Comparison $\left(R^2_{\mathrm{comp}}\right)$

Critical Value, Tukey Test for Unplanned Comparisons (F_T)

11.10 Using SPSS

One-Way Analysis of Variance (ANOVA): The Chess Study (11.1)

Analytical Comparisons: Planned Comparisons—Explanation vs. Prediction (11.5)

11.11 Exercises

Answers to Learning Checks

Learning Check 2

Learning Check 3

Learning Check 4

Answers to Odd-Numbered Exercises

Chapter 12 Two-Way Analysis of Variance (ANOVA)

12.1 An Example From the Research: Vote—or Else!

12.2 Introduction to Factorial Research Designs

Testing Interaction Effects

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

An Example of Interaction Effect

An Example of No Interaction Effect

Testing Main Effects

The Relationship Between Main Effects and Interaction Effects

Factorial Designs vs. Single-Factor Designs: Advantages and Disadvantages

12.3 The Two-Factor (A × B) Research Design

Notational System for Components of the Two-Factor Research Design

Analyzing the Two-Factor Research Design: Descriptive Statistics

Notational System for Means in the Two-Factor Research Design

Examining the Descriptive Statistics

12.4 Introduction to Analysis of Variance (ANOVA) for the Two-Factor Research Design

F-Ratios in the Two-Factor Research Design

12.5 Inferential Statistics: Two-Way Analysis of Variance (ANOVA)

State the Null and Alternative Hypotheses (H_0 and H_1)

Main Effect of Factor A

Main Effect of Factor B

A × B Interaction Effect

Make Decisions About the Null Hypotheses

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Values, and State Decision Rules

Calculate Statistics: F-Ratios for the Two-Way ANOVA

Make Decisions Whether to Reject the Null Hypotheses

Determine the Levels of Significance

Calculate Measures of Effect Size (R^2)

Draw Conclusions From the Analyses

Relate the Results of the Analyses to the Research Hypothesis

12.6 Investigating a Significant A × B Interaction Effect: Analysis of Simple Effects

12.7 Looking Ahead

12.8 Summary

12.9 Important Terms

12.10 Formulas Introduced in This Chapter

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Degrees of Freedom for Main Effect of Factor A, Two-Way ANOVA (df_A)
Degrees of Freedom for Main Effect of Factor B, Two-Way ANOVA (df_B)
Degrees of Freedom for A × B Interaction Effect, Two-Way ANOVA ($df_{A\times B}$)
Degrees of Freedom for Within-Group Variance, Two-Way ANOVA (dfWG)
Between-Group Variance for Main Effect of Factor A, Two-Way ANOVA (MS_A)
Between-Group Variance for Main Effect of Factor B, Two-Way ANOVA (MS_B)
Sums of Squares for Between-Group Variance, Two-Way ANOVA (SS_{BG})
Between-Group Variance for A × B Interaction, Two-Way ANOVA ($MS_{A\times B}$)
Within-Group Variance, Two-Way ANOVA (MS_{WG})
F-Ratio for Main Effect of Factor A, Two-Way ANOVA (F_A)
F-Ratio for Main Effect of Factor B, Two-Way ANOVA (F_B)
F-Ratio for A × B Interaction, Two-Way ANOVA ($F_{A\times B}$)
Measure of Effect Size for Main Effect of Factor A, Two-Way ANOVA(R^2_A)
Measure of Effect Size for Main Effect of Factor B, Two-Way ANOVA(R^2_B)
Measure of Effect Size for A × B Interaction, Two-Way ANOVA(R^2_{AB})
12.11 Using SPSS
Two-Way Analysis of Variance (ANOVA): The Voting Message Study (12.1)
12.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 13 Correlation
13.1 An Example From the Research: It's Good for You!
13.2 Introduction to the Concept of Correlation
Describing the Relationship Between Variables
Nature of the Relationship
Direction of the Relationship
Strength of the Relationship

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Measuring the Relationship Between Variables: Correlational Statistics
Introduction to the Pearson Correlation Coefficient (r)
Calculating Correlational Statistics: The Role of Variance and Covariance
13.3 Inferential Statistics: Pearson Correlation Coefficient
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: Pearson Correlation Coefficient (r)
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (r^2)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the Pearson Correlation
A Computational Formula for the Pearson r
Represent the Covariance Between X and Y (SPXY)
Represent the Variance of Variable X (SSX)
Represent the Variance of Variable Y (SSY)
Calculate the Pearson Correlation Coefficient (r)
13.4 Correlating Two Sets of Ranks: The Spearman Rank-Order Correlation
Introduction to Ranked Variables
An Example From the Research: Do Students and Teachers Think Alike?
Inferential Statistics: The Spearman Rank-Order Correlation Coefficient
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
13.5 Correlational Statistics vs. Correlational Research
Does ANOVA = Causality?
“Correlation Is Not Causality”—What Does That Mean?
13.6 Looking Ahead

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

13.7 Summary
13.8 Important Terms
13.9 Formulas Introduced in This Chapter
Degrees of Freedom for Pearson Correlation (df)
Sum of Products (Definitional Formula) (SPXY)
Sum of Squares for the X Variable (Definitional Formula) (SSX)
Sum of Squares for the Y Variable (Definitional Formula) (SSY)
Pearson Correlation (Definitional Formula) (r)
Measure of Effect Size for the Pearson Correlation Coefficient (r^2)
Sum of Products (Computational Formula) (SPXY)
Sum of Squares for the X Variable (Computational Formula) (SSX)
Sum of Squares for the Y Variable (Computational Formula) (SSY)
Pearson Correlation (Computational Formula) (r)
Spearman Rank-Order Correlation (rs)
Measure of Effect Size for the Spearman Rank-Order Correlation $\left(r_s^2\right)$
13.10 Using SPSS
Pearson Correlation: The Picky Eater Study (13.1)
Spearman Rank-Order Correlation: The Business Student Study (13.4)
13.11 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Answers to Odd-Numbered Exercises
Chapter 14 Linear Regression and Multiple Correlation
14.1 Predicting One Variable From Another: Linear Regression
The Linear Regression Equation
Calculating the Linear Regression Equation
Calculate the Slope of the Equation (b)
Calculate the Y-Intercept of the Equation (a)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Report the Linear Regression Equation

Drawing the Linear Regression Equation

A Second Example: Put Down Your Pencils

Calculating the Linear Regression Equation

Drawing the Linear Regression Equation

14.2 Correlation With Two or More Predictors: Introduction to Multiple Correlation and Regression

The Limitations of Using Only One Predictor

Introduction to Multiple Correlation

Multiple Correlation With Two Predictors: An Example

Measuring the Relationship Between Two Predictors and a Criterion: The Multiple Correlation (R)

Testing the Multiple Correlation for Statistical Significance

Predicting a Criterion From Two Predictors: The Multiple Regression Equation

Conclusion

14.3 Looking Ahead

14.4 Summary

14.5 Important Terms

14.6 Formulas Introduced in This Chapter

Linear Regression Equation

Estimate of Slope (b)

Estimate of Y-Intercept (a)

Multiple Correlation Coefficient (R)

F-Ratio for the Multiple Correlation Coefficient (F)

Multiple Regression Equation

Regression Coefficient

Estimate of Y-Intercept (a) (Multiple Regression)

14.7 Using SPSS

Linear Regression: The Picky Eater Study (14.1)

Multiple Regression: The College Admission Example (14.2)

14.8 Exercises

Answers to Learning Checks

Learning Check 1

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 2

Answers to Odd-Numbered Exercises

Chapter 15 Chi-Square

15.1 An Example From the Research (One Categorical Variable): Are You My Type?

15.2 Introduction to the Chi-Square Statistic

Observed and Expected Frequencies: The Basis of the Chi-Square Statistic

Characteristics of the Distribution of Chi-Square Statistics

15.3 Inferential Statistic: Chi-Square Goodness-of-Fit Test

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Value, and State a Decision Rule

Calculate a Statistic: Chi-Square (χ^2)

Make a Decision Whether to Reject the Null Hypothesis

Determine the Level of Significance

Calculate a Measure of Effect Size (Cramér's φ)

Draw a Conclusion From the Analysis

Relate the Result of the Analysis to the Research Hypothesis

Assumptions Underlying the Chi-Square Statistic

Conducting the Chi-Square Goodness-of-Fit Test With Unequal Hypothesized Proportions

Calculate Expected Frequencies (f_e)

Calculate the Chi-Square Statistic (χ^2)

15.4 An Example From the Research (Two Categorical Variables): Seeing Red

15.5 Inferential Statistic: Chi-Square Test of Independence

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Value, and State a Decision Rule

Calculate a Statistic: Chi-square (χ^2)

Make a Decision Whether to Reject the Null Hypothesis

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Determine the Level of Significance
Calculate a Measure of Effect Size (Cramér's φ)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
15.6 Parametric and Nonparametric Statistical Tests
Parametric vs. Nonparametric Statistical Tests
Reasons for Using Nonparametric Statistical Tests
Examples of Nonparametric Statistical Tests
15.7 Looking Ahead
15.8 Summary
15.9 Important Terms
15.10 Formulas Introduced in This Chapter
Degrees of Freedom for the Chi-Square Goodness-of-Fit Test
Expected Frequency for the Chi-Square Goodness-of-Fit Test
Chi-Square Statistic (χ^2)
Cramér's φ for the Chi-Square Goodness-of-Fit Test
Degrees of Freedom for the Chi-Square Test of Independence
Expected Frequency for the Chi-Square Test of Independence
Cramér's φ for the Chi-Square Test of Independence
15.11 Using SPSS
Chi-Square Goodness-of-Fit Test—Equal Hypothesized Proportions: The Personality Study (15.1)
Chi-Square Goodness-of-Fit Test—Unequal Hypothesized Proportions: The Personality Study (15.3)
Chi-Square Test of Independence: The Color Preference Study (15.4)
Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Learning Check 6

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Odd-Numbered Exercises

Appendix A: Tables

Appendix B: Review of Basic Mathematics

I. Symbols

II. Terms

III. Addition, Subtraction, Multiplication, Division

A. All Positive Numbers

B. Adding All Negative Numbers

C. Adding a Combination of Positive and Negative Numbers

D. Subtracting Negative Numbers

E. Multiplying Two Negative Numbers

F. Multiplying a Combination of Positive and Negative Numbers

G. Dividing Two Negative Numbers

H. Dividing One Positive Number and One Negative Number

IV. Order of Operations

V. Fractions, Decimals, Percentages

A. Fraction

B. Decimal

C. Percentage

VI. Solving Simple Equations With One Unknown (X)

A. X Has a Value Subtracted From It (Solve for X Using Addition)

B. X Has a Value Added to It (Solve for X Using Subtraction)

C. X Is Multiplied by a Value (Solve for X Using Division)

D. X Is Divided by a Value (Solve for X Using Multiplication)

VII. Solving More Complex Equations With One Unknown (X)

A. X Is Multiplied by a Value and Then Has a Value Either Added to It or Subtracted From It (Solve for X Using Either Subtraction or Addition and Then Division)

B. X Is Divided by a Value and Then Has a Value Either Added to It or Subtracted From It (Solve for X Using Either Addition or Subtraction and Then Multiplication)

C. X Has a Value Added to It or Subtracted From It and Is Then Multiplied by a Value (Solve for X Using Division and Then Either Subtraction or Addition)

D. X Has a Value Either Added to It or Subtracted From It and Is Then Divided by a Value (Solve for X Using Multiplication and Then Either Subtraction or

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Addition)
Glossary
References
Index

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

PREFACE

A Few (?) Words About Me

When I entered college, I thought my path would take me to law school; however, taking an Intro to Psych class made me realize I enjoy the challenge of trying to understand the human mind. The college I attended, U.C. Santa Cruz, was a fairly unconventional university, but somehow in the midst of a sea of humanistic psychologists I became attracted to the empirical and methodological aspects of psychology by taking classes from instructors such as David Harrington and Dane Archer, who showed me that statistics could appeal to students if taught with a gentle, guiding approach that addresses questions relevant to students' lives. After graduating from college, I was able to get a job as a research assistant for a human resource consulting firm. Despite my lack of work experience, I was hired primarily as a result of having taken statistics and research methods courses, which taught me that learning statistics has benefits both inside and outside of the classroom.

Several years later, I started grad school at U.C. Berkeley, where two events critical to this book took place. First, serving as a teaching assistant, I found I really enjoyed helping students, particularly in statistics and research methods classes often approached with fear and suspicion. Second, I took graduate classes from Geoff Keppel, who had developed his own unique system for analyzing experimental research designs. His lectures and books showed me that statistics can be taught in a systematic way that highlights similarities rather than differences between different research situations. Geoff managed to transform something as daunting-sounding as a "3 x 2 x 4 research design" into the mathematical equivalent of playing with wooden toy alphabet blocks labeled "A," "B," and "C."

After leaving Cal, I was hired by San Jose State, where I teach an introductory statistics course aimed at students with a wide range of backgrounds, ability, and motivation. When I went to select a textbook for the course, for the first time I looked carefully at the wide range of offerings. What I found striking (and still find striking) was that the majority of books focused on providing formulas and very small sets of data designed to demonstrate how to correctly calculate the correct numbers from these formulas. Little emphasis, however, was given to what these numbers meant or implied. Given my own experiences learning and using statistics, I thought a book was needed that focusses on how statistics may be used to answer questions and test ideas within the larger research process.

Goal of this Book

This textbook, aimed primarily at students majoring in the behavioral and social sciences, includes many of the same topics and chapters as the majority of existing introductory statistics books. However, what I believe distinguishes this book is its thematically-based approach that places statistics within the research process. The primary purpose of this

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

book is to not just teach students how to calculate statistics but how to interpret the results of statistical analyses in light of a study's research hypothesis, and to communicate one's results and interpretations to a broader audience. Hopefully, this book will not only help students understand the purpose and use of statistics, but it will also give them a greater appreciation of how research studies are conceived, conducted, and communicated.

The book is intended for students with a limited knowledge of mathematics, statistics, and research design and methodology. Some common obstacles faced by students are their math anxiety, fear of conducting arithmetic calculations, and an inability to see the relevance of the material to their daily lives. To address these concerns, this books attempts to judiciously use mathematical terminology, write in a conversational manner, and utilize actual research situations rather than contrived sets of numbers.

Organization of the Book

The 15 chapters of this book may be placed into three general categories. Chapters 1 to 4 are designed to introduce students to the research process and how data that have been collected may be organized, presented, and summarized:

Chapter 1 (Introduction to Statistics) defines the concept of statistics, explains why statistics are important to learn, and describes the steps in the research process that are the framework for the rest of the book.

Chapter 2 (Examining Data: Tables and Figures) discusses the importance of examining data that have been collected and describes how tables and figures may be created to represent data.

Chapter 3 (Measures of Central Tendency) presents measures of central tendency (the mode, median, and the mean) and compares them in terms of their relative strengths and weaknesses.

Chapter 4 (Measures of Variability) presents measures of variability (range, interquartile range, variance, standard deviation, index of qualitative variation) and the concept of parameters (population variance and standard deviation).

Chapters 5 through 10 discuss the process of conducting statistical analyses to test research questions and hypotheses, as well as issues and controversies regarding this process:

Chapter 5 (Normal Distributions) introduces theoretical distributions, starting with normal distributions, before moving on to the standard normal distribution (normal curve).

Chapter 6 (Probability and Introduction to Hypothesis Testing) presents the structure for later chapters by introducing, defining, and illustrating steps in hypothesis testing.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Chapter 7 (Testing One Sample Mean) tests one sample mean using the *z* -test and *t* -test and introduces sampling distributions, the central limit theorem, and the Student *t* distribution.

Chapter 8 (Estimating the Mean of a Population) introduces interval estimation by developing the confidence interval for the mean and discusses interval estimation as an alternative to null hypothesis significance testing (NHST).

Chapter 9 (Testing the Difference Between Two Means) covers the *t* -test for independent means for both equal and unequal sample size and introduces within-subjects research designs using the *t* -test for dependent means.

Chapter 10 (Errors in Hypothesis Testing, Statistical Power, and Effect Size) discusses Type I and Type II error and introduces measures of effect size as a way of drawing appropriate conclusions about an analysis.

Chapters 11 to 15 discuss different statistical procedures used in research situations that vary in the number of independent variables as well as how the variables have been measured:

Chapter 11 (One-way Analysis of Variance (ANOVA)) discusses the one-way ANOVA, the relationship between the *t* -statistic and the F-ratio, and planned and post-hoc comparisons.

Chapter 12 (Two-way Analysis of Variance (ANOVA)) introduces factorial research designs, the calculation of F-ratios for main effects and interaction effects, and illustrates the use of simple effects analyses to test research hypotheses.

Chapter 13 (Correlation) defines and illustrates the concept of correlation, discusses the Pearson and Spearman correlation coefficients, and makes the distinction between correlational statistics and correlational research methods.

Chapter 14 (Linear Regression and Multiple Correlation) introduces the concept of regression, discusses linear regression, and works through an example of multiple regression with two predictor variables.

Chapter 15 (Chi-Square) describes the analysis of categorical variables using the chi-square goodness-of-fit test and the chi-square test of independence before ending the book with a brief introduction to nonparametric statistics.

In most cases, the chapters in this book share a uniform format and structure, centered on the research process, based on examples of actual published research studies addressing topics relevant to students' lives. For each example, the research hypothesis is clearly identified; after a brief description of how the data in the study were collected, the calculation and interpretation of both descriptive and inferential statistics are presented,

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ending with the extent to which the results of the statistical analyses support the research hypothesis and how these findings may be communicated to a broader audience.

Pedagogical Features

The chapters in this textbook contain several reoccurring features designed to enhance students' comprehension of the material. These include the following:

Periodic learning checks within chapters. Short learning checks, in the form of conceptual review questions and computational exercises, are provided throughout each chapter.

Explicit presentation of formulas and calculations. This book provides both definitional formulas to highlight their relation to the concept being measured and computational formulas to facilitate calculations. In applying these formulas, all of the steps in the calculation of statistics such as the *t* -test and F-ratio are included to help students in completing exercise and homework assignments.

Illustration of the presentation and communication of analyses. Every chapter includes a discussion and illustration of how statistical analyses may be reported as well how they may be communicated to a larger audience in both written and visual form.

Highlighted definition of terms. Every term and concept introduced in this book is boldfaced, immediately followed by an explicit definition.

End of chapter summaries and exercises. The end of each chapter contains a summary of critical points, a list of terms and formulas introduced in the chapter, and a large number of exercises, many of which refer to published research studies.

Calculation of statistics using SPSS. Many chapters include end-of-chapter content that provides an annotated presentation of how statistics may be calculated using the SPSS software. Screen shots of data files, menu selections, and output files are included and annotated to locate and identify relevant aspects of the analysis.

What's New to the 3rd Edition

I'd like to start by expressing my most heartfelt appreciation to the instructors and reviewers of the first two editions of this book as their evaluation, insights, criticisms, and suggestions were absolutely vital to the changes made in this third edition. I'd like to particularly thank the instructors who've adopted the book who were kind enough to spontaneously send me their thoughts and suggestions. It is my sincere hope that all of you see the incorporation of your input in this edition.

There were two main goals in this edition. The first goal was to make the writing as brief and concise as possible, manifested in the goal of reducing the length of each chapter by 5% to

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

10%. This goal was achieved by simplifying the language and shortening or eliminating what I, the book's adopters, and the book's reviewers felt was unnecessary detail or discussion. The second goal was to increase and update the exercises at the end of many of the chapters, including the use of recent published research as the basis of these exercises. Another goal was to update the demonstration of the use of SPSS to analyze data using the most recent version (Version 31 at the time of this book's publication).

More specifically, in terms of content, the following changes were made to these chapters:

Chapter 1 (Introduction to Statistics): The published research studies used to illustrate concepts in the text, learning checks, and exercises were added to and updated using topics relevant to students (e.g., fear of missing out (FOMO), cryptocurrency, legalization of cannabis, Gen Z, COVID, fake news).

Chapter 6 (Probability and Introduction to Hypothesis Testing): The Super Bowl example used to introduce and illustrate the process of hypothesis testing was updated.

Chapter 8 (Estimating the Mean of a Population): The main research example, which involves data from an actual national salary survey, has been updated throughout this chapter with more current salary data.

As with the first two editions, any and all questions, concerns, suggestions, and complaints you have about the book are most welcome and can be sent to me at howard.tokunaga@sjsu.edu .

A Few Tips for Students

To you, the college student about to read this book as part of taking a statistics course: "Welcome!" and "Great job!" We welcome you because you're about to embark on a semester-long journey that we hope enhances your skills and widens your perspective; we congratulate you because it's a journey not everyone is willing to take. We say to our students that by taking just one statistics course they are ahead of 95% of the population in their knowledge of statistics.

We also tell our students that our primary goal for them is not to become a statistics genius but just to become comfortable with the process of analyzing and interpreting data. In that regard, we agree with a 2012 *Forbes* magazine article entitled *The Six Classes That Will Make Any College Grad Employable* that noted "a good statistics class will help the student become comfortable pushing numbers around and avoiding false interpretations."

At the present moment, we know our encouragement may be of little comfort because you might be anxious about having to learn statistics. Some of you might be anxious about having to learn statistical concepts and formulas; others might be anxious because the research process seems pretty complicated. Talking with students who have taken our courses over the years has helped us assemble the following advice:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Master the Material Presented in the Early Chapters

Chapters 1 through 6 discuss how research is typically conducted, how data are summarized and described, and the process by which researchers conduct statistical analyses to test their ideas. It is absolutely critical for you to have a firm understanding of these chapters as they lay the groundwork for later chapters that discuss a variety of statistical procedures used by researchers to test hypotheses.

What does it mean to master this material? First, *read the chapters both before and after they're discussed in class* . Reading the chapters beforehand helps you identify anything that's unclear and have your questions addressed by your instructor; re-reading the chapters afterward helps confirm your understanding of the material. Next, *be able to define and explain key conceptspresented in these chapters* . These concepts are highlighted in bold-faced letters, and it's important to learn them when they're first introduced because they'll appear throughout the remainder of the book. Next, *do the learning checks within the chapters and the exercises at the end of the chapters* , because they include exercises and review questions you can ask yourself to assess your understanding of the material. Most importantly, *do not miss class during the early part of the semester* ! It's been our experience that students who miss critical lectures at the beginning of the term often have difficulty keeping up as the semester continues.

Review High School Algebra

If you happen to move beyond the introductory statistics course for which you're reading this book, you'll find that statistics are often calculated using computers and statistical software rather than hand calculations. Consequently, some statistics textbooks no longer include mathematical formulas but instead have students use statistical software or Microsoft Excel. However, this book takes a different approach for two main reasons. First, we believe that in order to understand the results of statistical analyses, one must understand their underlying foundation. Learning statistics via computer software can lead to a "brain-dead" approach in which students are at the mercy (rather than control) of their computers. We refer to this as the "I'm only as smart as my printout" method of learning statistics, also known as "Because this is what my computer told me."

Second, this book contains mathematical formulas and calculations for one simple reason: The statistics in this book are not hard to calculate. To assess whether you're prepared to read this book, take the following test:

1. Do you know how to add, subtract, multiply, and divide?
2. Do you know how to calculate the square or square root of a number?
3. Do you understand the order of operations and parentheses within mathematical formulas?

If your answers to these questions are "Yes," congratulations! You possess the ability to conduct *every* mathematical calculation in this book because *none* of the formulas involve geometry, trigonometry, or calculus. However, if you're unsure of your mathematical ability, we suggest you review your high school algebra. Also, there's an appendix at the end of this

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

book that includes the mathematical concepts and operations needed to conduct the statistical analyses in this book. We highly recommend you read this appendix to review and assess your mathematical skills.

Learn to Use a Statistical Calculator

Although learning the statistics in this book doesn't require the use of computers or software, you will need a hand calculator. Calculators with statistical capabilities are typically identified as "statistical" or "scientific" calculators. Although there's a wide selection of brands and models, know that the calculators most appropriate for this book cost only $10 to $15. We do *not* recommend you use or purchase a more expensive calculator! We prefer these simpler calculators because students are able to do their calculations more quickly and with fewer errors. We do suggest you purchase a calculator similar to the one your instructor uses as they will be more familiar with its features and idiosyncrasies.

In selecting a calculator, look closely at the keyboard; you'll need a calculator whose keys have labels such as $\bar{X}$, σ, and ΣX (Chapters 3 and 4 discuss what these symbols represent). When you go calculator shopping, you'll find calculators that have keys with labels such as ΣY and XY. Although calculators with these Y symbols are only slightly more expensive than calculators with just the X symbols, we do *not* recommend their purchase because these features are not needed to perform the calculations in this book. In fact, we've found students with Y calculators have a more difficult time performing simpler calculations.

Ask Questions

Conducting research is the process of asking and answering questions. Accordingly, throughout this book issues are often framed in the form of questions, which we hope encourages you to ask questions as well. The first person you should direct any questions to is yourself. As you work your way through the chapters, get into the habit of asking yourself, "What does this mean?" A good test of your level of comprehension is whether you can explain the material to yourself in a meaningful way.

You should also be sure to work with your instructor to confirm your level of understanding and clarify unanswered questions. Instructors often say to students, "The only bad question is the one you don't ask." Don't be hesitant about asking questions in class. Instructors will tell you they use one student's question to assess an entire class's level of understanding. This is what we call the "pencil test"—when one student asks a question, we look to see how many other pencils get raised, which reflects the number of students who had the same question. Asking questions enhances the learning experience for yourself, your classmates, and your instructor.

Form a Study Group

No matter how much we encourage students to ask questions in class, it seems they hesitate for fear of drawing attention to themselves. As a result, we turn to another well-

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

traveled saying: “Misery loves company.” Forming a study group with other students will help you keep up with reading assignments, confirm your comprehension of the material, clarify any confusion you have, prepare for exams, and provide social support. We do suggest that your group contain at least one person who will raise the group’s questions and concerns to the instructor during class or office hours.

Practice, Practice, Practice

The best way to learn the topics covered by this book is to obtain as much practice as possible. Throughout the chapters are examples in which the calculations have been worked out for you; in the middle and end of each chapter are a number of exercises. The answers to some of the exercises are provided at the back of this book, and your instructor has access to the answers to the other exercises. Do the end-of-chapter exercises even if your instructor does not assign them as homework in order to identify any reoccurring mistakes you make. We find that the vast majority of errors students make are simple computational errors, committed when students do the calculations too quickly.

Also note that the publisher of this book, Sage Publications, has developed a webpage for this book that contains a huge amount of supplemental tools such as videos and quizzes for each chapter. Find it by going to the Sage Publishing website **collegepublishing.sagepub.com** , searching for my name (Tokunaga), and looking for the **Resources** tab.

Enough cheerleading and preaching – let’s get started!

WELCOME!

GREAT JOB!

Additional Teaching Resources

Visit **collegepublishing.sagepub.com** , and navigate to the Resources tab on your book’s page to find the teaching materials designed to accompany this textbook. On this site you will find an array of materials that will save you time and to help you keep students engaged, including the following:

Learning management system cartridges that easily integrate with your course management system so student test results and graded assignments seamlessly flow into your gradebook

Test banks , aligned to Bloom’s Taxonomy, that provide a diverse range of test items, including multiple choice, true/false, and essay questions

Lecture notes that provide an outline and the key concepts in each chapter to aid in lecture preparation

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

PowerPoint ® slides that offer a flexible, accessible, and customizable solution for creating multimedia lectures

Tables and Figures from the book to support lecture preparation and class discussions

Sample course syllabi , including suggested models for structuring your course

Answers to even-numbered questions from the text help facilitate grading

SPSS datasets for use with exercises from the text.

Multimedia content , including original SAGE videos featuring tutorials with author Howard T. Tokunaga that bring concepts to life and appeal to diverse learners

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ACKNOWLEDGMENTS

This book has traveled a long journey, with many people playing both transient and persevering roles in its development. First and foremost, I'd like to thank my kids, Meagan and Will, for providing me perspective and inspiration amid all of the twists and turns we faced while this book was being written. I'm so very blessed to have you as my "burden." Away from home, I'm grateful to my graduate students at San Jose State—working on this book has given me much empathy for their thesis-related travails. However, they should be aware that the completion of this book ends the unspoken agreement that I not ask, "So how's your thesis?" if they don't ask, "So how's your book?" A number of friends have provided most welcome relief after long days staring at a computer screen, with particular thanks going to Ellie, Eileen, Susan, Bonnie, Ann, Lisa, and Terry. I would like to thank Charles Linsmeier and David Shirley for their many helpful comments and advice on earlier versions of this book, Julie West for her help in developing chapter exercises, and Youn-Jeong Choi and Naomi Ran for their careful examination of the chapters and exercises.

I would like to express my appreciation to everyone at SAGE who has played a role in shepherding this book to completion. First and foremost, I must express my gratitude to my editor, Vicki Knight. It has been a joy and an honor to work with someone with such a truly exceptional combination of experience, knowledge, perspective, and humor. In production I was fortunate enough to work with others on the SAGE team who made the process a pleasure: Yvonne McDuffee, editorial assistant; Katie Guarino, assistant editor; Nicole Elliott, executive marketing manager; Candice Harman, cover designer; and the copyeditor, Michelle Ponce. Furthermore, I am grateful to the reviewers who took the time to read the chapters, raise questions, and provide constructive comments and suggestions to assist me in clarifying and strengthening the manuscript. The following are the reviewers for third edition:

Erin M. Fekete, University of Indianapolis

Brooke Magnus, Boston College

Loretta Neal McGregor; Arkansas State University

Michael T. Moore, Adelphi University

Megan M. Ringel, St. Edward's University

Bettina Viereck, University of Hartford

Janice Weaver, Community College of Allegheny County

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ABOUT THE AUTHOR

Howard T. Tokunaga is professor of Psychology at San Jose State University, where he serves as coordinator of the MS Program in Industrial/Organizational (I/O) Psychology and teaches undergraduate and graduate courses in statistics, research methods, and I/O psychology. He received his bachelor's degree in psychology at U.C. Santa Cruz and his PhD in psychology at U.C. Berkeley. In addition to his teaching, he has consulted with a number of public sector and private sector organizations on a wide variety of management and human resource issues. He is author of *Moving from IBM SPSS to R and RStudio: A Statistics Companion* , and coauthor (with G. Keppel) of *Introduction to Design and Analysis* : *A Student's Handbook* . In his spare time, he enjoys the outdoors, museums, live theater, and being as physically, socially, and politically active as possible.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CHAPTER 1 INTRODUCTION TO STATISTICS

Learning Objectives

1.1 Define the concept of statistics.

1.2 Describe different reasons for learning statistics.

1.3 Describe the main steps in the research process and the tasks and concepts within each step.

In introducing this book to you, we assume you're a college student taking what is perhaps your first course in statistics to fulfill a requirement for your major or a general education requirement. If so, you may be asking yourself two questions:

What is statistics?

Why learn statistics?

The ultimate goal of this book is to help you begin to answer these two questions.

1.1 What Is Statistics?

You may not know it, but you encounter a variety of "statistics" in your day-to-day activities: the typical cost of going to college, the yearly income of the average college graduate, the average price of a car, and so on. So what exactly is "statistics"?

The Merriam-Webster dictionary defines **statistics** as *a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data* . When most people think about statistics, they focus on the "analysis" aspect of this definition—that is to say, they focus on numbers that result from analyzing data. However, statistics not only is concerned about how data are *analyzed* but also how data are *collected* and how the results of statistical analyses are *interpreted* and *communicated* . The main purpose of this book is to introduce, describe, and illustrate the role of statistics within the larger research process.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1.2 Why Learn Statistics?

There are many reasons why you should learn statistics. First, not only do you currently encounter statistics in your daily activities, but *throughout your life you are affected by the results of research and statistical analyses* . Which college you attend is based in part on test scores developed by researchers. You may have to take a personality or intelligence test to get a job. The choices of drugs and medicines available to you are based on medical research. Learning about statistics will help you become a more informed and aware consumer of research and statistical analyses that affect many aspects of your life.

A second reason for learning statistics is that *you may be required to interpret the results of statistical analyses* . Many college courses require students to read academic research journal articles. Evaluating published research is complicated by the fact that different people studying the same topic may come up with diverse or even opposing conclusions. Understanding statistics and their role in the research process will help you decide whether conclusions drawn in research articles are appropriate and justified.

Another reason for learning statistics is that *you might be required to do your own research* . College courses sometimes have students design and conduct mini-research studies; undergraduate majors might encourage students to do senior honors theses; graduate programs often require masters' theses and doctoral dissertations. Learning to collect and analyze data will help you address your own questions in an objective, systematic manner.

A final reason for learning statistics is that it may *help you in your future career* . The website Careercast.com conducts an annual survey in which they evaluate jobs on four dimensions: environmental factors, income, employment outlook, and stress. Recently, their #1 rated job was "data scientist," and the #2 rated job was "statistician." They noted, "A statistician's skill set can be used to break down and analyze large quantities of data. The demand for these skills spans a variety of industries, including marketing, banking, government, sports, retail, and even healthcare."

It's generally a good idea for students to maintain a healthy level of curiosity or even skepticism in regard to their education. However, we find that when it comes to learning statistics, students' feelings may be characterized as fear or anxiety. We understand these feelings, and we hope the benefits of learning statistics will become clear to you and help overcome any concerns you have.

1.3 Introduction to the Stages of the Research Process

Much of scientific research involves asking questions. Throughout this book, we'll examine a broad range of questions regarding human attitudes and behavior that contemporary researchers have asked and attempted to answer. The following are research questions we'll address in this chapter to introduce the stages of the research process:

> Does appealing to one's fear of missing out (FOMO) affect the intention to invest in cryptocurrencies?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Do pro-cannabis messages on social media lead to cannabis use?

Do parents differ in their awareness of their children's online activities?

Do nonbinary college students have the same experiences as cis men or cis women?

How might you try to answer questions such as these? You could base your answers on your personal beliefs, or you could adopt the answers given to you by others. But rather than relying on subjective beliefs and feelings, researchers in many academic disciplines test their ideas using the scientific method. The **scientific method** is a method of investigation that uses the objective and systematic collection and analysis of empirical data to test theories and hypotheses.

At its simplest, this book will portray the scientific method as consisting of five main steps or phases:

developing a research hypothesis to be tested,

collecting data,

analyzing the data,

drawing a conclusion regarding the research hypothesis, and

communicating the findings of the study.

Accomplishing each of these five steps requires completing a number of tasks, as shown in Figure 1.1 . Because this sequence of steps serves as the model for the research process we'll discuss throughout this book, each step is briefly introduced in the following sections. It's important to understand that the research process depicted in Figure 1.1 represents an ideal way of doing research. The "real" way, as you may discover in your own efforts, is often anything but a smooth ride but rather is filled with false starts, dead ends, and wrong turns.

Developing a Research Hypothesis to Be Tested

The initial stage—and the first step—of the research process is to develop a research hypothesis to be tested. A **research hypothesis** may be defined as a statement regarding an expected or predicted relationship between variables. A **variable** is a property or characteristic of an object, event, or person that can take on different values. One example of a variable is "U.S. state," a variable with 50 possible values (Alabama, Arkansas, etc.).

Research hypotheses are usually developed through the completion of several tasks:

identifying a question or issue to be examined,

reviewing and evaluating relevant theories and research, and

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

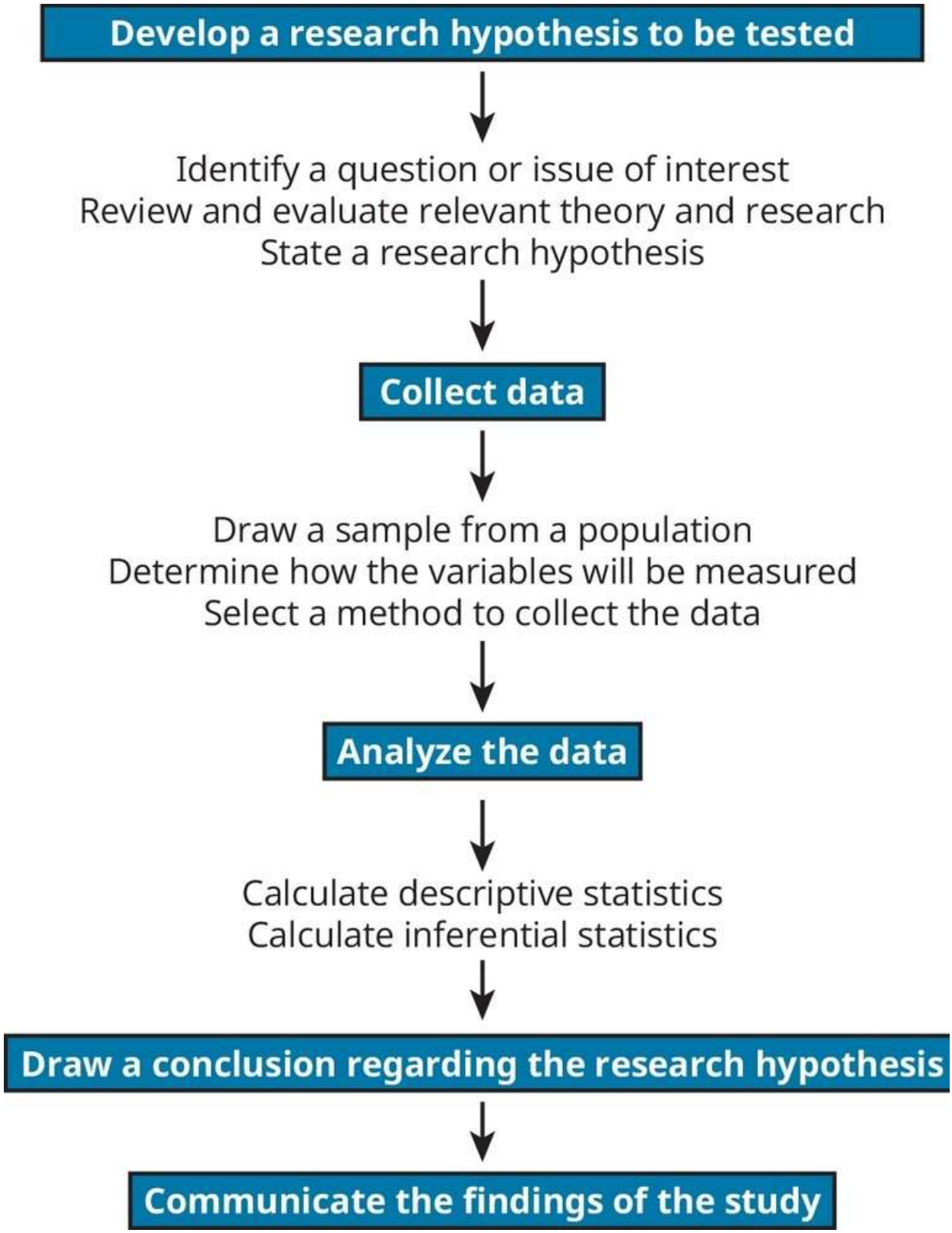
stating a research hypothesis.

Identifying a Question or Issue to Be Examined

Most research starts with a question posed by the researcher that often comes from the researcher's own ideas and daily observations. Although this may not seem terribly scientific, there's an advantage in using one's own experience as a starting point: People are generally much more motivated to explore issues that concern them personally. In teaching statistics, we advise students developing their own research projects to study something of interest to them. Conducting research can be tedious, difficult, and frustrating. At various points, you may ask yourself, "Why am I doing this?" Being able to provide a satisfactory answer to this question will help you overcome obstacles you encounter along the way.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

▶ Long Description: Figure 1.1 Steps in the Research Process Within the Scientific Method

Figure 1.1 Steps in the Research Process Within the Scientific Method

Reviewing and Evaluating Relevant Theories and Research

Beyond your own curiosity, research questions often arise from studying the theories, ideas, and research of others. A **theory** is a set of propositions used to describe or explain a phenomenon. The purpose of a theory is to summarize and explain specific facts using a logically consistent framework. Placing a question within a theoretical framework provides guidance and structure to research.

Reviewing and evaluating existing theories and research helps you decide whether it's worth the time and energy required to conduct a study. By seeing what others have done, you may conclude that the current way of thinking on a topic is incomplete or mistaken. By doing this review, you're able to ensure that the research you conduct adds to and improves upon an existing body of knowledge.

Stating a Research Hypothesis: Independent and Dependent Variables

Understanding and evaluating an existing literature not only helps develop a question of interest, but it may also lead to a predicted answer to that question. Within the scientific method, this answer is stated as a research hypothesis, defined earlier as a statement regarding an expected or predicted relationship between variables. Table 1.1 lists the research questions and research hypotheses for the studies listed at the beginning of this section. For example, the first research hypothesis states that "consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos."

One characteristic of research hypotheses such as those listed in Table 1.1 is that they identify the variables that are the focus of their research. As mentioned earlier, a variable is a property or characteristic with different values. Variables can be classified in several ways. In specifying a research hypothesis, researchers often speak in terms of "independent" and "dependent" variables. An **independent variable** may be defined as a variable manipulated or controlled by the researcher. A **dependent variable** , on the other hand, is a variable measured by the researcher that is expected to change or vary as a function of the independent variable. The independent variable is seen as the "cause" or "treatment," whereas the dependent variable is the "effect" or "outcome."

Researchers are interested in examining the effect of the independent variable on the dependent variable. Consider the first research hypothesis provided in Table 1.1 : "Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos." Here, the independent variable is the type of appeal consumers are exposed to: FOMO or non-FOMO. The dependent variable is the intention to invest in cryptos. In this study, the effect of the independent variable on the dependent variable is that

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

differences in the intention to invest in cryptos (the dependent variable) may “depend” upon which type of appeal (FOMO or non-FOMO) a consumer receives. Table 1.2 lists the independent and dependent variables for each of the research hypotheses in Table 1.1 .

In addition to identifying variables, a second characteristic of research hypotheses is that they specify the nature and direction of the relationship between variables. For example, the first research hypothesis in Table 1.1 states that “Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *higher* intention to invest in cryptos.” The word *higher* indicates the nature and direction of the relationship between the independent variable (type of appeal) and the dependent variable (intention to invest in cryptos). The direction of the relationship would not have been stated if the hypothesis had included the less specific phrase “will have a *different* intention,” which simply indicates that the two groups of consumers are not expected to be the same. Table 1.3 provides directional and nondirectional research hypotheses for the research studies from Table 1.1 .

Table 1.1 Examples of Research Questions and Hypotheses

Research Question	Research Hypothesis
Does appealing to one’s fear of missing out (FOMO) affect the intention to invest in cryptocurrencies? (Friederich et at., 2024)	Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.
Do pro-cannabis messages on social media lead to cannabis use? (Willoughby et al., 2024)	Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.
Do parents differ in their awareness of their children’s online activities? (Symons et al., 2017)	Mothers have more accurate knowledge of their children’s online activities than do fathers.
Do nonbinary college students have the same experiences as cis men or cis women? (Marx et al., 2024)	Nonbinary college students experience higher levels of victimization than do cis men or cis women.

The ability to form a directional research hypothesis is dependent on the state of the literature on the question of interest. If little or perhaps conflicting research has been conducted, researchers may not be able to form a directional hypothesis. For example, one study examining the relationship between exercise deprivation (not allowing people to exercise) and depression. reported, “We did not have a directional hypothesis, and when participants asked what we expected to find in this study, we replied: ‘We really are not sure since the results of earlier work on exercise deprivation are mixed’” (Mondin et al., 1996, p. 1200).

Table 1.2 Research Hypotheses and Their Independent and Dependent Variables

Research Hypothesis	Independent Variable	Dependent Variable
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher	Type of appeal	Intention to invest in cryptos

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Independent Variable	Dependent Variable
intention to invest in cryptos.		
Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.	Exposure to pro-cannabis messages	Intention to use cannabis
Mothers have more accurate knowledge of their children's online activities than do fathers.	Parent	Knowledge of children's online activities
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	Gender identity	Amount of victimization

Collecting Data

Once a research hypothesis has been formulated, researchers are ready to proceed to the second stage in the research process: collecting data. This step is seen as being composed of three tasks:

drawing a sample from a population,

determining how the variables will be measured, and

selecting a method by which to collect the data.

Table 1.3 Directional and Nondirectional Research Hypotheses

Directional Research Hypothesis	Nondirectional Research Hypothesis
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *higher* intention to invest in cryptos.	Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *different* intention to invest in cryptos.
Exposure to pro-cannabis messages on social media is related to *increased* intentions to use cannabis.	Exposure to pro-cannabis messages on social media is *related to* intentions to use cannabis.
Mothers have *more* accurate knowledge of their children's online activities than do fathers.	Mothers have *different* knowledge of their children's online activities than do fathers.
Nonbinary college students experience *higher* levels of victimization than do cis men or cis women.	Nonbinary college students experience *different* levels of victimization than do cis men or cis women.

Drawing a Sample From a Population

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The first step in collecting data is to identify the group of participants to which the research hypothesis applies. The group to which the results of a study may be applied or generalized is called a population. A **population** is the total number of possible units or elements that could potentially be included in a study. Although people often think the word "population" refers to everyone in the world, researchers typically define their populations more narrowly. For example, the population for the first research hypothesis in Table 1.1 could be "consumers," "U.S. consumers," "Georgia consumers," or "University of Georgia consumers." Researchers typically try to define their populations as broadly as possibly (e.g., "consumers" rather than "University of Georgia consumers") to maximize the applications or implications of their research.

Learning Check 1

Reviewing What You've Learned So Far



1. Review questions
    a. What are the five steps involved in the research process?
    b. Why is it useful to review and evaluate theories and research before conducting a study?
    c. What are the two main characteristics of research hypotheses?
    d. What is the difference between an independent variable and a dependent variable?
2. The following are several research hypotheses from published studies. For each research hypothesis, identify the independent and dependent variable.
    a. "College students will rate instructors who dress formally (i.e., business suit and tie) as having more expertise than instructors who dress casually (i.e., slacks and shirt)" (Sebastian & Bristow, 2008, p. 197).
    b. "It was expected tha. . . . greater amounts of television viewin. . . . would predict greate. . . . posttraumatic stress symptoms" (McLeish & Del Ben, 2008, p. 417).
    c. "Do borrowers who did not complete college have greater financial distress compared with . . . non-borrowers who did not complete college [or] borrowers who earned a degree?" (Jabbari et al., 2023, p. 6).
    d. "We expected that drivers would believe that texting while driving is more detrimental to their driving performance than browsing social media" (Hashash et al., 2019, p. 70).

It's extremely difficult to collect data from all members of a population. Imagine, for example, the time and money needed to collect information from every consumer in the United States! For this reason, researchers typically draw conclusions about populations based on information collected from a sample of the population. A **sample** is a subset or portion of a

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

population. Table 1.4 describes the samples used in the four studies introduced in Table 1.1 . As you can see from Table 1.4 , samples vary greatly in terms of their targeted population and size (the number of participants). The extent to which the results of a study can be generalized to the target population depends on the extent to which the sample is representative of the population.

Determining How Variables Will Be Measured: Levels of Measurement

The research hypotheses described in Table 1.1 involve variables such as type of appeal, intentions to use cannabis, and gender identity. To conduct a research study, the researcher must decide how to measure the variables stated in the research hypothesis. **Measurement** is the assignment of numbers or categories to objects or events according to rules.

For example, to measure the variable “height,” a researcher might use “number of inches from the ground in bare feet” as a form of measurement. To measure “success in college,” a student’s grade point average (GPA) may be obtained from school transcripts. “Self-esteem” might be measured by having people complete a questionnaire and be categorized as having either “low” or “high” self-esteem. As these examples demonstrate, different types of variables have different forms of measurement. Consequently, researchers have identified four distinct levels for measuring variables: nominal, ordinal, interval, and ratio.

The values of variables measured at the **nominal level of measurement** differ in category or type. The word *nominal* implies having to do with “names,” such that we use first and last names to distinguish between people. Gender identity is an example of a nominal variable, in that it consists of categories or types (such as cis male and cis female) rather than numeric values. In the first research hypothesis in Table 1.1 , the independent variable of “type of appeal” is measured at the nominal level with two categories: FOMO and non-FOMO.

Variables measured at the **ordinal level of measurement** have values that can be placed in an order relative to the other values. Rankings (such as finishing first, second, or third in a race), size (small, medium, large, or extra-large), and ratings (below average, average, or above average) are examples of ordinal variables. The values of ordinal variables indicate that one value represents relatively more or less of the variable than other values, but they don’t specify the precise difference between values. For example, although you can say that a “medium” drink is smaller than a “large” drink which is smaller than an “extra-large” drink, you can’t specify the exact difference in size between the three drinks.

Table 1.4 Research Hypotheses and Study Samples

Research Hypothesis	Study Sample
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.	“We recruited 150 participants from Amazon MTurk” (Friederich et at., 2024, p. 105).
Exposure to pro-cannabis messages on social media is	“Teens (N = 350) . . . needed to be between 13 and 17 years old and from Washington state” (Willoughby et

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Study Sample
related to increased intentions to use cannabis.	al., 2024, p. 185).
Mothers have more accurate knowledge of their children's online activities than do fathers.	"triadic data including reports from children aged 13 to 18, mothers and fathers . . . recruited in Flanders, the northern, Dutch speaking region of Belgium . . . a total of 357 valid triads was achieved" (Symons et al., 2017, p. 405).
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	"undergraduate college students aged 18–30 who were enrolled at 4-year institutions. This yielded a sample of 46,510 students" (Marx et al., 2024, p. 334).

The values of variables measured at the **interval level of measurement** are numbers equally spaced along a continuum. One example of an interval variable is the Fahrenheit scale of temperature. Here, a difference of 5 degrees has the same meaning anywhere along the scale; for example, the difference between 45°F and 50°F is the same as the difference between 65°F and 70°F. Many variables studied in the behavioral sciences (e.g., personality characteristics or attitudes) are considered to be measured at the interval level of measurement.

Variables at the **ratio level of measurement** are identical to interval variables, with one exception: Ratio scales possess what is known as a true zero point, for which the value of zero (0) represents the complete absence of the variable. Variables that describe a physical dimension (such as height, weight, distance, and time duration) typically have a true zero point. In the last research hypothesis in Table 1.1 , the variable "number of times victimized in last 12 months" is a ratio variable because it has a true zero point, where zero would indicate the complete absence of victimizations.

One advantage of ratio measurement versus interval is that ratio variables allow for a greater number of comparisons among values. Consider, for example, the value 6 for the ratio variable "inches." Not only is the difference between 4 and 6 inches the same as the difference between 6 and 8 inches (thereby involving addition and subtraction, which also can be used with interval variables), 6 inches is twice as long as 3 inches and half as long as 12 inches (involving multiplication and division). Multiplication and division comparisons cannot be made with interval variables. You cannot, for example, say that a temperature of 90°F is three times as much temperature as 30°F.

The boldfaced and italicized text in Table 1.5 illustrates how the four studies described in the earlier tables measured their independent and dependent variables. As you can see, the variables in these studies have been measured in a variety of ways at different levels of measurement.

Table 1.5 Research Hypotheses and Measurement of Variables

Research Hypothesis	Independent Variable	Dependent Variable

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Independent Variable	Dependent Variable
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.	Type of appeal **FOMO, non-FOMO (nominal)**	Intention to invest in cryptos **Likelihood of investing** **1-7 scale (1 = Highly unlikely to 7 = Highly likely) (ordinal)**
Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.	Exposure to pro-cannabis messages on social media **How often saw pro-cannabis messages (ratio)**	Intentions to use cannabis **Likelihood of using cannabis in the future (ratio)**
Mothers have more accurate knowledge of their children's online activities than do fathers.	Type of parent **Mother, Father (nominal)**	Accuracy of online activities **Accurate, somewhat accurate, inaccurate (nominal)**
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	Gender identity **Nonbinary, cis man, cis woman** **(nominal)**	Number of times victimized in last 12 months **(ratio)**

What difference does the level of measurement make?

A variable's level of measurement has important implications for researchers in that it influences how research hypotheses are stated and how data are analyzed. For example, imagine you study the variable "success in college." You could measure this variable at the *nominal* level of measurement by classifying students into one of two groups: successful or unsuccessful. Researchers typically use nominal variables when they are interested in studying differences between groups, such as, "Do successful and unsuccessful students differ in their study habits?"

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

You could instead measure success in college as an *ordinal* variable by ranking students from top to bottom, which enables a researcher to study relative differences with less concern for the precise size of these differences. For example, using ranked data, you could ask, “How similar are younger and older faculty members’ rankings of their students?”

To study success in college at the *interval* level, you could rate each student on a scale from 1 (low) to 100 (high). You could then use these ratings to address a research question such as, “Is there a relationship between faculty ratings of their students and students’ ratings of themselves?”

Finally, to measure success in college using the *ratio* level of measurement, you could use students’ salary after graduation (in dollars) as an indicator of success. Salary is a ratio variable because it contains a true zero point. In practice, ratio variables can be used to ask many of the same questions as those involving interval variables. For example, you could ask, “Is there a relationship between faculty members’ ratings of students and students’ salaries after graduation?”

As you see, how researchers measure their variables influences how they state their research questions and hypotheses. The level of measurement has another important implication: It helps determine the statistics researchers use to analyze their data. Certain statistics can be applied to variables measured at the interval or ratio levels of measurement, whereas others are appropriate for nominal or ordinal variables. As we move from chapter to chapter of this book, the level of measurement of the variables being analyzed will be noted.

Selecting a Method to Collect the Data: Experimental and Nonexperimental Research Methods

In addition to drawing a sample from a population and determining how the variables will be measured, the third step in collecting data is to determine the type of research method to use to collect the data on these variables. Research methods may be classified into two main types:

experimental research methods, and

nonexperimental research methods.

Experimental research methods.

Experimental research methods are methods designed to test causal relationships between variables—more specifically, whether changes in independent variables produce or cause changes in dependent variables. To make inferences about cause-effect relationships, researchers conducting an experiment must first eliminate other possible causes or explanations for changes in the dependent variable besides the independent variable. If it can be claimed that a variable other than the independent variable created the changes in the dependent variable, the research results are said to be confounded. A **confounding variable** is a variable related to an independent variable that provides an alternative explanation for the relationship between the independent and dependent variables.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

To illustrate the concept of confounding variables, consider the following question: Is there a relationship between a mother's ethnicity and the birth weight of her baby? Although research has shown differences in the birth weights of babies of different ethnicities, one study noted that some of this research did not eliminate the possibility that these differences may be partly due to differences between ethnic groups on factors associated with low birth weights, such as the mother's average age at the time of pregnancy and behaviors such as smoking and drinking (Ma, 2008). These other factors are considered confounding variables in that they provide alternative explanations for the relationship between ethnicity and babies' birth weights. Confounding variables make it difficult for researchers to draw firm conclusions regarding the relationship between the independent and dependent variable.

Learning Check 2

Reviewing What You've Learned So Far



1. Review questions
    a. What is the difference between a population and a sample?
    b. Within the research process, what is the relationship between populations and samples?
    c. What are the four levels of measurement? How do they differ?
2. For each of the following variables, name the scale of measurement (nominal, ordinal, interval, or ratio).
    a. Type of school attended (public, private)
    b. Probability of graduating college in 4 years (0% to 100%)
    c. Rating of television (good, better, best)
    d. Number of computers in home

Researchers minimize the influence of confounding variables in two main ways. First, researchers exercise experimental control by making the research setting (i.e., characteristics of the research participants, the instruments or measures administered, the methods used to collect data) the same for all participants. In the birth weight study, for example, the research design might control for the effect of parental smoking on birth weight by only including nonsmokers in the study sample, thereby excluding those who smoke.

Another way researchers exert control over the research setting is by including a condition known as a **control group** , a group of participants in an experiment not exposed to the independent variable. For example, if you conducted an experiment examining the effects of caffeine on one's health, one group (the experimental group) might be instructed to drink coffee while a second group (the control group) drinks decaffeinated coffee. By comparing the results of the two groups, researchers can then assess the impact of caffeinated coffee on the health of those who drink it.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Because researchers cannot possibly identify and control for all potential confounding variables, a second strategy used to minimize the influence of confounding variables is **random assignment** , which is assigning participants to each category of an independent variable in such a way that each participant has an equal chance of being assigned to each category. For example, a researcher could randomly assign a participant to the experimental or control condition by flipping a coin: heads for experimental, tails for control. The purpose of random assignment is to equalize or neutralize the effects of confounding variables by distributing them equally over the levels of the independent variable.

Nonexperimental (correlational) research methods.

Experimental research designs are one of the best tools researchers have for making causal inferences. However, the ability to make causal inferences requires a great deal of control that may not always be possible or desirable. For this reason, researchers often employ **nonexperimental research methods** , which are research methods designed to measure relationships between variables without having the ability to infer cause-effect relationships. These research methods are sometimes referred to as *correlational* research methods. It is important to understand that correlational research methods are different from correlational statistics we'll introduce in Chapter 13. Correlational research methods pertain to how data are *collected* ; correlational statistics pertain to how data are *analyzed* . Some of the most common types of nonexperimental research designs include quasi-experiments, survey research, observational research, and archival research.

Quasi-experimental research compares naturally formed or preexisting groups rather than randomly assigning participants to conditions. One example of this type of research was a study examining the effects of the COVID-19 pandemic on three groups of students: those living on campus, those living off campus but within commuting distance to school, and those living far away in another city, state, or country (Bisconer & McGill, 2023). This study is a quasi-experiment because the students were not randomly assigned to one of the three living conditions before collecting data.

Survey research methods obtain information directly from a group of people regarding their opinions, beliefs, or behavior. The goal of survey research, which can include using questionnaires and interviews, is to obtain information from a sample to represent or estimate the views of a larger population. One example of survey research examined college students' mental health by having them fill out a survey in which they reported their spirituality (the desire for truth, meaning, and significance of one's life) and their self-acceptance (knowledge and acceptance of one's self and limitations) (Nonis et al., 2024). Survey research is nonexperimental in that the variables being measured are not manipulated by the researcher.

Observational research is the systematic and objective observation of naturally occurring behavior or events. Observational research is often used to study phenomena that the researcher cannot or should not deliberately manipulate. For example, researchers studying interpersonal conflict among children would never force children to fight each other. One study resolved this issue by videotaping groups of adolescents discussing topics such as competition and peer pressure and then counting the number of times the adolescents teased or mocked each other (Connolly et al., 2015).

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Rather than observing or measuring behavior directly, **archival research** is the use of archives (records or documents of the activities of individuals, groups, or organizations) to examine research questions or hypotheses. One example of an archival research study studied the question of whether men are perceived to be more creative than women (Proudfoot et al., 2015). In this study, the 100 most viewed talks on TED.com were identified, and the researchers compared viewers' posted ratings of the talks given by men versus women.

Labeling variables in nonexperimental research.

One issue that often confuses students when learning about nonexperimental research is the practice of referring to the variables as the "independent variable" and the "dependent variable." For example, in the study described previously that involved biological sex and TED.com ratings, biological sex may be called the "independent variable." However, because biological sex isn't manipulated by the researcher, it would be inappropriate to say that biological sex "causes" differences in TED.com ratings. To avoid confusion, researchers sometimes refer to the independent variable and the dependent variable in a nonexperimental study as the **predictor variable** and **criterion variable**, respectively.

Analyzing the Data

Once data collection has been completed, the next step in the research process involves data analysis. This part of the research process addresses the primary topic of this book: statistics. Because this is the first chapter of this book, we won't describe analyzing data in detail. Instead, we introduce the notion that there are two main purposes of analyzing data: (1) to organize, summarize, and describe data that has been collected and (2) to test and draw conclusions about ideas and hypotheses. These two purposes are met by calculating two main types of statistics: descriptive statistics and inferential statistics.

Calculating Descriptive Statistics

Descriptive statistics are statistics used to summarize and describe a set of data for a variable. For example, you've probably heard of statistics used to describe or summarize aspects of our society, such as crime statistics and unemployment statistics. Using a research-related example, Caitlin Abar, a researcher at Brown University, conducted a study examining increases in students' alcohol-related behaviors after entering college (Abar, 2012). To measure students' level of alcohol use, she created a variable called "typical weekend drinking," which "was measured as the sum of drinks consumed on a typical Friday and Saturday within the past 30 days" (p. 22). One way to summarize students' responses to the typical weekend drinking variable (0, 1, 2, etc.) would be to calculate the mean, which is the mathematical average of a set of scores. The mean, described in Chapter 3, is an example of a descriptive statistic. The first part of this book describes a variety of descriptive statistics used by researchers to summarize data they've collected.

Calculating Inferential Statistics

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Besides summarizing and describing data, a second purpose of statistics is to test hypotheses. **Inferential statistics** are statistical procedures used to test hypotheses and draw conclusions from data collected during research studies. Using inferential statistics, researchers make inferences about what's believed to exist in populations based on data gathered from samples of the population.

As an example of an inferential statistic, let's return to the example of students' alcohol-related behaviors (Abar, 2012). This study was interested in seeing whether these behaviors were related to aspects of these students' relationships with their parents. One aspect was "alcohol communications," defined as "how often they discussed alcohol related topics with their parents" (p. 22). To test a hypothesis regarding the relationship between students' alcohol use and their alcohol communications, we could use an inferential statistic known as the Pearson correlation coefficient, which will be discussed in Chapter 13. The chapters in the last half of this book discuss a broad variety of inferential statistics, along with detailed instructions for analyzing and drawing conclusions from statistical data and presenting research findings.

Drawing a Conclusion Regarding the Research Hypothesis

Once statistical analyses have been completed, the next step is to interpret the results of the analyses as they relate to the study's research hypothesis. More specifically, do the findings of the analyses support or not support the research hypothesis? The word *support* in the previous sentence is very important. Students and researchers are sometimes tempted to say their findings "prove" their hypotheses are either "true" or "false." However, as we'll discuss in Chapter 5, because researchers typically don't collect data from the entire population, they can't know with 100% certainty whether their hypothesis has been proven to be true or false. A *huge* part of this book is learning how to draw appropriate conclusions from statistical analyses.

Communicating the Findings of the Study

Conducting research requires a variety of different skills: conceptual skills to develop research hypotheses, methodological skills to collect data, and mathematical skills to analyze these data. Another integral part of the research process is the communication skills needed to inform others about a study. Researchers must be not only skilled scientists but also effective writers.

Learning Check 3

Reviewing What You've Learned So Far



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1. Review questions
    a. What are the main differences between experimental and nonexperimental research methods?
    b. What are confounding variables? What do researchers do to minimize the effects of confounding variables?
    c. What are the main types of nonexperimental research methods?
    d. What are the relative strengths and weaknesses of experimental and nonexperimental research?
    e. What are the main purposes of descriptive and inferential statistics?

For many years, the American Psychological Association (APA), the professional association for psychologists, has provided researchers guidance on how to communicate the results of their research. In 1929, the APA published a seven-page article entitled, “Instructions in Regard to Preparation of Manuscript.” In contrast, the seventh edition of the *Publication Manual of the American Psychological Association* , published in 2020, consists of 428 pages! This increase in length highlights the challenge of communicating research effectively and appropriately.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1.4 Plan of the Book

The primary purpose of this chapter was to introduce statistics and place it within the larger research process. The remainder of this book discusses different statistics used by researchers to examine various questions of interest. Although it's critical for you to learn to correctly calculate statistics, it is equally important to understand the role of statistics within the research process and appreciate issues related to the use (and sometimes misuse) of statistics.

The first half of the book (Chapters 2 through 6) introduces conceptual and mathematical issues that are the foundation of statistical analyses. Chapters 2, 3, and 4 discuss how data may be examined, described, summarized, and presented in numeric, visual, and graphic form. Descriptive statistics are introduced and discussed in these chapters. Chapters 5 and 6 introduce characteristics of the inferential statistics used to test research hypotheses. The chapters in the second half of the book cover a number of different inferential statistics used with different types of variables.

In most cases, the chapters in this book share a uniform format and structure and are centered on the research process, using examples of actual published research studies that address questions with which you may be familiar. For each example, we'll clearly state the research hypothesis, briefly describe how the data in the study were collected, illustrate the calculation and interpretation of descriptive and inferential statistics, discuss the extent to which the statistical analyses support the research hypothesis, and show how to communicate one's research to a broader audience.

1.5 Looking Ahead

We began this chapter by defining statistics and providing reasons why learning about them may be of value to you. Next, we placed statistics within the stages of the larger research process, a process that will be the foundation of this textbook. As we've explained, the research process is centered on a research hypothesis, which is a predicted relationship between variables. Once a hypothesis has been stated, the next step is to collect data to be analyzed. However, before conducting statistical analyses on a set of data, there is preliminary work that must be completed. The next chapter will focus on methods used by researchers to examine data.

1.6 Summary

Statistics may be defined as a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data. As such, statistics not only is concerned about how data are analyzed but also recognizes the importance of understanding how data are collected and how the results of analyses are interpreted and communicated. There are a variety of reasons why students should learn statistics: Statistics are encountered in a wide variety of daily activities, students may be asked or required to read and interpret the results of statistical analyses in their courses, students may use statistics in conducting their own research, and statistics may help one's career.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Researchers conduct research using the **scientific method** of inquiry, a method of investigation that uses the objective and systematic collection and analysis of empirical data to test theories and hypotheses. The research process used within the scientific method of inquiry consists of five main steps: developing a research hypothesis to be tested, collecting data, analyzing the data, drawing a conclusion regarding the research hypothesis, and communicating the findings to the study.

A **research hypothesis** is a statement regarding an expected or predicted relationship between variables. Developing a research hypothesis involves identifying a question or issue to be examined, reviewing and evaluating relevant theories and research, and stating the research hypothesis to be tested in the study. A **theory** is a set of propositions used to describe or explain a phenomenon. A research hypothesis contains **variables** , which are properties or characteristics of some object, event, or person that can take on different values. More specifically, a research hypothesis states the nature and direction of a proposed relationship between an **independent variable** (a variable manipulated or controlled by the researcher) and a **dependent variable** (a variable measured by the researcher that is expected to change or vary as a function of the independent variable).

Collecting data involves drawing a sample from a population, determining how the variables will be measured, and selecting a method to collect the data. A **population** is the total number of possible units or elements that could potentially be included in a study; a **sample** is a subset or portion of a population. Variables can be measured at one of four levels of measurement: **nominal** (values differing in category or type), **ordinal** (values placed in an order relative to the other values), **interval** (values equally spaced along a numeric continuum), or **ratio** (values equally spaced along a numeric continuum with a true zero point).

There are two main types of methods used to collect data: **experimental research methods** , designed to test whether changes in independent variables produce or cause changes in dependent variables, and **nonexperimental (correlational) research methods** , designed to examine the relationship between variables without the ability to infer cause-effect relationships.

In experimental research, researchers are concerned about possible **confounding variables** , which are variables related to independent variables that provide an alternative explanation for the relationship between independent and dependent variables. To minimize the impact of confounding variables, a research study may exercise experimental control over various aspects of the situation, including the use of a **control group** , which is a group of participants not exposed to the independent variable the research is designed to test, or **random assignment** , which involves assigning participants to each category of an independent variable in such a way that each participant has an equal chance of being assigned to each category.

Examples of nonexperimental research methods are **survey research** , in which information is directly obtained from a group of people regarding their opinions, beliefs, or behavior; **observational research** , which involves the systematic and objective observation of naturally occurring events; and **archival research** , which uses archives (records or documents of the activities of individuals, groups, or organizations) to examine research questions or hypotheses. In nonexperimental research, the independent variable and

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

dependent variable may be referred to as the **predictor variable** and **criterion variable**, respectively.

Once data have been collected, the next step is to analyze them using two main types of statistics: **descriptive statistics**, which summarize and describe a set of data, and **inferential statistics**, which are statistical techniques used to test hypotheses and draw conclusions from data collected during research studies.

Once the statistical analyses have been completed, the results of the analyses are interpreted regarding whether they support or not support the study's research hypothesis. The final step in the research process is to communicate the study to a broader audience.

1.7 Important Terms

archival research

confounding variable

control group

criterion variable

dependent variable

descriptive statistics

experimental research methods

independent variable

inferential statistics

interval level of measurement

level of measurement (nominal, ordinal, interval, ratio)

measurement

nonexperimental research methods

observational research

population

predictor variable

quasi-experimental research

random assignment

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

research hypothesis

sample

scientific method

statistics

survey research

theory

variable

1.8 Exercises

1. Identify the independent and dependent variable for each of the following research questions and hypotheses from published articles.
    a. “We hypothesize that finance majors will embody narcissistic traits more than other business majors” (Wood et al., 2021, p. 924).
    b. “Men are expected to employ physical characteristics (particularly those that are directly related to sex) more often than women in selecting dating candidates” (Hetsroni, 2000, p. 91).
    c. “It is hypothesized that perceptions of emotional pain and support needs will be higher for the community of the Black victim than those of the officers’” (Kahn et al., 2024).
    d. “Gen Z students are more likely to believe that ‘Group projects bring out the worst in people’ than Millennials” (Schlee et al., 2020, p. 145).
    e. “We hypothesized to find . . . higher stress and anxiety post COVID in the Portuguese group in contrast with the British ” (Penalba-Sanchez et al., 2022, p. 322).
    f. “We hypothesized that consumers would think that a ‘sale price’ presentation would generate a greater monetary savings than an ‘everyday low price’ presentation” (Tom & Ruiz, 1997, p. 403).
2. The following are a number of research questions and hypotheses from actual published articles. For each hypothesis, identify the independent and dependent variable.
    a. “The use of color in a Yellow Pages advertisement will increase the perception of quality of the products for a particular business when compared with noncolor advertisements” (Lohse & Rosen, 2001, p. 75).
    b. “We hypothesized that parents who use more frequent corporal and verbal punishmen. . . . will report more problem behaviors in their children” (Brenner & Fox, 1998, p. 252).
    c. “It was hypothesized that adolescents with anorexia nervosa woul. . . . be more respectful when compared with peers with bulimia nervosa” (Pryor & Wiederman, 1998, p. 292).
    d. “The purpose of the present research was to assess brand name recognition as a function of humor in advertisements. . . . It was predicted that participants would recognize product brand names that had been presented with humorous

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

advertisements more often than brand names presented with nonhumorous advertisements" (Berg & Lippman, 2001, p. 197).

e. "The purpose of this study was to explore the relation between time spent in daycare and the quality of exploratory behaviors in 9-month-old infant. . . . it was hypothesized tha. . . . infants who spent greater amounts of time in center-based care would demonstrate more advanced exploratory behaviors than infants who did not spend as much time in center-based care" (Schuetze et al., 1999, p. 269).

f. "It was hypothesized that students would score higher on test items in which the narrative contains topics and elements that resonate with their daily experiences versus test items comprised of material that is unfamiliar" (Erdodi, 2012, p. 172).

3. Identify the independent and dependent variable for each of the following research questions and hypotheses from published articles.
    a. "It was hypothesized that the number of selfies posted . . . would be positively correlated with dimensions of narcissism" (Barry et al., 2017, p. 51).
    b. "We hypothesized that preschoolers' self-reported preferences for aggressive media . . . would be positively related to parent reports of children's aggressive behaviors" (Jamnik & DiLalla, 2018, p. 159).
    c. "We hypothesized that helicopter parenting would lead to higher levels of depression" (Darlow et al., 2017, p. 3).
    d. "We predicted that weight-related shame would be associated with greater binge eating" (Craven & Fekete, 2019, p. 45).
    e. "We hypothesize that men with strong gender identities will be more sensitive to manipulated threats to their masculinity compared with men with gender identities that are less central to their identity" (Harrison & Michelson, 2019, p. 65).
    f. "After exposure to fake news story with user comments identifying the news as fake, people are less likely to share the fake news than after exposure to the fake news story with user comments supportive of the content" (Colliander, 2019, p. 204).
4. Name the scale of measurement (nominal, ordinal, interval, ratio) for each of the following variables:
    a. The amount of time needed to react to a sound
    b. Gender
    c. Score on the Scholastic Aptitude Test (SAT)
    d. Political orientation (not at all conservative, conservative, very conservative)
    e. Political affiliation (Democrat, Republican, Independent)
5. Name the scale of measurement (nominal, ordinal, interval, ratio) for each of the following variables:
    a. One's age (in years)
    b. Size of soft drink (small, medium, large, extra large)
    c. Voting behavior (in favor vs. against)
    d. IQ score
    e. Parent (mother vs. father)
6. A faculty member wishes to assess the relationship between students' scores on the SAT and their performance in college.
    a. What is a possible research hypothesis in this situation?
    b. What are the independent and dependent variables?
    c. How could you measure the variable "performance in college" at each of the four levels of measurement?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

7. A researcher hypothesizes that drivers who use cellular phones will get into a greater number of traffic accidents than drivers who do not use these phones. For this example,
    a. What are the independent and dependent variables?
    b. How could you measure the dependent variable?
    c. How could you conduct this study using an experimental research method? How could you conduct this study using a nonexperimental method?

Answers to Learning Checks

Learning Check 1

1.

    a. IV: Instructor's dress; DV: Expertise

    b. IV: Amount of television viewing; DV: Posttraumatic stress symptoms

    c. IV: Student group; DV: Financial distress

    d. IV: Activity while driving; DV: Driving performance

Learning Check 2

2.

    a. Nominal

    b. Ratio

    c. Ordinal

    d. Ratio

Answers to Odd-Numbered Exercises

1.

    a. IV: Type of business major; DV: Narcissistic traits

    b. IV: Biological sex; DV: Employment of physical characteristics

    c. IV: Community; DV: Perceptions of emotional pain and support needs

    d. IV: Generation; DV: Beliefs about group projects

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

e. IV: Country; DV: Stress and anxiety

f. IV: Price presentation; DV: Perception of savings

3.

a. IV: Number of selfies posted; DV: Narcissism

b. IV: Preference for aggressive media; DV: Parent reports of children's aggressive behaviors

c. IV: Helicopter parenting; DV: Depression

d. IV: Weight-related shame; DV: Binge eating

e. IV: Strength of gender identity; DV: Sensitivity to threat to masculinity

f. IV: Type of user comments to news story; DV: Sharing of news story

5.

a. Ratio

b. Ordinal

c. Nominal

d. Interval

e. Nominal

7.

a. IV: Cellular phone use; DV: Number of traffic accidents

b. Example: Determine the number of traffic accidents each person has experienced in the past year.

c. Example of experimental: Use a driving simulation with the "driver" talking on a phone and measure the number of potential driving mistakes or accidents.

d. Example of nonexperimental: Take a survey of the number of accidents people have been in and whether or not they were using a phone during the accident.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CHAPTER 2 EXAMINING DATA: TABLES AND FIGURES

Learning Objectives

2.1 Describe the main purpose and the research hypothesis of the “lottery” study.

2.2 Describe different reasons for examining data.

2.3 Describe the steps used to create a frequency distribution table.

2.4 Describe the steps used to create a grouped frequency distribution table.

2.5 Explain the differences between bar charts, pie charts, histograms, and frequency polygons.

2.6 Describe the three aspects and types of distributions of data.

Chapter 1 described a process by which research may be conducted, which begins by stating a research hypothesis. The next step involves collecting data that will ultimately be used in statistical analyses designed to test the research hypothesis. However, researchers don’t conduct analyses until they’ve first examined the data to make sure appropriate conclusions can be drawn from the analyses. This chapter describes *why* researchers examine data and *how* data may be examined using tables, charts, and graphs. As with many of the chapters in this book, this discussion will be guided by the findings from a published research study.

2.1 An Example From the Research: Winning the Lottery

A friend of yours flips a coin and asks you to guess whether it’s “heads” or “tails.” You guess “tails,” but the answer is “heads.” Your friend flips the coin a second time, and again you predict “tails.” But again the answer is “heads.” On the third coin flip, you guess “tails” once more—but, once again, you’re wrong. As your friend prepares to flip the coin a fourth time, you tell yourself there’s little chance the coin could be “heads” again. But the reality is that the fourth coin flip is equally likely to be “heads” or “tails” because the outcome of the first three coin flips has absolutely no impact on the fourth flip. In other words, the outcome of a coin flip is a random event that can’t be controlled or predetermined. But do you really believe this?

The impact of random events can take on greater significance than the outcome of a coin flip. For example, government-run lotteries provide the opportunity to win huge amounts of money through the random selection of numbers. In 2001, Dr. Karen Hardoon, a researcher at McGill University in Canada, studied the thought processes people use when purchasing lottery tickets (Hardoon et al., 2001). As previous research had found gamblers believe they can control random events such as the rolling of dice or the spinning of a roulette wheel, Dr. Hardoon and her colleagues were interested in

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

determining “whether individuals with gambling problems perceive the purchase of lottery tickets in a similar manner as non-problem gamblers” (p. 752). On the basis of earlier findings, Dr. Hardoon hypothesized that problem gamblers are less likely to believe lottery outcomes are random than are nonproblem gamblers.

To test their research hypothesis, Dr. Hardoon and her associates located a group of gamblers and asked whether they had ever experienced problems as a result of their gambling. Each participant was classified into one of two groups: problem gamblers and nonproblem gamblers.

The participants in the study, which we’ll refer to as the lottery study, were shown four lottery tickets and asked, “If you were to buy a ticket to play in the lottery, which one would you select?” Each ticket contained six numbers selected from among the numbers 1 through 49; the four tickets were given a particular label. The *sequence* ticket had numbers in consecutive order (30-31-32-33-34-35); numbers on the *pattern* ticket increased by 5s (5-10-15-20-25-30); the *nonequilibrated* numbers (35-37-40-43-44-49) were clustered at the upper end of the 49 numbers, and the *random* ticket (7-8-23-34-36-42) shared none of the characteristics of the other three tickets. We’ll call the variable in this study “lottery ticket,” a variable with four categories: sequence, pattern, nonequilibrated, and random.

The researchers recorded which of the four tickets each participant chose. Table 2.1 lists the data for the lottery ticket variable for the nonproblem gamblers (the data for the problem gamblers are provided in the exercises at the end of this chapter). Looking at this table, we see that the nonproblem gamblers differed in their lottery ticket choices; however, it’s difficult to describe these differences given the unorganized nature of the collected data. Consequently, before conducting statistical analyses on their data, researchers typically organize and examine them. In the following sections, we will discuss why and how data are organized and examined.

2.2 Why Examine Data?

There are a variety of reasons why researchers examine data they’ve collected before analyzing the data. These reasons include the following:

To gain an initial sense of the data

To identify outliers

To evaluate research methodology

To determine whether data meet statistical criteria and assumptions

Each of these reasons is discussed in the following sections.

Table 2.1 The Lottery Ticket Choices of 22 Nonproblem Gamblers

Participant	Lottery Ticket
1	**Random**
2	**Random**
3	**Nonequilibrated**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Participant	Lottery Ticket
4	**Pattern**
5	**Random**
6	**Sequence**
7	**Random**
8	**Random**
9	**Pattern**
10	**Pattern**
11	**Random**
12	**Pattern**
13	**Random**
14	**Nonequilibrated**
15	**Random**
16	**Pattern**
17	**Nonequilibrated**
18	**Random**
19	**Random**
20	**Pattern**
21	**Random**
22	**Random**

Gaining an Initial Sense of the Data

Because researchers spend a great amount of time and energy designing their research studies, they're eager to analyze the data they've collected. Students taking courses in statistics, on the other hand, face a different challenge: They're given sets of data with which they have no prior experience. Although both researchers and students may be motivated to analyze their data as quickly as possible, it's important to first examine the data in order to gain an initial sense of them.

As an example, imagine you conduct a study regarding the fuel efficiency of automobiles and collect data on the miles per gallon (MPG) of different types of cars. You might expect the scores in your data to range from 10 to 50 MPG, with the majority between 25 and 35 MPG. Examining data helps researchers gain an initial sense of whether the data are in line with their expectations. Examining data also provides a way of confirming the results of later statistical analyses. We've found that students' errors in their calculations could have been avoided had they first looked at the set of data. For example, examining data will help you avoid making statements such as, "The average MPG in my sample was 127 MPG."

Identifying Outliers

Researchers also examine their data to identify **outliers** , defined as rare, extreme scores that lie outside of the range of the majority of scores in a set of data. Imagine, for example, that a group of 15 students is asked to take a 20-item test of mathematical calculations; each student's score is the number of test items answered correctly. In scoring the tests, a researcher finds that 14 students

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

correctly answered between 8 and 17 questions. However, one student provided the wrong answer to every question for a score of 0. In this case, the score of 0 would be considered an outlier.

If not identified, outliers may distort conclusions researchers draw about their sample and their data, particularly when the sample is relatively small. In the previous example, the single score of 0 might lead an observer to conclude the class as a whole performed more poorly on the test than they actually did.

Evaluating Research Methodology

Examining data is also useful in assessing the effectiveness of the methods used to collect the data. For the 20-item math test, for example, it might be reasonable to expect the majority of scores to fall in the middle of the 0 to 20 range. But what if every student answered at least 17 of the 20 questions correctly? You might conclude the test was too easy and that more difficult questions are needed to accurately assess students' math skills. By comparing the collected data for a variable with the range of possible values, researchers are able to evaluate and modify their measurement tools.

Determining Whether Data Meet Statistical Criteria and Assumptions

Researchers also examine their data to see whether they meet certain statistical criteria and assumptions. Many of the statistics in this book are based on assumptions regarding the shape and nature of a set of data that has been collected. For example, these procedures assume that in most data sets, the majority of scores will fall in the middle of the range of possible values, with a small amount of scores at the highest and lowest possible values (you may have heard this referred to as "the bell curve"). In the example of the 20-item math test, it might be assumed that most scores will be between 7 and 13, with smaller number of scores less than 7 or greater than 13. When assumptions such as this are not met, it's possible to draw inappropriate conclusions from statistical analyses that are conducted.

2.3 Examining Data Using Tables

One part of examining data that have been collected for a variable involves answering a series of questions:

> What are the values of the variable?
>
> How many participants in the sample have each value of the variable?
>
> What percentage of the sample has each value of the variable?

As it may be difficult to answer questions such as these just by looking at the data in Table 2.1 , researchers typically organize their data by creating a table known as a **frequency distribution table** , a table that summarizes the number and percentage of participants for the different values of a variable. This section discusses the steps involved in creating frequency distribution tables using these questions.

What Are the Values of the Variable?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The first step in organizing the data for a variable is to identify all of the possible values for the variable. Table 2.2 begins the construction of a frequency distribution table by identifying the four values for the lottery ticket variable: sequence, pattern, nonequilibrated, and random.

How Many Participants in the Sample Have Each Value of the Variable?

Once the values of the variable have been identified, the next step in organizing a set of data is to determine the **frequency** of each value, which is the number of participants in a sample corresponding to a value of a variable. These frequencies help indicate the nature and shape of the distribution of scores for a variable in a sample.

Table 2.2 Identifying the Values of the Lottery Ticket Variable

Number	Lottery Ticket
1	Sequence
2	Pattern
3	Nonequilibrated
4	Random

Table 2.3 Using the Tick Mark Method to Determine the Frequencies of the Lottery Ticket Variable

Lottery Ticket	Tick Marks	*f*
Sequence	|	1
Pattern	||||— |	6
Nonequilibrated	|||	3
Random	||||— ||||— ||	12
Total	||||— ||||— ||||—||||—||	22

One way to determine the frequencies for a variable is to sort the data into the different values and then count the number of participants having each value using a “tick mark method,” where a tick mark is placed next to a value of the variable each time that value appears in the sample. As shown in Table 2.3 , the tick marks are then counted to determine the frequency for each value. For example, the number of tick marks and therefore the frequency (f) for the sequence ticket is 1 because it was picked by one participant (Participant 6 in Table 2.1). Similarly, the tick marks for each of the other three types of tickets can be recorded and counted until all of the data in the sample are accounted for.

What Percentage of the Sample Has Each Value of the Variable?

Although it’s important to determine the frequency for each value of a variable, these frequencies by themselves are not always particularly informative. For example, how you interpret “There are 35 Democrats in my sample” is different when the sample consists of 40 voters rather than 400 voters.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Because the interpretation of frequencies depends on the size of the sample, it's useful to calculate the *percentage* of the sample having each value of a variable. These percentages can be calculated using the following formula:

$$\% = \frac{f}{total\,number\,of\,scores} * 100$$

where *f* is the frequency for a value of a variable.

Looking at Table 2.3 , 1 of the 22 nonproblem gamblers chose the sequence ticket. Therefore, the percentage of the sample choosing this ticket is

$$\begin{aligned} \% &= \frac{f}{total\,number\,of\,scores} * 100 \\ &= \frac{1}{22} * 100 = .05 * 100 \\ &= 5\% \end{aligned}$$

After calculating the percentage for each of the four tickets, Table 2.4 provides the frequency distribution table for the lottery ticket variable.

Once a frequency distribution table has been constructed for a variable, researchers can begin to draw conclusions about their sample. For the lottery ticket variable, the random ticket had the highest frequency and percentage (*f* = 12, 54%) compared to the pattern (*f* = 6, 27%), nonequilibrated (*f* = 3, 14%), and sequence (*f* = 1, 5%) tickets. As a result, the researchers in the lottery study made the following observation:

"The results of the present study indicated that, for the entire sample . . . the greatest percentage of tickets selected by the entire sampl. . . . were random tickets." (Hardoon et al., 2001, p. 760)

Table 2.4 Frequency Distribution Table for the Lottery Ticket Variable

Lottery Ticket	***f***	**%**
Sequence	1	**5%**
Pattern	6	**27%**
Nonequilibrated	3	**14%**
Random	12	**54%**
Total	22	**100%**

Learning Check 1

Reviewing What You've Learned So Far

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



1. Review questions
    a. What are some reasons for examining data before conducting statistical analyses?
    b. Why are outliers of concern to researchers?
    c. What questions do you address about a set of data for a variable in constructing frequency distribution tables?
    d. Why is it useful to calculate percentages in addition to frequencies for each value of a variable?
2. For each of the following situations, create a frequency distribution table.
    a. A college counselor asks a group of 116 seniors what their plans are after graduating from college; 71 say they are going to “work,” 34 are planning on going to “graduate school,” and 11 say they are “not sure” what their plans are.
    b. A high school instructor teaching a course finds that of the students she teaches, 69 are Freshmen, 18 are Sophomores, 12 are Juniors, and 9 are Seniors.
    c. A pollster stops 12 people and asks if they are either “in favor” or “against” a local proposition in Table 2.5 :

Table 2.5 Example of People’s Position

Person	Position
1	**In favor**
2	**In favor**
3	**Against**
4	**In favor**
5	**Against**
6	**Against**
7	**In favor**
8	**Against**
9	**In favor**
10	**Against**
11	**In favor**
12	**In favor**

    d. The company that makes M&Ms candy (www.mms.com) conducted a survey asking people which new color they would like to have added: purple, aqua, or pink. Table 2.6 are the votes of a hypothetical sample:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Table 2.6
Hypothetical Sample of New M&M Color

Person	Color
1	**Pink**
2	**Purple**
3	**Pink**
4	**Aqua**
5	**Purple**
6	**Purple**
7	**Purple**
8	**Aqua**
9	**Purple**
10	**Pink**
11	**Aqua**
12	**Pink**
13	**Pink**
14	**Purple**
15	**Aqua**
16	**Pink**
17	**Purple**
18	**Purple**
19	**Aqua**
20	**Purple**
21	**Purple**
22	**Pink**
23	**Pink**
24	**Purple**

2.4 Grouped Frequency Distribution Tables

The frequency distribution table in Table 2.4 was sufficient for presenting the frequencies for the lottery ticket variable. However, other situations may involve variables with a large number of possible values. Consider, for example, the variable grade point average (GPA). The values for GPA typically range from 0.00 to 4.00, with hundreds of possible values in between. Creating a frequency distribution table for GPA could result in an extremely large table, with many values having frequencies of zero.

To examine a variable with a large number of values, it is useful to construct a **grouped frequency distribution table** , a table that groups the values of a variable measured at the interval or ratio level

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

of measurement into a small number of intervals and then provides the frequency and percentage for each interval. To illustrate how grouped frequency distribution tables are constructed, Table 2.7 (a) provides a sample of 42 hypothetical GPAs. In Table 2.7 (b), each of these 42 GPAs has been placed into one of eight intervals that comprise a grouped frequency distribution table.

To illustrate how to read and interpret a grouped frequency distribution table, the lowest interval in Table 2.7 (b) consists of students with GPAs less than (<) 2.00. The frequency in this interval (*f* = 1) consists of one student: Student 29 (GPA = 1.96). The next interval consists of students with GPAs ranging from 2.00 to 2.29 and contains three students (4, 22, and 38). This grouping continues until all of the GPAs in the sample have been accounted for.

To create a grouped frequency distribution table, the real limits of each interval should be identified. The **real limits** of an interval are the values of the variable that are halfway between the top of one interval and the bottom of the next interval. For example, because of how values of GPA are rounded, the lowest value of GPA that would lead someone to be placed in the interval 2.30–2.59 is 2.295, which is halfway between the top of the 2.00–2.29 interval and the bottom of the 2.30–2.59 interval. The **real lower limit** is the smallest value of a variable that would be grouped into a particular interval. In this case, 2.295 represents the real lower limit for the 2.30–2.59 interval. On the other hand, the **real upper limit** is the largest value of a variable that would be grouped into a particular interval. For the 2.30–2.59 interval, the real upper limit is 2.594.

Once the grouped frequency table for the GPA example has been created (Table 2.7(b)), what preliminary conclusions might be drawn about these students? First, because the 2.90–3.19 interval has the greatest frequency (*f* = 10), you might conclude that the typical student in this sample has a GPA close to 3.00. Second, moving in both directions away from the 2.90–3.19 interval, the frequencies of the different intervals grow progressively smaller, with fewer and fewer students having either very high or very low GPAs. Table 2.8 provides a list of guidelines for creating grouped frequency distribution tables.

Cumulative Percentages

In addition to determining the percentage of the sample that has each individual value or interval for a variable, it is sometimes useful to combine these percentages. For the GPA example, you may be interested in knowing, “What percentage of the sample had GPAs less than 2.60?” To answer this question, the percentages in the “%” column of the frequency distribution table must be combined. This type of percentage is referred to as a **cumulative percentage** , defined as the percentage of a sample at or below a particular value of a variable.

Table 2.7a Grade Point Average (GPA), 42 Students: Individual GPAs

Student	GPA
1	**3.05**
2	**2.83**
3	**3.26**
4	**2.19**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Student	GPA
5	**3.52**
6	**3.34**
7	**3.02**
8	**2.97**
9	**3.71**
10	**2.50**
11	**2.77**
12	**3.10**
13	**3.06**
14	**3.83**
15	**2.39**
16	**3.16**
17	**3.92**
18	**2.37**
19	**3.65**
20	**3.70**
21	**3.00**
22	**2.26**
23	**2.65**
24	**3.40**
25	**2.87**
26	**2.92**
27	**2.89**
28	**3.41**
29	**1.96**
30	**3.77**
31	**3.20**
32	**2.86**
33	**2.94**
34	**2.71**
35	**3.43**
36	**3.95**
37	**3.32**
38	**2.28**
39	**3.08**
40	**3.25**
41	**2.40**
42	**3.22**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Table 2.7b Grade Point Average (GPA), 42 Students: Grouped Frequency Distribution Table

GPA	*f*	%
3.80+	3	7%
3.50−3.79	5	12%
3.20−3.49	9	21%
2.90−3.19	10	24%
2.60−2.89	7	17%
2.30−2.59	4	10%
2.00−2.29	3	7%
<2.00	1	2%
Total	42	100%

Table 2.8 Guidelines for Creating Grouped Frequency Distribution Tables

Guideline	Guideline Description
1. *Variables are grouped into approximately 10 intervals.*	The exact number of intervals depends on the data in each sample. For example, data that fall within a small range of values may be accurately represented with a relatively small number of intervals, whereas data extending across a wide range may require a larger number of intervals.
2. *The number of intervals should accurately represent the data.*	The intervals should represent the nature of the data as accurately as possible. For example, there should not be so large a number of intervals that many intervals have frequencies of zero (e.g., 3.00−3.05) or so few intervals that a large majority of the sample falls into only one or two intervals (e.g., 3.00−3.99).
3. *Intervals should be of equal size.*	To make the intervals comparable, they should be of equal size or width. For example, you would not want one interval of 2.01−2.50 (a width of .50) and another of 2.51−2.70 (a width of .20). One exception to this guideline is the lowest or highest interval (e.g., less than 2.00 or greater than 3.70), which is typically wider than the other intervals because it contains a relatively small frequency.
4. *Intervals should not overlap.*	The intervals should be distinct from each other, such that each score may be placed in only one interval. For example, intervals of 2.30− *2.60* and *2.60*− 2.90 are a problem because the GPA 2.60 is in both intervals leading to confusion regarding within which interval to place someone with a GPA of 2.60.

Table 2.9 includes the cumulative percentages for the GPA example. As an example of a cumulative percentage, the cumulative percentage for the 2.30–2.59 interval (19%) has been calculated by combining the percentages for three intervals: < 2.00, 2.00–2.29, and 2.30–2.59: 2% + 7% + 10% = 19%. As a result, we could conclude that 19% of this sample had GPAs less than 2.60. The accumulating of percentages continues until all (100%) of the scores in the sample are combined. As

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

shown in Table 2.9 , cumulative percentages are typically located in the final column of a frequency distribution table and labeled “Cum %.”

Table 2.9 Frequency Distribution Table With Cumulative Percentages for the GPA Variable

GPA	*f*	%	Cum %
3.80+	3	7%	**100%**
3.50−3.79	5	12%	**93%**
3.20−3.49	9	21%	**81%**
2.90−3.19	10	24%	**60%**
2.60−2.89	7	17%	**36%**
2.30−2.59	4	10%	**19%**
2.00−2.29	3	7%	**9%**
< 2.00	1	2%	**2%**
Total	42	100%	

2.5 Examining Data Using Figures

In addition to organizing data into frequency distribution tables, researchers also examine data visually using figures such as charts or graphs. There are four types of figures often used to visually portray data for variables: bar charts, pie charts, histograms, and frequency polygons. Which type of figure a researcher may use depends on the variable’s level of measurement (nominal, ordinal, interval, ratio). Data for variables measured at the nominal or ordinal level of measurement are typically displayed using bar charts or pie charts. The values of interval and ratio variables are graphed using histograms and frequency polygons. Each of these types of visual illustration is discussed in the following sections.

Displaying Nominal and Ordinal Variables: Bar Charts and Pie Charts

The values of variables measured at the *nominal* level of measurement differ in category or type; one example of a nominal variable is the lottery ticket variable discussed earlier, which consisted of four categories: sequence, pattern, nonequilibrated, and random. The values of variables measured at the *ordinal* level of measurement can be placed in an order relative to the other values; age (young, middle age, elderly) is one example of an ordinal variable.

One way to display nominal and ordinal variables is to use a bar chart. A **bar chart** is a figure that uses bars to represent the frequency or percentage of a sample corresponding to each value of a variable, with the different bars not touching each other. Figure 2.1 shows a bar chart for the data from the lottery ticket variable. In this figure, each bar corresponds to one of the four types of lottery tickets; the height of each bar corresponds to the frequency (*f*) of nonproblem gamblers who selected that type of ticket.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

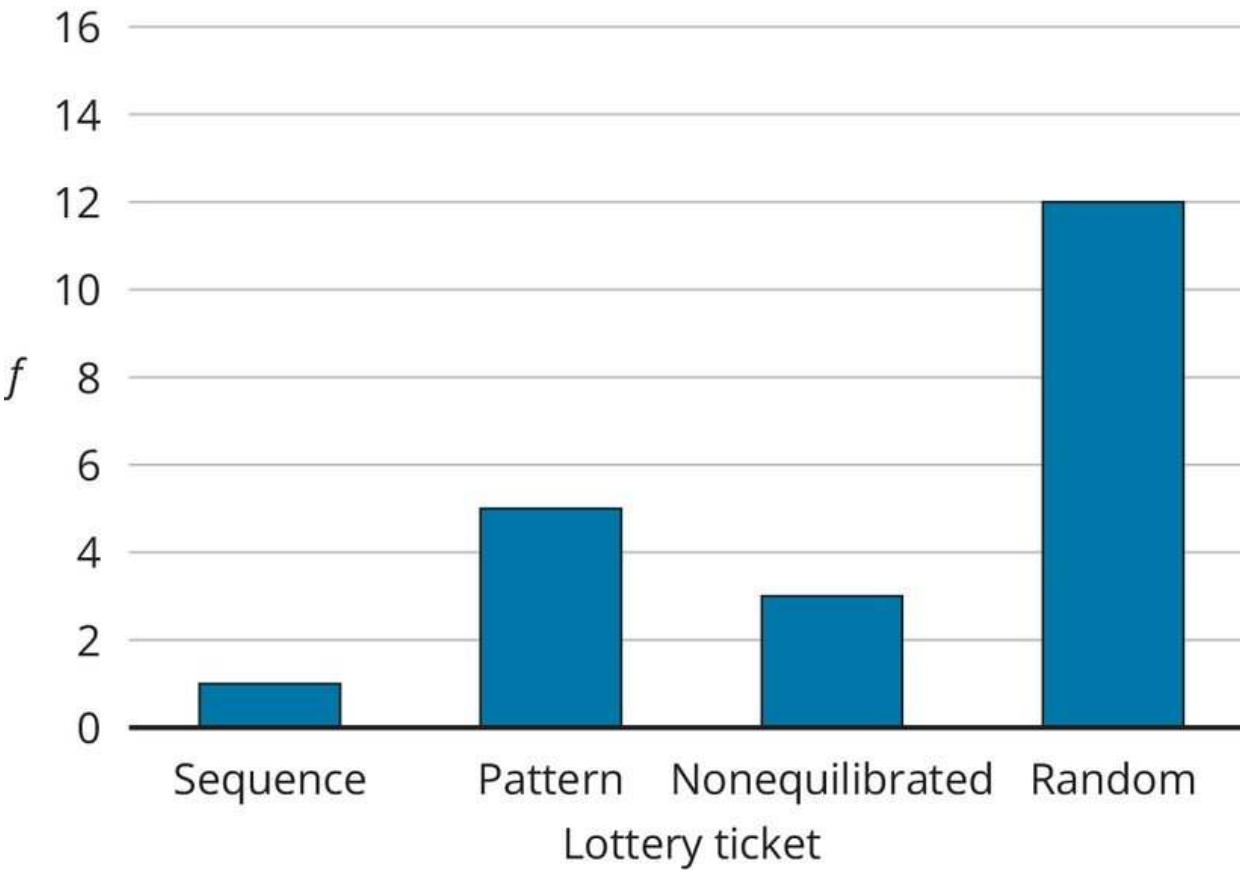


Figure 2.1 Bar Chart Showing the Frequencies of the Lottery Ticket Variable for the Nonproblem Gamblers

When creating a bar chart, the horizontal (X) axis contains the name of the variable and the values of the variable; the vertical (Y) axis contains the frequency (*f*) or percentage (%) for each value. Notice that the bars in the bar chart in Figure 2.1 do not touch each other; the gap between the bars indicates that the values of the variable represent distinct groups or categories that can't be placed along a numeric continuum.

A second type of figure used to display nominal or ordinal variables is a pie chart. A **pie chart** is a figure that uses a circle divided into proportions (slices) that represent the percentage of the sample for each value of a variable. Pie charts are created by dividing the 360 degrees of a circle into proportions corresponding to the percentage of each value.

Figure 2.2 shows a pie chart for the data for the lottery ticket variable. To illustrate how pie charts are constructed, the frequency distribution table in Table 2.4 shows that 5% of nonproblem gamblers chose the sequence ticket. Given that a pie chart consists of 360 degrees, 5% of the 360 degrees are needed to represent the sequence ticket. As 5% * 360 = 18, 18 degrees of the pie is assigned to the sequence ticket. Similarly, 97 degrees of the pie chart (27% of 360) are used to represent the pattern ticket. This process continues until all 360 degrees of the pie chart are accounted for.

The decision whether to use a bar chart or pie chart to illustrate a variable may depend on the number of values for the variable. For example, a pie chart for a variable that has 12 values may

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

result in many of the 12 “slices” being very thin and small, making them hard to see, meaning that a bar chart may be a better choice to illustrate this variable.

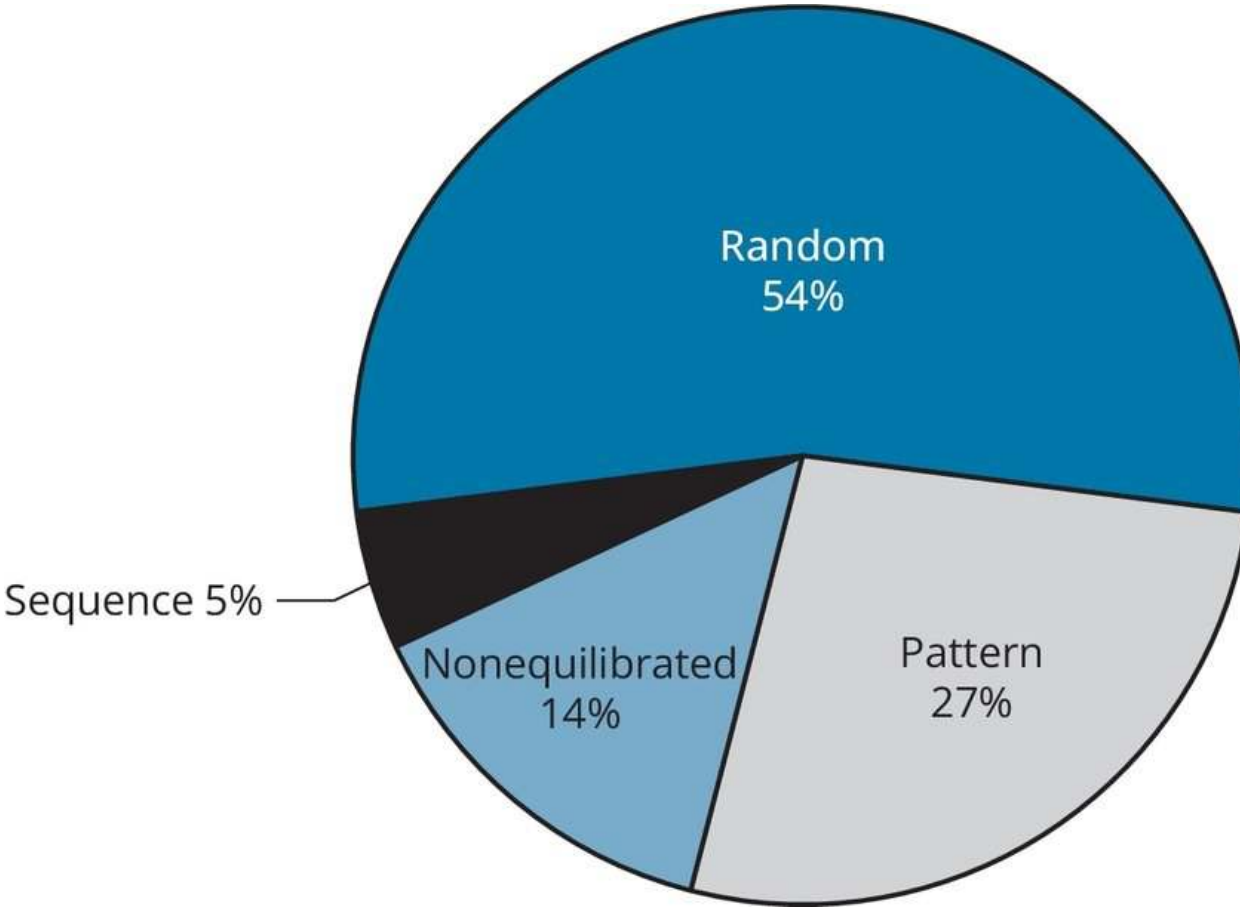


Figure 2.2 Pie Chart Showing the Percentages of the Lottery Ticket Variable for the Non−Problem Gamblers

Displaying Interval and Ratio Variables: Histograms and Frequency Polygons

Interval and ratio variables have values equally spaced along a numeric continuum and are identical to each other with one exception: Ratio scales possess a true zero point, for which the value of zero (0) represents the complete absence of the variable. The values of interval and ratio variables may be illustrated using histograms and frequency polygons, the choice of which depends on the number of values for the variable.

A **histogram** is a figure in which bars are used to represent the frequency of each value of a variable, with the different bars touching each other. The name and values of the variable are located on the X-axis, and the frequency or percentage for each value is on the Y-axis. Histograms are similar to bar charts, with one important difference: The bars in a histogram touch each other, indicating that the values of the variable are connected to each other along a numeric continuum.

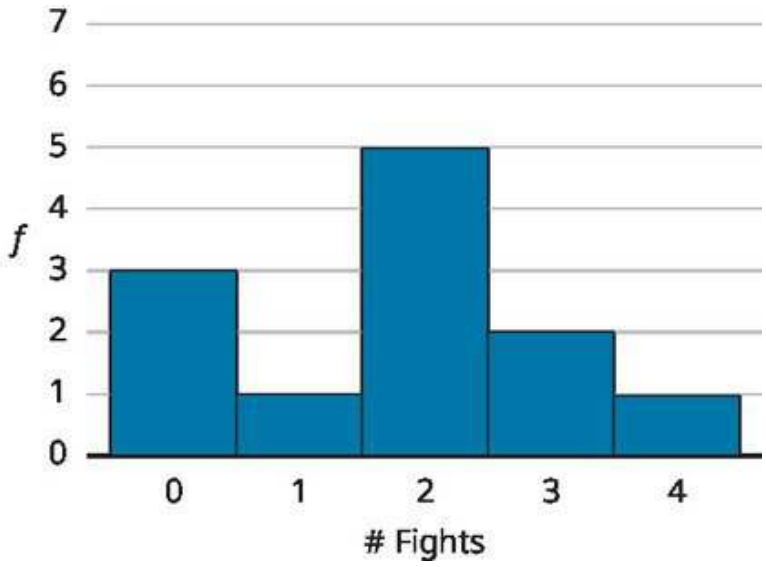
join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

As an example, imagine a teacher is interested in studying aggressive behavior in children and measures this variable by counting the number of fights each of her students is involved in over a 3-day period. “Number of fights” is a ratio variable rather than an interval variable because the value of 0 represents the complete absence of fights. Figure 2.3 provides a frequency distribution table and histogram for this variable.

A histogram is used when an interval or ratio variable consists of a relatively small number of values. But what if, for example, the number of fights could be anywhere from 0 to 20 rather than 0 to 5? In order to have the histogram fit on one page, the 21 bars would need to be very thin and crowded together, making the histogram confusing and difficult to read. In situations such as this, the values of the variable may be more clearly illustrated using a figure known as a **frequency polygon**, which is a line graph that uses data points to represent the frequency of each value of a variable, with straight lines connecting the data points. In a frequency polygon, the X-axis represents the name and values of the variable, while the Y-axis represents the frequency for each value.

# Fights	*f*
4	1
3	2
2	5
1	1
0	3
Total	12



▶ Long Description: Figure 2.3 Frequency Distribution Table and Histogram for the Number of Fights Variable

Figure 2.3 Frequency Distribution Table and Histogram for the Number of Fights Variable

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

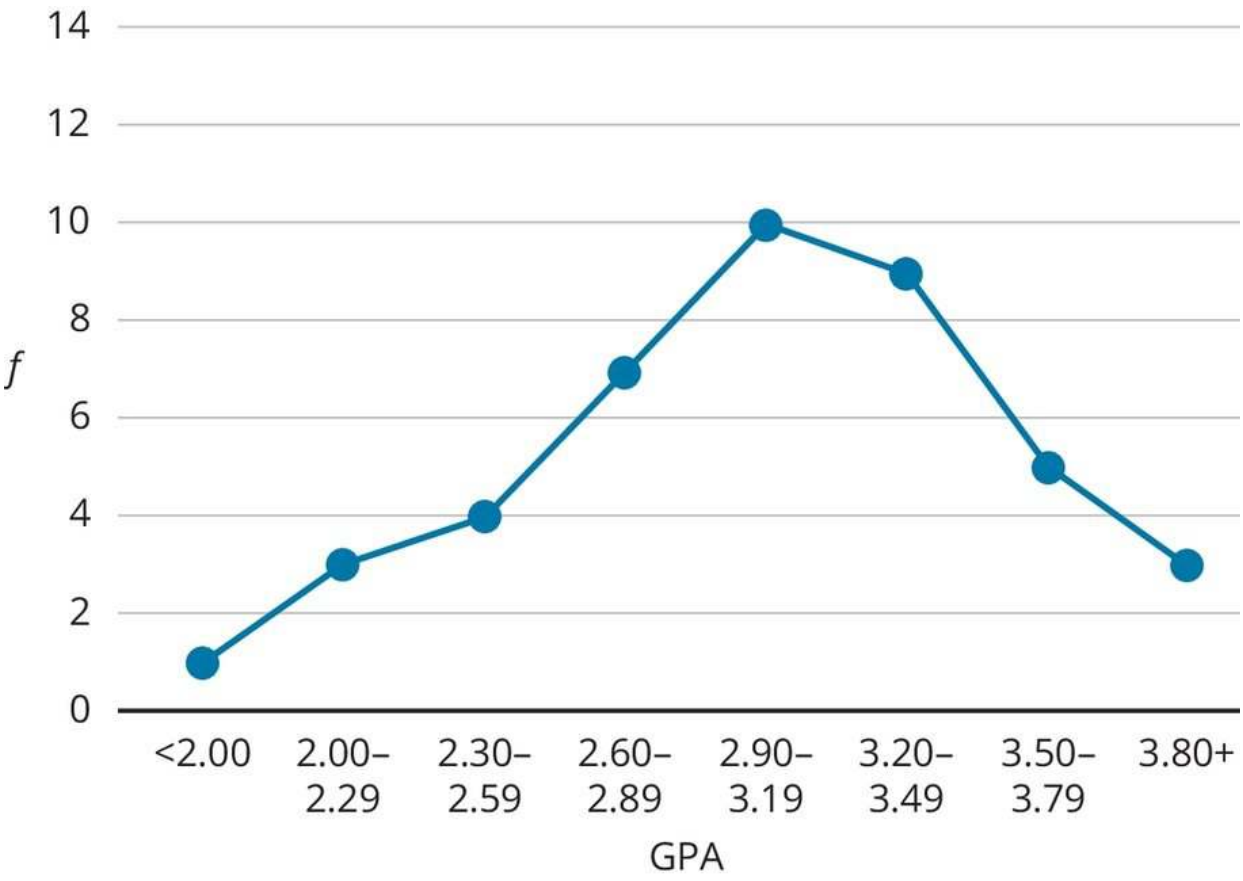


Figure 2.4 Frequency Polygon Showing the Frequency of Students in Each Interval of the Grade Point Average (GPA) Variable

Figure 2.4 shows a frequency polygon based on the grouped frequency distribution table of the earlier example of students' GPA (Table 2.7(b)). In this frequency polygon, the data points representing the frequencies of the eight GPA intervals are connected using straight, sharp lines. These lines indicate that the values of the variable can be connected together along a numeric continuum.

Although a frequency polygon like the one in Figure 2.4 may look jagged as it moves from one data point to the next, it provides a general sense of the shape of the distribution of scores for a variable. For example, looking at Figure 2.4 , you might conclude that the majority of GPAs in this sample are in the middle of the 1.00 to 4.00 range, with relatively low frequencies of students with either low or high GPAs.

Learning Check 2

Reviewing What You've Learned So Far

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



1. Review questions
    a. What is the difference between a frequency distribution table and a grouped frequency distribution table?
    b. For what levels of measurement would you create a pie chart, bar chart, histogram, or a frequency polygon?
    c. What is the difference between a bar chart and a histogram?
    d. Why would it be inappropriate to create a histogram or frequency polygon for a nominal or ordinal variable?
    e. Under what circumstances might you use a frequency polygon rather than a histogram to graph an interval or ratio variable?
2. For each of the following variables, identify whether you could use a bar chart, pie chart, histogram, and/or a frequency polygon to visually display the data.
    a. Miles per gallon (MPG)
    b. Living situation (on-campus, off-campus)
    c. Number of brothers and sisters
    d. Marital status (single, married, divorced)

Drawing Inappropriate Conclusions From Figures

A well-constructed figure can greatly facilitate the understanding of complex information. However, a poorly designed figure may not only be confusing, it can lead to inappropriate or misleading conclusions. Box 2.1 provides guidelines for constructing figures based on recommendations from the American Psychological Association (APA).

A simple example may demonstrate how altering the features of a figure can lead to different interpretations of the same set of data. Imagine a pollster assesses how voters with different political affiliations plan to vote on a ballot measure. She asks a sample of Democrats, Republicans, and Independents the question, “Do you plan on voting in favor of Measure A?” In examining her data, the pollster finds 56% of Democrats, 47% of Republicans, and 52% of Independents intend to vote “yes” on the measure.

Figure 2.5 graphs these three percentages using two bar charts; bar charts are used to display the data because the variable of interest (political affiliation) is a nominal variable (Democrat, Republican, or Independent). The bars don’t touch each other because the three political affiliations are distinct from each other and can’t be placed along a numeric continuum.

Although the two bar charts are similar, there is one important difference in how they are constructed. Following the guidelines in Box 2.1, the bar chart in Figure 2.5 (a) starts the Y-axis with the value 0% and ends with 100%. However, the bar chart in Figure 2.5 (b) starts with the value 45% and ends with 65%. By limiting the range of values of the percentages, the bar chart in Figure 2.5 (b) exaggerates differences in the heights of the three bars. As a result, someone looking at Figure 2.5 (b) may interpret the differences between the three political affiliations as being larger than someone looking at Figure 2.5 (a).

The saying “a picture is worth a thousand words” is particularly relevant for the presentation of statistical data. Because figures can provide a vivid representation of data, it’s the responsibility of researchers to portray their data in an accurate manner and to avoid creating images that may mislead or confuse.

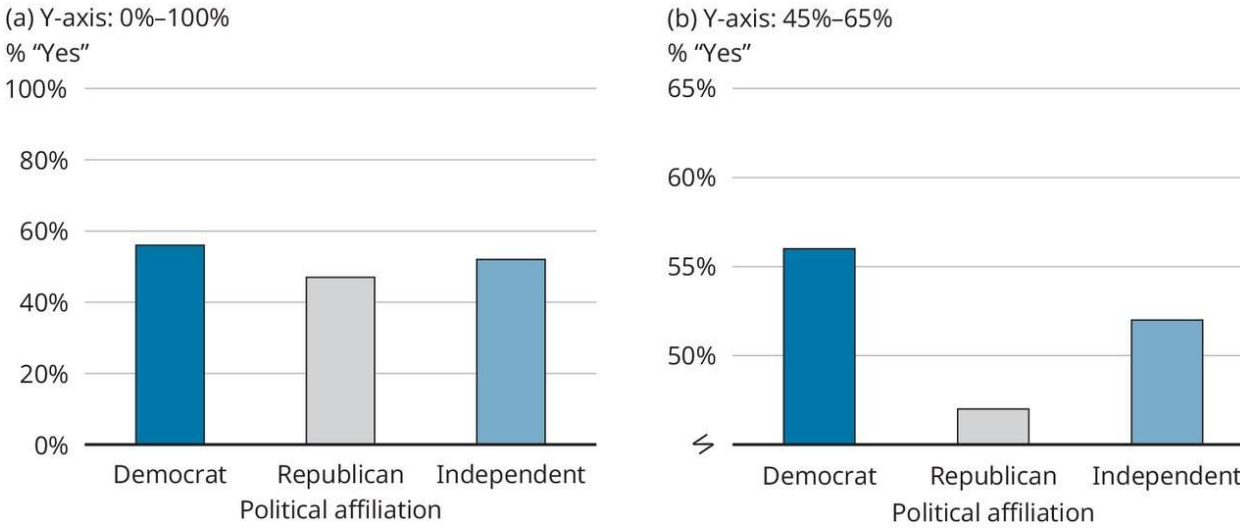
join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Box 2.1 Some APA Guidelines for Creating Figures

1. The figure should be formatted in a portrait rather than landscape page orientation (the reader should not turn the page sideways to read the table).
2. The figure is numbered using Arabic numbers (Figure 1, not Figure *I* or Figure *A*).
3. The figure includes a title that describes the contents of the figure. For example, it is not enough to name a figure simply " Figure 2.1 " or " Figure 2.1 . Bar chart."
4. Figures have a rectangular shape, with the height of the vertical (Y) axis two thirds to three fourths the length of the horizontal (X) axis.
5. The horizontal (X) axis contains the values of the variable; the vertical (Y) axis contains the frequency (*f*) or percentage (%) of the sample having each value of the variable.
6. The *X* - and *Y* -axes include descriptive labels written in both upper- and lowercase letters (as opposed to all capital letters).
7. The labels for the *X* - and *Y* -axes are short enough to fit on one line along the axes.
8. The label of the Y-axis is often placed parallel to the Y-axis (the letters face sideways rather than vertically) but can sometimes be seen on top of the y-axis for accessibility.
9. The Y-axis starts with the value zero (0) and is divided into equally spaced values or intervals.
10. The Y-axis is divided into 6 to 10 values (just enough values to accommodate bar length). The bars should neither be flattened along the bottom of the Y-axis nor bumped up against the top of the Y-axis.

Political affiliation	% "Yes"
Democrat	56%
Republican	47%
Independent	52%



▶ Long Description: Figure 2.5 Different Displays of the Same Data

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Figure 2.5 Different Displays of the Same Data

2.6 Examining Data: Describing Distributions

We've discussed how researchers construct tables and figures to organize their data. Looking at a table or figure enables researchers to describe the distribution of values or scores for a variable. Describing a distribution involves addressing questions such as, "Which values of the variable occur most often?," "How do the frequencies of the values of the variable change in relation to the most common values?," and "What is the amount of differences among the different values?"

Questions such as these pertain to three aspects of distributions: modality, symmetry, and variability. These three aspects of distributions are introduced in the following sections; they'll also be discussed in greater detail in the next two chapters of this book.

Modality

The first feature of a distribution of data for a variable is its modality. The **modality** of a distribution refers to the value or values of the variable that have the highest frequency or occur most often in a set of data. Using the lottery study as an example, identifying the modality of the lottery ticket variable involves answering the question, "Which of the four lottery tickets was chosen the most often?"

The modality of a distribution can be identified by looking at a frequency distribution table; the modality is the value(s) of the variable with the highest frequency (*f*) or percentage (%). The modality may also be determined by examining a figure and locating the tallest bar in a bar chart or histogram, the highest point in a frequency polygon, or the biggest "slice" in a pie chart. For the lottery ticket variable, the modality is the random ticket.

In identifying the modality of a distribution, it may be that one value or a limited number of values adjacent to one another clearly have the highest frequency. In such cases, the distribution is referred to as a unimodal distribution. A **unimodal distribution** is a distribution where one value of a variable occurs with the greatest frequency. Looking at the frequency distribution table (Table 2.4) or figure (Figure 2.1) for the lottery ticket variable, this distribution would be considered unimodal as one value, the random ticket, clearly occurred more often than the other three ticket choices.

In contrast to a unimodal distribution, it's possible that two distinct values may both have relatively high frequencies; if so, the distribution is referred to as a **bimodal distribution** , which is a distribution where two values clearly distinct from each other have the greatest frequency. A bimodal distribution sometimes resembles the two humps on a camel's back. As an example of a bimodal distribution, in April 2024 the Pew Research Center (www.pewresearch.org) asked a sample of 7,000 voters whether they felt Donald Trump had the mental fitness needed to be president, with five responses ranging from "extremely" to "not at all." The majority of voters chose one of the two extreme choices, either "extremely" (23%) or "not at all" (35%), with only 14% choosing the middle response of "somewhat."

And, in addition to unimodal and bimodal distributions, it's possible for a distribution to have more than two values with relatively equally high frequencies; this type of distribution is known as a **multimodal distribution** —a distribution where more than two values have the greatest frequency. In examining people's attitudes toward gun control, in May 2023, the news network CNN asked 1,227 people the question, "Do you think allowing gun owners to carry their guns in public places makes those places safer, less safe, or makes no difference in how safe a public place is?" They found people differed greatly in their responses, with 32% saying "makes public places safer," 36% saying

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

"makes public places less safe," and 32% saying "makes no difference," Figure 2.6 presents an example of a unimodal, bimodal, and multimodal distribution.

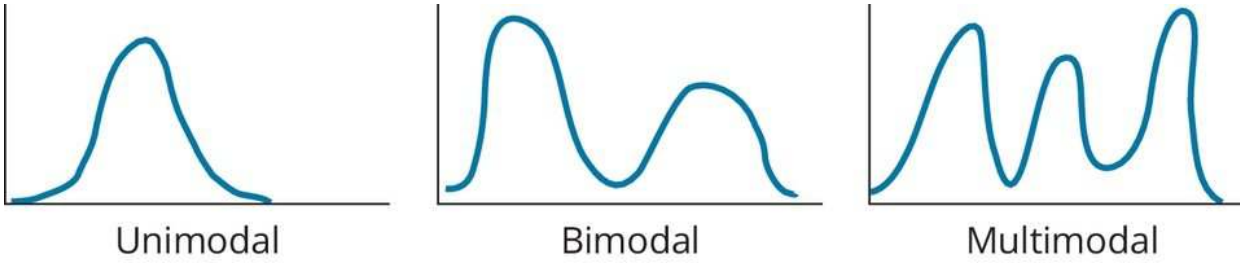


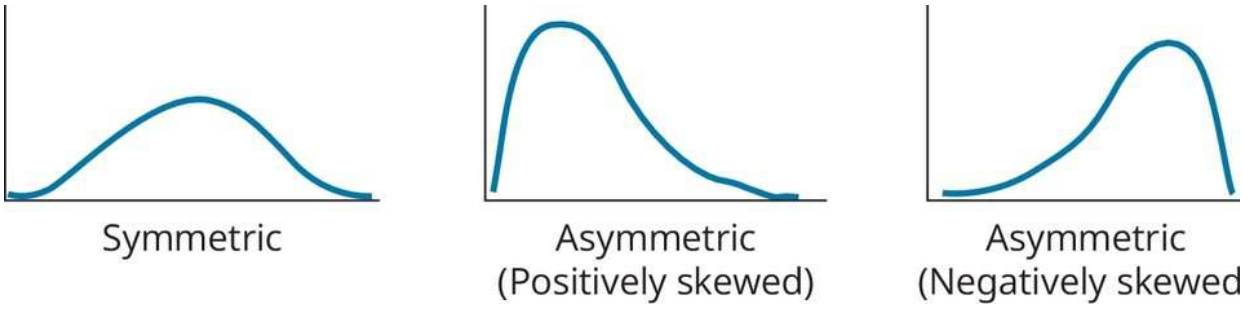
Figure 2.6 Distributions With Different Modalities (Unimodal, Bimodal, and Multimodal)

Symmetry

The second feature of a distribution of data is its symmetry. The **symmetry** of a distribution refers to how the frequencies of values of a variable change in relation to the most common or frequently occurring values. Describing the symmetry of a distribution means answering the question, "What is the shape of the distribution?"

The symmetry of a distribution can be determined by looking at the frequencies in a frequency distribution table or the bars or points in a figure. By doing so, a distribution may be described as being either symmetric or asymmetric. A **symmetric distribution** is one in which the frequencies change in a similar manner moving away in both directions from the most frequently occurring values. An example of a symmetric distribution is illustrated in Figure 2.7 . One way to determine whether a distribution is symmetric is to draw the distribution on a piece of paper and then fold the distribution vertically down the middle so that the two halves are on top of each other. A distribution is symmetric when the two halves are similar in shape such that they resemble mirror images of each other.

However, a distribution may also be asymmetric. An **asymmetric distribution** (also known as a *skewed* distribution) is one in which the frequencies change in a different manner moving away in both directions from the most frequently occurring values. In asymmetric distributions, the highest frequencies are located at one end of the distribution rather than in the middle, such that if you fold an asymmetric distribution down the middle, the shapes of the two halves do not look the same.



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Figure 2.7 Distributions With Different Symmetries (Symmetric, Positively Skewed, and Negatively Skewed)

In an asymmetric distribution, the frequencies take on the shape of a long tail as the values move toward the end of the distribution; scores in the tail of the distribution are often considered to be outliers. For example, in the 2022 Current Population Survey, the U.S. Census Bureau (www.census.gov) estimated that most Americans (63%) had annual incomes less than $50,000. However, a very small percentage (3%) were greater than $200,000; these higher incomes would be considered outliers.

Depending on the location of its tail, asymmetric distributions are referred to either as positively skewed or negatively skewed. A **positively skewed distribution** is one in which the higher frequencies are at the lower end of the distribution, with the tail on the upper (right) end of the distribution; a **negatively skewed distribution** is one in which the higher frequencies are at the upper end of the distribution, with the tail on the lower (left) end of the distribution. Figure 2.7 provides an illustration of a symmetric distribution and two asymmetric distributions (one positively skewed and one negatively skewed). Annual income, illustrated by the U.S. Census Bureau example mentioned earlier, is an example of a positively skewed distribution.

Variability

The third way in which distributions may be described is in terms of the amount or nature of differences among the different values for a variable. The **variability** of a distribution refers to the amount of differences in a distribution of data for a variable. The variability in a distribution is described by answering the question, “To what extent are the scores in the distribution different from each other?”

The amount of variability in a distribution can be estimated by looking at a frequency distribution table or figure. Imagine, for example, an instructor gives an 11-item quiz to her class and finds almost all of the students answer six, seven, or eight of the items correctly. In this situation, the distribution would resemble the shape of a tall mountain: thin and tall, with many of the scores residing in a few values of the variable. A distribution with a small amount of variability such as this one is referred to as a **peaked distribution** , defined as a distribution where much of the data are in a small number of values of a variable.

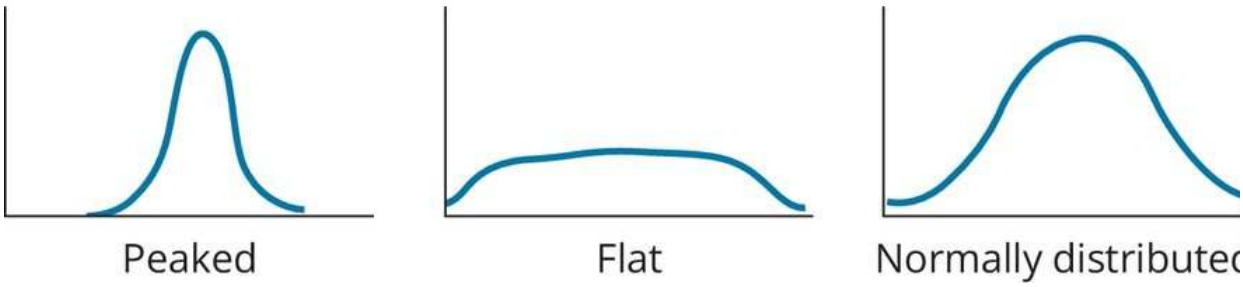


Figure 2.8 Distributions With Different Variability (Peaked, Flat, and Normally Distributed)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

On the other hand, what would the distribution look like if she had found her students' scores were distributed relatively evenly across the entire range of possible scores? In this situation, the distribution would not be tall and thin like a mountain but rather would be low and wide like a plateau. This type of distribution is referred to as a **flat distribution** , which is a distribution where the data are spread evenly across the values of a variable. The CNN gun control example described earlier is an example of a flat distribution.

In discussing modality, symmetry, and variability, distributions may have more than one mode, may be asymmetric (skewed), and may be peaked or flat. But what if a distribution has none of these qualities? A distribution that is unimodal, symmetric, and neither peaked nor flat may be described as "bell shaped." If you look at a bell, you'll notice that it is the highest in the middle, and as you move away from the top of a bell in both the left and right directions, the shape of the bell slopes downward in the same manner. A bell-shaped distribution for a variable is given a special name: a normally distributed variable.

A **normally distributed variable** is a distribution for a variable that is considered unimodal, symmetric, and neither peaked nor flat. Note that the word *normal* is a statistical concept and not a value judgment. In other words, asymmetric, peaked, or flat distributions are not considered "abnormal." The concept of normal distributions will be discussed in greater detail in Chapter 5. Figure 2.8 provides an illustration of three distributions with different amounts of variability: peaked, flat, and normally distributed.

Learning Check 3

Reviewing What You've Learned So Far



1. Review questions
    a. What are the three aspects used to describe distributions of variables?
    b. What is the difference between unimodal, bimodal, and multimodal distributions?
    c. What is the difference between symmetric and asymmetric (skewed) distributions?
    d. What is the difference between positively and negatively skewed distributions?
    e. What is the difference between peaked and flat distributions?
    f. What are the characteristics of a normally distributed variable?
2. For each of the following sets of data, determine whether the distribution of scores is unimodal, bimodal, or multimodal.
    a. 4, 2, 11, 6, 10, 2, 6, 13, 2, 10, 6, 10, 1
    b. 15, 17, 13, 18, 17, 14, 20, 18, 17, 14, 17
    c. 20, 10, 50, 10, 80, 30, 70, 10, 60, 10, 80, 0, 30, 80, 10, 100, 80
3. For each of the following sets of data, determine whether the distribution of scores is symmetric or asymmetric (skewed).
    a. 4, 5, 1, 4, 2, 4, 3, 5, 3, 4, 4, 3
    b. 220, 230, 250, 230, 200, 240, 210, 230, 270, 220, 240, 230, 240, 230
    c. 3, 1, 9, 2, 3, 6, 2, 7, 3, 5, 2, 6, 3, 2, 2
4. For each of the following sets of data, determine whether the distribution of scores is peaked or flat.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

a. 16, 10, 19, 12, 13, 15, 14, 17, 15, 20, 11, 13, 19, 16, 15
b. 12, 15, 12, 11, 13, 11, 12, 12, 14, 10, 12, 11, 16, 12, 13, 12, 11
c. 14, 12, 16, 14, 13, 15, 12, 14, 14, 15, 14, 14, 13, 14

2.7 Looking Ahead

This chapter has discussed why and how researchers examine data. Examining data by creating tables and figures requires a number of different skills, including organizational skills (to prepare a set of data for statistical analysis), communication skills (to present data in tables or figures), and conceptual skills (to describe a distribution). The next two chapters focus on mathematical skills related to the statistical analysis of data, describing how researchers describe distributions numerically rather than visually. Given that this is a statistics textbook, we will focus on calculating statistics to describe and analyze data. However, and we cannot say this strongly enough, it is *extremely* important to examine a set of data before conducting statistical analyses as this will prevent errors both in your calculations of statistics as well as your interpretation of the results of these calculations.

2.8 Summary

Before conducting statistical analyses, researchers typically examine their data to gain an initial sense of the data, detect data coding or data entry errors, identify ***outliers*** (rare, extreme scores that lie outside of the range of the majority of scores in a set of data), evaluate research methodology, and determine whether the data meet statistical criteria and assumptions.

One common way of examining variables involves creating tables known as ***frequency distribution tables*** (tables that summarize the number and percentage of participants that express the different values for a variable) and ***grouped frequency distribution tables*** (tables that group the values of a variable measured at the interval or ratio level of measurement into a small number of intervals and then provide the frequency and percentage within each interval). To create a grouped frequency table, the ***real limits*** of each interval (the values of the variable that fall halfway between the top of one interval and the bottom of the next interval) need to be identified.

Data for a variable may also be examined by creating figures such as bar charts, pie charts, histograms, and frequency polygons. ***Bar charts*** and ***pie charts*** are used for variables measured at the nominal or ordinal level of measurement; a bar chart uses bars to represent the frequency or percentage of a sample corresponding to each value of a variable, with the different bars not touching each other; a pie chart is a circle divided into proportions that represent the percentage of the sample corresponding to each value of a variable. ***Histograms*** and ***frequency polygons*** are used for variables measured at the interval or ratio level of measurement; a histogram is a figure in which bars are used to represent the frequency of each value of a variable, with the different bars touching each other; a frequency polygon is a line graph that uses data points to represent the frequency of each value of a variable, with lines connecting the data points.

From examining their data, researchers can describe the distribution of values or scores for a variable. There are three main aspects of distributions: modality, symmetry, and variability.

The ***modality*** of a distribution is the value(s) of the variable that have the highest frequencies or occur most often in a set of data. ***Unimodal distributions*** are distributions where one value occurs with the greatest frequency, ***bimodal distributions*** have two values with the greatest frequency, and ***multimodal distributions*** have more than two values with the greatest frequencies.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The ***symmetry*** of a distribution refers to how the frequencies of values of a variable change in relation to the most common or frequently occurring values of the variable. In ***symmetric distributions*** , the frequencies change in a similar manner, moving away in both directions from the most common values of the variable. In ***asymmetric (skewed) distributions*** , the frequencies change in a different manner, moving away in both directions from the most common values of the variable.

The ***variability*** of a distribution refers to the amount of differences in a distribution of data. In ***peaked distributions*** , much of the data are in a few values of a variable; in ***flat distributions*** , the data are spread evenly across the values of a variable. A ***normally distributed variable*** has a symmetric distribution that is neither peaked nor flat.

2.9 Important Terms

asymmetric (skewed) distribution

bar chart

bimodal distribution

cumulative percentage

flat distribution

frequency (*f*)

frequency distribution table

frequency polygon

grouped frequency distribution table

histogram

modality

multimodal distribution

negatively skewed distribution

normally distributed variable

outlier

peaked distribution

pie chart

positively skewed distribution

real limits

real lower limit

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

real upper limit

symmetric distribution

symmetry

unimodal distribution

variability

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

2.10 Formula Introduced in This Chapter

Percentage (%)

$$\% = \frac{f}{total\ number\ of\ scores} * 100$$

2.11 Using IBM SPSS Software *

Creating Data Files: Defining Variables and Entering Data

* IBM SPSS software. SPSS is a registered trademark of International Business Machines Corporation.

1. Open a new data file (See Figure 2.9).

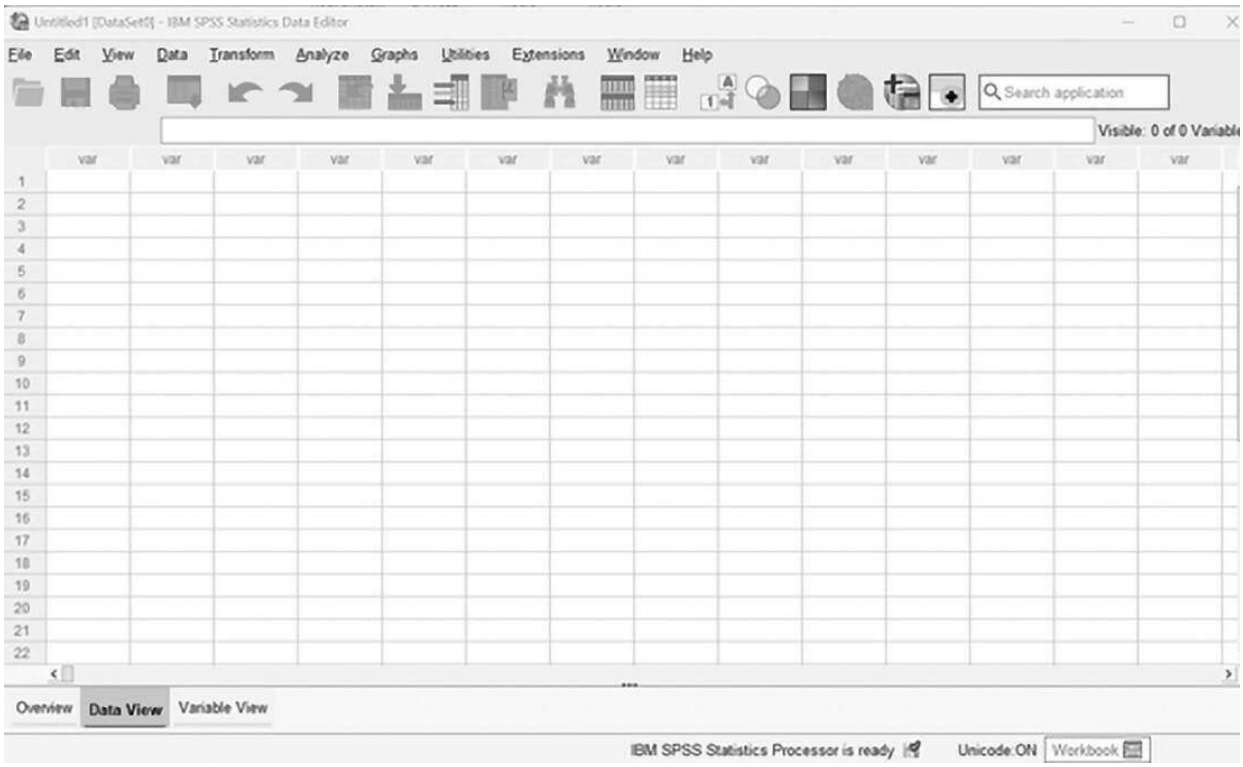


Figure 2.9 Data View

2. Define and name variable(s) (See Figure 2.10).

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

How? (1) Click **Variable View** tab, (2) click below **Name** , and (3) type name of variable in the cell.

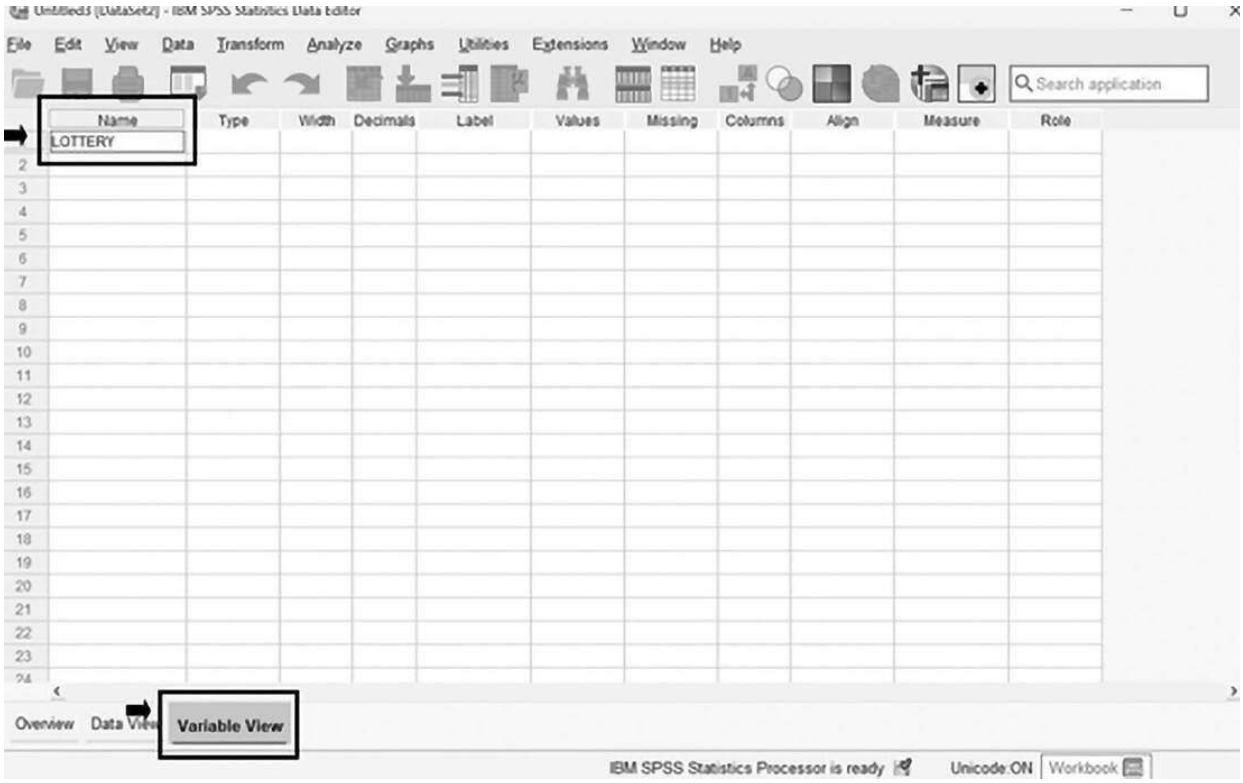


Figure 2.10 Variable View

3. Determine number of decimals for your variable(s) (See Figure 2.11).

How? Click desired number of decimal places in the **Decimals** box.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

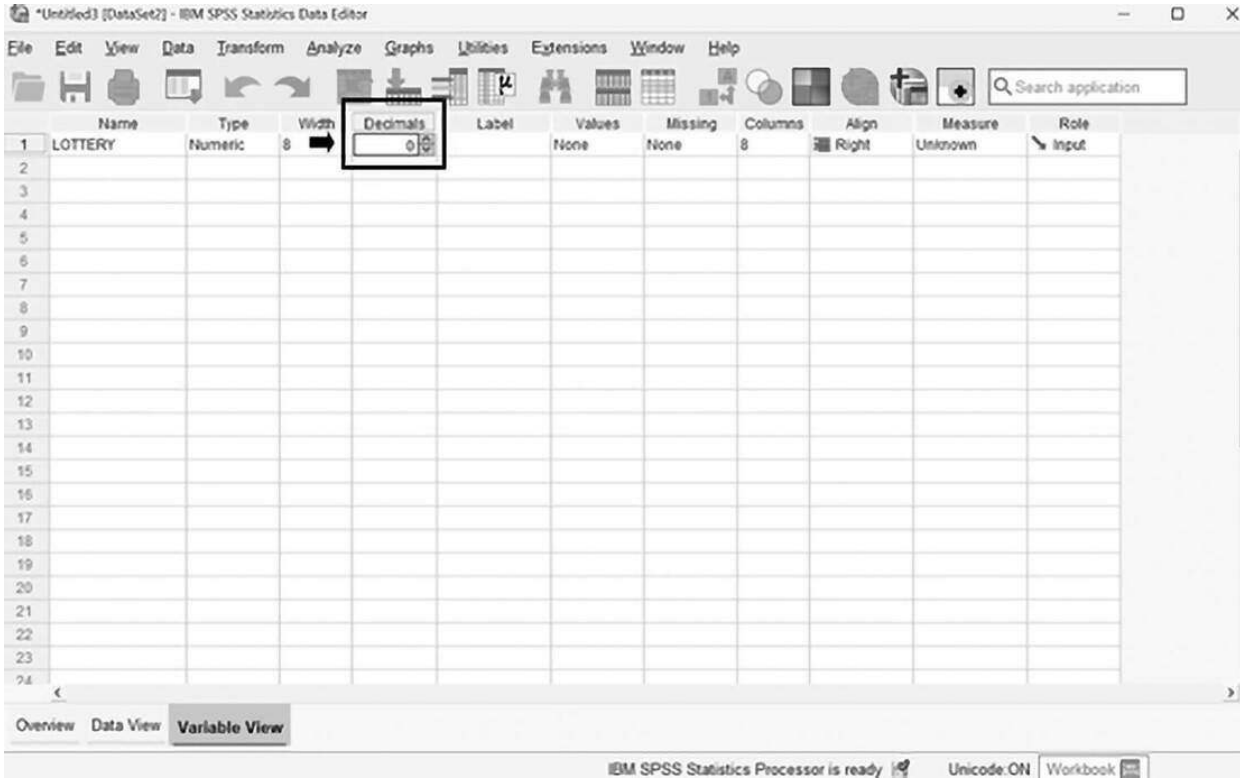


Figure 2.11 Decimals Box

4. Provide labels for your variable(s) (See Figure 2.12).

How? (1) Click below **Label** , and (2) type the label for the variable in the cell.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

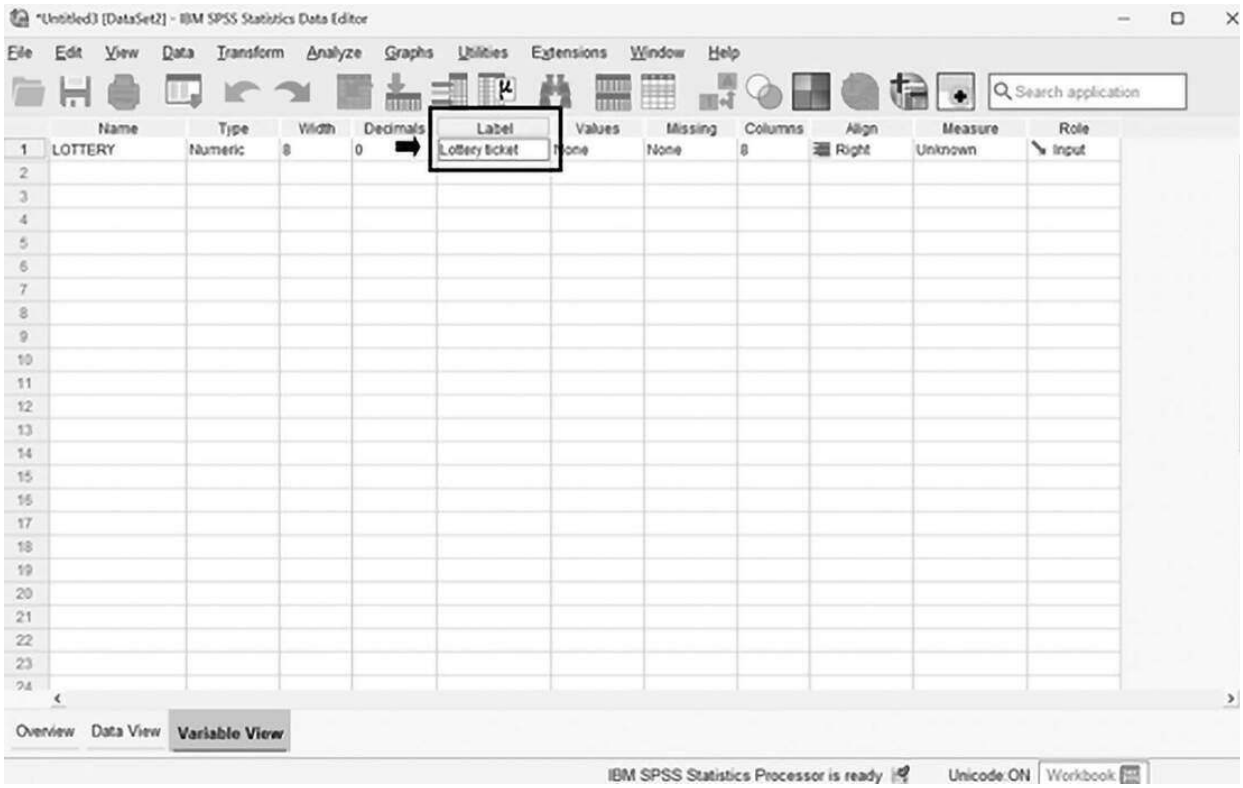


Figure 2.12 Label Box

5. Provide labels for values of nominal (categorical) variable(s) (See Figure 2.13).

How? (1) Click below **Values** , and (2) type labels for coded values of the variable.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

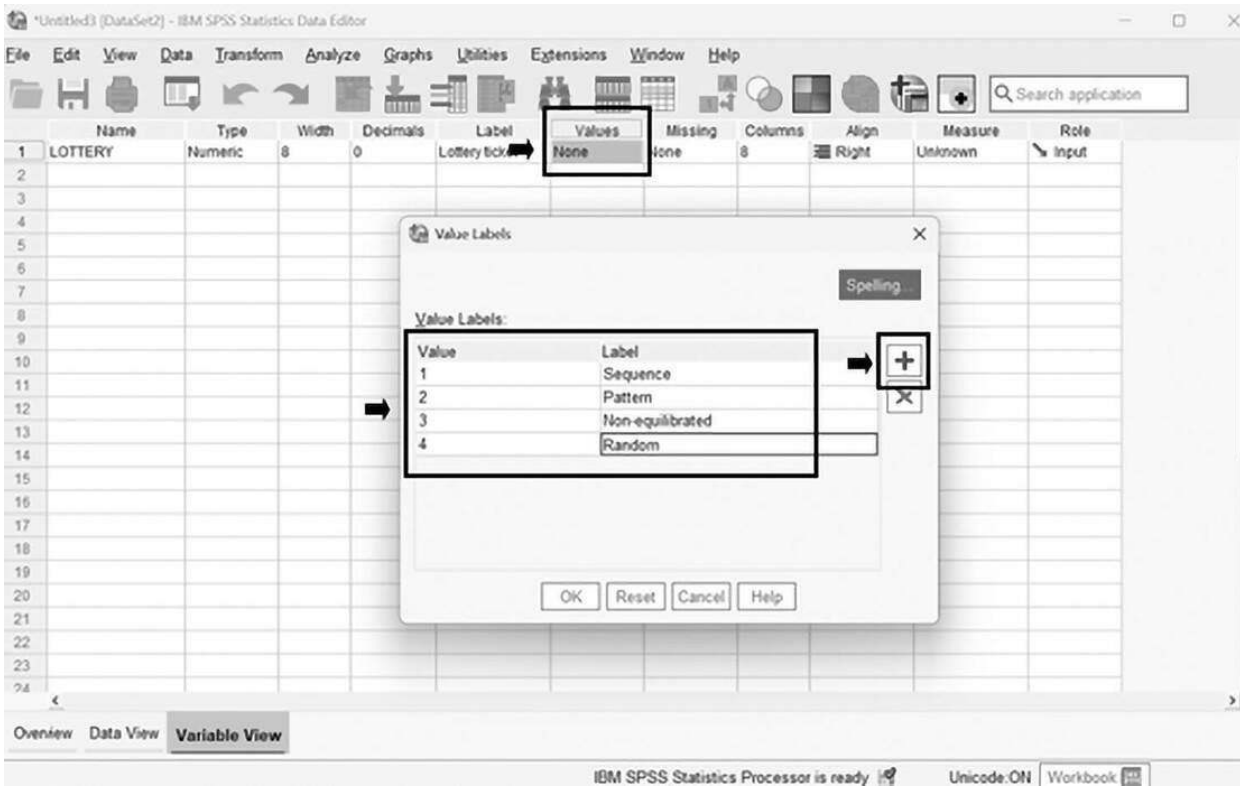


Figure 2.13 Nominal Variables

6. Enter data (See Figure 2.14).

How? (1) Click **Data View** tab and (2) enter data for the variable in the cells.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

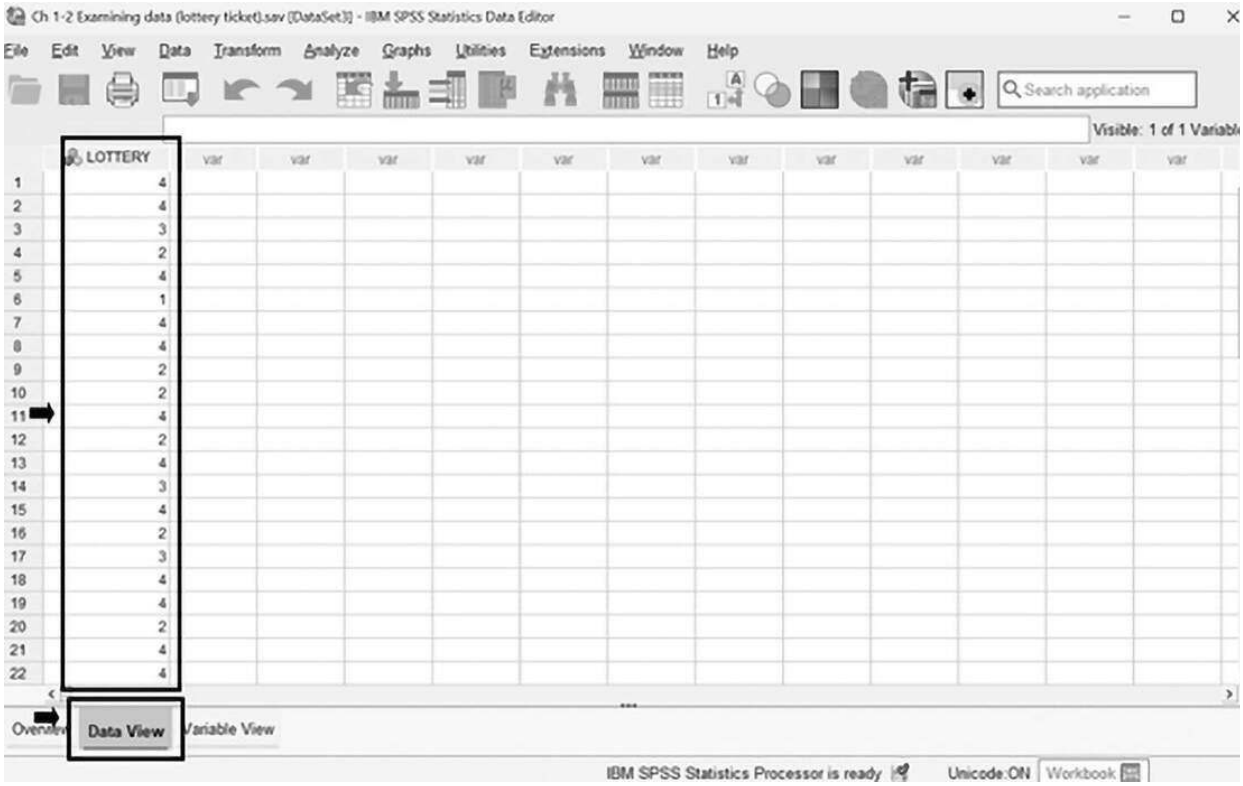


Figure 2.14 Enter Data View

Examining Data: Frequency Distribution Tables and Figures: The Lottery Study (2.1)

1. Select the frequency distribution procedure within SPSS (See Figure 2.15).

   How? (1) Click **Analyze** menu, (2) click **Descriptive Statistics** , and (3) click **Frequencies** .

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

To Jennifer , my kids Meagan and Eika , and my parents, Katsumi and Grayce Tokunaga.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

As a global academic publisher, Sage is driven by the belief that research and education are critical in shaping society. Our mission is building bridges to knowledge—supporting the development of ideas into scholarship that is certified, taught, and applied in the real world.

Sage's founder, Sara Miller McCune, transferred control of the company to an independent trust, which guarantees our independence indefinitely. This enables us to support an equitable academic future over the long term by building lasting relationships, championing diverse perspectives, and co-creating social and behavioral science resources that transform teaching and learning.

▶ Long Description:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

FUNDAMENTAL STATISTICS FOR THE SOCIAL AND BEHAVIORAL SCIENCES

Third Edition

Howard T. Tokunaga

San Jose State University



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

eISBN: 978-1-0718-6160-8

Copyright © 2027 by Sage.

All rights reserved. Except as permitted by U.S. copyright law, no part of this work may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without permission in writing from the publisher.

All third party trademarks referenced or depicted herein are included solely for the purpose of illustration and are the property of their respective owners. Reference to these trademarks in no way indicates any relationship with, or endorsement by, the trademark owner.



FOR INFORMATION:

2455 Teller Road, Thousand Oaks, California 91320, E-mail: orders@sagepub.com

3rd floor, HYLO, 103-105 Bunhill Row, London, EC1Y 8LZ, United Kingdom

10th Floor, Emaar Capital Tower 2, MG Road Sikanderpur, Sector 26, Gurugram, Haryana – 122002, India

18 Cross Street #10-10/11/12, China Square Central, Singapore 048423

Library of Congress Control Number: 2026943277

ISBN: 978-1-0718-6144-8

Printed in the United States of America

This book is printed on acid-free paper.

26 27 28 29 30 10 9 8 7 6 5 4 3 2 1

Acquisitions Editor: Anna Villarruel

Content Development Editor: Kenzie Offley

Editorial Assistant: Daria Terry

Production Editor: Syeda Aina Rahat Ali

Copy Editor: Michelle Ponce

Typesetter: Lumina Datamatics

Cover Designer: Candice Harman

Marketing Manager: Victoria Velasquez

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CONTENTS

Preface
Acknowledgments
About the Author
Chapter 1 Introduction to Statistics
Chapter 2 Examining Data: Tables and Figures
Chapter 3 Measures of Central Tendency
Chapter 4 Measures of Variability
Chapter 5 Normal Distributions
Chapter 6 Probability and Introduction to Hypothesis Testing
Chapter 7 Testing One Sample Mean
Chapter 8 Estimating the Mean of a Population
Chapter 9 Testing the Difference Between Two Means
Chapter 10 Errors in Hypothesis Testing, Statistical Power, and Effect Size
Chapter 11 One-Way Analysis of Variance (ANOVA)
Chapter 12 Two-Way Analysis of Variance (ANOVA)
Chapter 13 Correlation
Chapter 14 Linear Regression and Multiple Correlation
Chapter 15 Chi-Square
Appendix A: Tables
Appendix B: Review of Basic Mathematics
Glossary
References
Index

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

DETAILED CONTENTS

Preface

Acknowledgments

About the Author

Chapter 1 Introduction to Statistics

- 1.1 What Is Statistics?
- 1.2 Why Learn Statistics?
- 1.3 Introduction to the Stages of the Research Process
  - Developing a Research Hypothesis to Be Tested
    - Identifying a Question or Issue to Be Examined
    - Reviewing and Evaluating Relevant Theories and Research
    - Stating a Research Hypothesis: Independent and Dependent Variables
  - Collecting Data
    - Drawing a Sample From a Population
    - Determining How Variables Will Be Measured: Levels of Measurement
    - Selecting a Method to Collect the Data: Experimental and Nonexperimental Research Methods
  - Analyzing the Data
    - Calculating Descriptive Statistics
    - Calculating Inferential Statistics
  - Drawing a Conclusion Regarding the Research Hypothesis
  - Communicating the Findings of the Study
- 1.4 Plan of the Book
- 1.5 Looking Ahead
- 1.6 Summary
- 1.7 Important Terms
- 1.8 Exercises

Answers to Learning Checks

Learning Check 1

Learning Check 2

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Odd-Numbered Exercises

Chapter 2 Examining Data: Tables and Figures

2.1 An Example From the Research: Winning the Lottery

2.2 Why Examine Data?

Gaining an Initial Sense of the Data

Identifying Outliers

Evaluating Research Methodology

Determining Whether Data Meet Statistical Criteria and Assumptions

2.3 Examining Data Using Tables

What Are the Values of the Variable?

How Many Participants in the Sample Have Each Value of the Variable?

What Percentage of the Sample Has Each Value of the Variable?

2.4 Grouped Frequency Distribution Tables

Cumulative Percentages

2.5 Examining Data Using Figures

Displaying Nominal and Ordinal Variables: Bar Charts and Pie Charts

Displaying Interval and Ratio Variables: Histograms and Frequency Polygons

Drawing Inappropriate Conclusions From Figures

2.6 Examining Data: Describing Distributions

Modality

Symmetry

Variability

2.7 Looking Ahead

2.8 Summary

2.9 Important Terms

2.10 Formula Introduced in This Chapter

Percentage (%)

2.11 Using IBM SPSS Software *

Creating Data Files: Defining Variables and Entering Data

Examining Data: Frequency Distribution Tables and Figures: The Lottery Study (2.1)

2.12 Exercises

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 3 Measures of Central Tendency
3.1 An Example From the Research: The 10% Myth
3.2 Understanding Central Tendency
3.3 The Mode
Determining the Mode
3.4 The Median
Calculating the Median: Odd Number of Scores
Calculating the Median: Even Number of Scores
Calculating the Median When Multiple Scores Have the Median Value
3.5 The Mean
Calculating the Mean
Reporting Calculated Numbers
Rounding Calculated Numbers
Calculating the Mean From Frequency Distribution Tables
The Mean as a Balancing Point in a Distribution
The Sample Mean vs. the Population Mean: Statistics vs. Parameters
3.6 Comparison of the Mode, Median, and Mean
Strengths and Weaknesses of the Mode
Strengths and Weaknesses of the Median
Strengths and Weaknesses of the Mean
Choosing a Particular Measure of Central Tendency
3.7 Measures of Central Tendency: Drawing Conclusions
3.8 Looking Ahead
3.9 Summary
3.10 Important Terms
3.11 Formulas Introduced in This Chapter
Median (Odd Number of Scores)
Median (Even Number of Scores)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Mean
Mean (Frequency Distribution Table)
Mean (Balancing Point in a Distribution)
3.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 4 Measures of Variability
4.1 An Example From the Research: How Many Sometimes" in an Always"?
4.2 Understanding Variability
4.3 The Range
Strengths and Weaknesses of the Range
4.4 The Interquartile Range
Strengths and Weaknesses of the Interquartile Range
4.5 The Variance (s^2)
Definitional Formula for the Variance
Computational Formula for the Variance
Why Not Use the Absolute Value of the Deviation in Calculating the Variance?
Why Divide by N − 1 Rather Than N in Calculating the Variance?
4.6 The Standard Deviation (s)
Definitional Formula for the Standard Deviation
Computational Formula for the Standard Deviation
Eyeball Estimating" the Mean and Standard Deviation
4.7 Measures of Variability for Populations
The Population Variance (σ^2)
The Population Standard Deviation (σ)
Measures of Variability for Samples vs. Populations
4.8 Measures of Variability for Nominal Variables
The Index of Qualitative Variation
4.9 Measures of Variability: Drawing Conclusions

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

4.10 Looking Ahead
4.11 Summary
4.12 Important Terms
4.13 Formulas Introduced in This Chapter
Range
Interquartile Range
Variance (s^2) (Definitional Formula)
Variance (s^2) (Computational Formula)
Standard Deviation (s) (Definitional Formula)
Standard Deviation (s) (Computational Formula)
Population Variance (σ^2) (Definitional Formula)
Population Standard Deviation (σ) (Definitional Formula)
Index of Qualitative Variation (IQV)
4.14 Using SPSS
Calculating Measures of Central Tendency and Variability: The Frequency Rating Study (4.1)
4.15 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 5 Normal Distributions
5.1 Normal Distributions
What Are Normal Distributions?
Why Are Normal Distributions Important?
Is There Only One Normal Distribution?
5.2 The Standard Normal Distribution
What Is the Standard Normal Distribution?
Interpreting z-Scores: Distance From the Mean
Interpreting z-Scores: Relative Position Within the Distribution
Finding the Area Between the Mean and a z-Score

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Finding the Area Less or Greater Than a z-Score
Finding the Area Between Two z-Scores
5.3 Applying z-Scores to Normal Distributions
Transforming Normal Distributions Into the Standard Normal Distribution
Understanding the z-Score Formula
Example: Transforming SAT Scores Into z-Scores
Interpreting z-Scores: Distance From the Mean
Interpreting z-Scores: Relative Position Within Distribution
Finding the Area Between the Mean and a Score
Finding the Area Less or Greater Than a Score
Finding the Area Between Two Scores
Transforming Scores Into z-Scores: Drawing Conclusions
5.4 Standardizing Frequency Distributions
Example: Measuring Politicians' Behavior
Standardizing a Frequency Distribution
Interpreting Standardized Scores
Using Standardized Scores to Compare or Combine Variables
Characteristics of Standardized Distributions
When Can You Apply the Normal Curve Table to a Standardized Distribution?
5.5 Looking Ahead
5.6 Summary
5.7 Important Terms
5.8 Formulas Introduced in This Chapter
z-Score
Standardized Score
5.9 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 6 Probability and Introduction to Hypothesis Testing

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

6.1 A Brief Introduction to Probability
- What Is Probability?
- Why Is Probability Important to Researchers?
- Applying Probability to Normal Distributions
- Applying Probability to Binomial Distributions

6.2 Example: Making Heads or Tails of the Super Bowl

6.3 Introduction to Hypothesis Testing
- State the Null and Alternative Hypotheses (H_0 and H_1)
  - The Null Hypothesis (H_0)
  - The Alternative Hypothesis (H_1)
- Make a Decision About the Null Hypothesis
  - Defining a Low" Probability of a Statistic: Setting Alpha (α)
  - Identifying the Values of the Statistic With a Low Probability
  - Stating a Decision Rule
  - Calculating a Value of a Statistic
  - Making the Decision Whether to Reject the Null Hypothesis
- Draw a Conclusion From the Analysis
- Relate the Result of the Analysis to the Research Hypothesis
- Summary of Steps in Hypothesis Testing

6.4 Issues Related to Hypothesis Testing: An Introduction
- The Issue of Proof" in Hypothesis Testing
- Errors in Decision Making
- Factors Influencing the Decision About the Null Hypothesis
  - Sample Size
  - Alpha (α)
  - Directionality of the Alternative Hypothesis

6.5 Looking Ahead

6.6 Summary

6.7 Important Terms

6.8 Formula Introduced in This Chapter

Probability

6.9 Exercises

Answers to Learning Checks

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 7 Testing One Sample Mean
7.1 An Example From the Research: Do You Read Me?
7.2 The Sampling Distribution of the Mean
Characteristics of the Sampling Distribution of the Mean
7.3 Inferential Statistics: Testing One Sample Mean (σ Known)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: z-Test for One Mean
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the z-Test for One Mean
Summary
7.4 A Second Example From the Research: Unique Invulnerability
7.5 Introduction to the t-Distribution
Characteristics of the Student t-Distribution
7.6 Inferential Statistics: Testing One Sample Mean (σ Not Known)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: t-Test for One Mean
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the t-Test for One Mean

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Summary
7.7 Factors Affecting the Decision About the Null Hypothesis
Sample Size
Alpha (α)
Directionality of the Alternative Hypothesis
7.8 Looking Ahead
7.9 Summary
7.10 Important Terms
7.11 Formulas Introduced in This Chapter
z-Test for One Mean
Population Standard Error of the Mean $\left(\mathbf{i}t\sigma_{\mathbf{i}t\overline{X}}\right)$
Degrees of Freedom (df) for One Sample
t-Test for One Mean
Standard Error of the Mean $\left(\mathbf{i}ts_{\mathbf{i}t\overline{X}}\right)$
7.12 Using SPSS
Testing One Sample Mean (σ Not Known): The Unique Invulnerability Study (7.6)
7.13 Exercises
Answers to Learning Checks
Learning Check 2
Learning Check 3
Learning Check 5
Learning Check 6
Answers to Odd-Numbered Exercises
Chapter 8 Estimating the Mean of a Population
8.1 An Example From the Research: Salary Survey
8.2 Introduction to the Confidence Interval for the Mean
Using the Sampling Distribution of the Mean to Create the Confidence Interval for the Mean
8.3 The Confidence Interval for the Mean (σ Not Known)
State the Desired Level of Confidence
Calculate the Confidence Interval and Confidence Limits
Calculate the Standard Error of the Mean $\left(s_{\overline{X}}\right)$

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Calculate the Degrees of Freedom (df) and Identify the Critical Value of t

Calculate the Confidence Interval and Confidence Limits

Draw a Conclusion About the Confidence Interval

Probability of an Interval or the Probability of a Population Mean?

Summary

8.4 The Confidence Interval for the Mean (σ Known)

State the Desired Level of Confidence

Calculate the Confidence Interval and Confidence Limits

Calculate the Population Standard Error of the Mean $\left(\sigma_{\overline{X}}\right)$

Identify the Critical Value of z

Calculate the Confidence Interval and Confidence Limits

Draw a Conclusion About the Confidence Interval

Summary

8.5 Factors Affecting the Width of the Confidence Interval for the Mean

Sample Size

Estimating the Sample Size Needed for a Desired Confidence Interval

Level of Confidence

Choosing Between Different Levels of Confidence: Probability Versus Precision

8.6 Interval Estimation and Hypothesis Testing

Concerns About Hypothesis Testing

The Interpretation of Statistical Significance and Nonsignificance

The Influence of Sample Size on the Decision About the Null Hypothesis

Interval Estimation as an Alternative to Hypothesis Testing

8.7 Looking Ahead

8.8 Summary

8.9 Important Terms

8.10 Formulas Introduced in This Chapter

Confidence Interval for the Mean (σ Not Known)

Confidence Interval for the Mean (σ Known)

Estimated Sample Size for the Confidence Interval for the Mean

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

8.11 Using SPSS
Calculating the confidence interval for the mean (The salary survey study (8.1))
8.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 9 Testing the Difference Between Two Means
9.1 An Example From the Research: You Can Just Wait
9.2 The Sampling Distribution of the Difference
Characteristics of the Sampling Distribution of the Difference
9.3 Inferential Statistics: Testing the Difference Between Two Sample Means
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: t-Test for Independent Means
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the t-Test for Independent Means
Summary
9.4 Inferential Statistics: Testing the Difference Between Two Sample Means (Unequal Sample Sizes)
An Example From the Research: Kids' Motor Skills and Fitness
Inferential Statistics: Testing the Difference Between Two Sample Means (Unequal Sample Sizes)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Relate the Result of the Analysis to the Research Hypothesis

Assumptions of the t-Test for Independent Means (Unequal Sample Sizes)

Summary

9.5 Inferential Statistics: Testing the Difference Between Paired Means

Research Situations Appropriate for Within-Subjects Research Designs

An Example From the Research: Web-Based Family Interventions

Inferential Statistics: Testing the Difference Between Paired Means

Calculate the Difference Between the Paired Scores

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Draw a Conclusion From the Analysis

Relate the Result of the Analysis to the Research Hypothesis

Assumptions of the t-Test for Dependent Means

Summary

9.6 Looking Ahead

9.7 Summary

9.8 Important Terms

9.9 Formulas Introduced in This Chapter

Degrees of Freedom (df), Difference Between Two Sample Means

t-Test for Independent Means

Standard Error of the Difference $\left(s_{\overline{X}_1-\overline{X}_2}\right)$

Standard Error of the Difference (Unequal Sample Sizes)

Degrees of Freedom (df), Difference Between Paired Means

t-Test for Dependent Means

Standard Error of Difference Scores (s D)

9.10 Using SPSS

Testing the Difference Between Two Sample Means: The Parking Lot Study (9.1)

Testing the Difference Between Paired Means: The Web-Based Intervention Study (9.5)

9.11 Exercises

Answers to Learning Checks

Learning Check 1

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Answers to Odd-Numbered Exercises
Chapter 10 Errors in Hypothesis Testing, Statistical Power, and Effect Size
10.1 Hypothesis Testing vs. Criminal Trials
10.2 An Example From the Research: Truth or Consequences
Correct Decisions and Mistakes in Criminal Trials
10.3 Two Errors in Hypothesis Testing: Type I and Type II Error
Type I Error: The Risk in Rejecting the Null Hypothesis
What Is the Probability of Making a Type I Error?
What Is the Probability of Not Making a Type I Error?
Why Is Type I Error a Concern?
Type II Error: The Risk in Not Rejecting the Null Hypothesis
What Is the Probability of Making a Type II Error?
What Is the Probability of Not Making a Type II Error?
Why Is Type II Error a Concern?
Why Does Type II Error Occur?
Type I and Type II Error: Summary
10.4 Controlling Type I and Type II Error
Controlling Type I Error
Concerns About Controlling Type I Error
Controlling Type II Error
Increasing Sample Size
Raising Alpha (α)
Using a Directional Alternative Hypothesis (H1)
Increasing Between-Group Variability
Decreasing Within-Group Variability
Controlling Type I and Type II Error: Summary
10.5 Measures of Effect Size
Interpreting Inferential Statistics: The Influence of Sample Size
The Inappropriateness of “Highly Significant”

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Effect Size as Variance Accounted for

Measure of Effect Size for the Difference Between Two Sample Means: r^2

Interpreting Measures of Effect Size

Presenting Measures of Effect Size

Reasons for Calculating Measures of Effect Size

A Second Measure of Effect Size for the Difference Between Two Sample Means: Cohen's d

10.6 Looking Ahead

10.7 Summary

10.8 Important Terms

10.9 Formulas Introduced in This Chapter

Probability of Type I Error

Probability of Type II error

r Squared (r^2)

Cohen's d

10.10 Exercises

Answers to Learning Checks

Learning Check 1

Learning Check 2

Learning Check 3

Learning Check 4

Answers to Odd-Numbered Exercises

Chapter 11 One-Way Analysis of Variance (ANOVA)

11.1 An Example From the Research: It's Your Move

11.2 Introduction to Analysis of Variance (ANOVA)

Review: Testing the Difference Between the Means of Two Samples

Understanding Between-Group and Within-Group Variability

Relating Between-Group and Within-Group Variability

Introduction to Analysis of Variance (ANOVA) and the F-Ratio

Characteristics of the F-Ratio Distribution

11.3 Inferential Statistics: One-Way Analysis of Variance (ANOVA)

State the Null and Alternative Hypotheses (H_0 and H_1)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Value, and State a Decision Rule
Calculate a Statistic: F-Ratio for the One-Way ANOVA
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (R^2)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Summary
11.4 A Second Example: The Parking Lot Study Revisited
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Value, and State a Decision Rule
Calculate a Statistic: F-Ratio for the One-Way ANOVA
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (R^2)
Draw a Conclusion From the Analysis
The Relationship Between the t-Test and the F-Ratio
11.5 Analytical Comparisons Within the One-Way ANOVA
Planned Versus Unplanned Comparisons
Between-Group and Within-Group Variance in Analytical Comparisons
Conducting Planned Comparisons
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Conducting Unplanned Comparisons
Concerns Regarding Unplanned Comparisons
Methods of Controlling Familywise Error

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Controlling Familywise Error: The Tukey Test

11.6 Looking Ahead

11.7 Summary

11.8 Important Terms

11.9 Formulas Introduced in This Chapter

Degrees of Freedom for Between-Group Variance, One-Way ANOVA (df_{BG})

Degrees of Freedom for Within-Group Variance, One-Way ANOVA (df_{WG})

Between-Group Variance, One-Way ANOVA (MS_{BG})

Total Sample Mean, One-Way ANOVA $\left(\overline{X}_T\right)$

Within-Group Variance, One-Way ANOVA (MS_{WG})

F-Ratio, One-Way ANOVA (F)

Measure of Effect Size, One-Way ANOVA (R^2)

Relationship Between t-Test and F-Ratio for the One-Way ANOVA

F-Ratio, Analytical Comparison (F_{comp})

Between-Group Variance, Analytical Comparison $\left(MS_{\mathrm{BG}_{\mathrm{comp}}}\right)$

Measure of Effect Size, Analytical Comparison $\left(R^2_{\mathrm{comp}}\right)$

Critical Value, Tukey Test for Unplanned Comparisons (F_T)

11.10 Using SPSS

One-Way Analysis of Variance (ANOVA): The Chess Study (11.1)

Analytical Comparisons: Planned Comparisons—Explanation vs. Prediction (11.5)

11.11 Exercises

Answers to Learning Checks

Learning Check 2

Learning Check 3

Learning Check 4

Answers to Odd-Numbered Exercises

Chapter 12 Two-Way Analysis of Variance (ANOVA)

12.1 An Example From the Research: Vote—or Else!

12.2 Introduction to Factorial Research Designs

Testing Interaction Effects

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

An Example of Interaction Effect

An Example of No Interaction Effect

Testing Main Effects

The Relationship Between Main Effects and Interaction Effects

Factorial Designs vs. Single-Factor Designs: Advantages and Disadvantages

12.3 The Two-Factor (A × B) Research Design

Notational System for Components of the Two-Factor Research Design

Analyzing the Two-Factor Research Design: Descriptive Statistics

Notational System for Means in the Two-Factor Research Design

Examining the Descriptive Statistics

12.4 Introduction to Analysis of Variance (ANOVA) for the Two-Factor Research Design

F-Ratios in the Two-Factor Research Design

12.5 Inferential Statistics: Two-Way Analysis of Variance (ANOVA)

State the Null and Alternative Hypotheses (H_0 and H_1)

Main Effect of Factor A

Main Effect of Factor B

A × B Interaction Effect

Make Decisions About the Null Hypotheses

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Values, and State Decision Rules

Calculate Statistics: F-Ratios for the Two-Way ANOVA

Make Decisions Whether to Reject the Null Hypotheses

Determine the Levels of Significance

Calculate Measures of Effect Size (R^2)

Draw Conclusions From the Analyses

Relate the Results of the Analyses to the Research Hypothesis

12.6 Investigating a Significant A × B Interaction Effect: Analysis of Simple Effects

12.7 Looking Ahead

12.8 Summary

12.9 Important Terms

12.10 Formulas Introduced in This Chapter

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Degrees of Freedom for Main Effect of Factor A, Two-Way ANOVA (df_A)
Degrees of Freedom for Main Effect of Factor B, Two-Way ANOVA (df_B)
Degrees of Freedom for A × B Interaction Effect, Two-Way ANOVA ($df_{A\times B}$)
Degrees of Freedom for Within-Group Variance, Two-Way ANOVA (dfWG)
Between-Group Variance for Main Effect of Factor A, Two-Way ANOVA (MS_A)
Between-Group Variance for Main Effect of Factor B, Two-Way ANOVA (MS_B)
Sums of Squares for Between-Group Variance, Two-Way ANOVA (SS_{BG})
Between-Group Variance for A × B Interaction, Two-Way ANOVA ($MS_{A\times B}$)
Within-Group Variance, Two-Way ANOVA (MS_{WG})
F-Ratio for Main Effect of Factor A, Two-Way ANOVA (F_A)
F-Ratio for Main Effect of Factor B, Two-Way ANOVA (F_B)
F-Ratio for A × B Interaction, Two-Way ANOVA ($F_{A\times B}$)
Measure of Effect Size for Main Effect of Factor A, Two-Way ANOVA(R^2_A)
Measure of Effect Size for Main Effect of Factor B, Two-Way ANOVA(R^2_B)
Measure of Effect Size for A × B Interaction, Two-Way ANOVA(R^2_{AB})
12.11 Using SPSS
Two-Way Analysis of Variance (ANOVA): The Voting Message Study (12.1)
12.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 13 Correlation
13.1 An Example From the Research: It's Good for You!
13.2 Introduction to the Concept of Correlation
Describing the Relationship Between Variables
Nature of the Relationship
Direction of the Relationship
Strength of the Relationship

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Measuring the Relationship Between Variables: Correlational Statistics
Introduction to the Pearson Correlation Coefficient (r)
Calculating Correlational Statistics: The Role of Variance and Covariance
13.3 Inferential Statistics: Pearson Correlation Coefficient
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: Pearson Correlation Coefficient (r)
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (r^2)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the Pearson Correlation
A Computational Formula for the Pearson r
Represent the Covariance Between X and Y (SPXY)
Represent the Variance of Variable X (SSX)
Represent the Variance of Variable Y (SSY)
Calculate the Pearson Correlation Coefficient (r)
13.4 Correlating Two Sets of Ranks: The Spearman Rank-Order Correlation
Introduction to Ranked Variables
An Example From the Research: Do Students and Teachers Think Alike?
Inferential Statistics: The Spearman Rank-Order Correlation Coefficient
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
13.5 Correlational Statistics vs. Correlational Research
Does ANOVA = Causality?
“Correlation Is Not Causality”—What Does That Mean?
13.6 Looking Ahead

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

13.7 Summary
13.8 Important Terms
13.9 Formulas Introduced in This Chapter
Degrees of Freedom for Pearson Correlation (df)
Sum of Products (Definitional Formula) (SPXY)
Sum of Squares for the X Variable (Definitional Formula) (SSX)
Sum of Squares for the Y Variable (Definitional Formula) (SSY)
Pearson Correlation (Definitional Formula) (r)
Measure of Effect Size for the Pearson Correlation Coefficient (r^2)
Sum of Products (Computational Formula) (SPXY)
Sum of Squares for the X Variable (Computational Formula) (SSX)
Sum of Squares for the Y Variable (Computational Formula) (SSY)
Pearson Correlation (Computational Formula) (r)
Spearman Rank-Order Correlation (rs)
Measure of Effect Size for the Spearman Rank-Order Correlation $\left(r_s^2\right)$
13.10 Using SPSS
Pearson Correlation: The Picky Eater Study (13.1)
Spearman Rank-Order Correlation: The Business Student Study (13.4)
13.11 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Answers to Odd-Numbered Exercises
Chapter 14 Linear Regression and Multiple Correlation
14.1 Predicting One Variable From Another: Linear Regression
The Linear Regression Equation
Calculating the Linear Regression Equation
Calculate the Slope of the Equation (b)
Calculate the Y-Intercept of the Equation (a)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Report the Linear Regression Equation

Drawing the Linear Regression Equation

A Second Example: Put Down Your Pencils

Calculating the Linear Regression Equation

Drawing the Linear Regression Equation

14.2 Correlation With Two or More Predictors: Introduction to Multiple Correlation and Regression

The Limitations of Using Only One Predictor

Introduction to Multiple Correlation

Multiple Correlation With Two Predictors: An Example

Measuring the Relationship Between Two Predictors and a Criterion: The Multiple Correlation (R)

Testing the Multiple Correlation for Statistical Significance

Predicting a Criterion From Two Predictors: The Multiple Regression Equation

Conclusion

14.3 Looking Ahead

14.4 Summary

14.5 Important Terms

14.6 Formulas Introduced in This Chapter

Linear Regression Equation

Estimate of Slope (b)

Estimate of Y-Intercept (a)

Multiple Correlation Coefficient (R)

F-Ratio for the Multiple Correlation Coefficient (F)

Multiple Regression Equation

Regression Coefficient

Estimate of Y-Intercept (a) (Multiple Regression)

14.7 Using SPSS

Linear Regression: The Picky Eater Study (14.1)

Multiple Regression: The College Admission Example (14.2)

14.8 Exercises

Answers to Learning Checks

Learning Check 1

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 2

Answers to Odd-Numbered Exercises

Chapter 15 Chi-Square

15.1 An Example From the Research (One Categorical Variable): Are You My Type?

15.2 Introduction to the Chi-Square Statistic

Observed and Expected Frequencies: The Basis of the Chi-Square Statistic

Characteristics of the Distribution of Chi-Square Statistics

15.3 Inferential Statistic: Chi-Square Goodness-of-Fit Test

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Value, and State a Decision Rule

Calculate a Statistic: Chi-Square (χ^2)

Make a Decision Whether to Reject the Null Hypothesis

Determine the Level of Significance

Calculate a Measure of Effect Size (Cramér's φ)

Draw a Conclusion From the Analysis

Relate the Result of the Analysis to the Research Hypothesis

Assumptions Underlying the Chi-Square Statistic

Conducting the Chi-Square Goodness-of-Fit Test With Unequal Hypothesized Proportions

Calculate Expected Frequencies (f_e)

Calculate the Chi-Square Statistic (χ^2)

15.4 An Example From the Research (Two Categorical Variables): Seeing Red

15.5 Inferential Statistic: Chi-Square Test of Independence

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Value, and State a Decision Rule

Calculate a Statistic: Chi-square (χ^2)

Make a Decision Whether to Reject the Null Hypothesis

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Determine the Level of Significance
Calculate a Measure of Effect Size (Cramér's φ)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
15.6 Parametric and Nonparametric Statistical Tests
Parametric vs. Nonparametric Statistical Tests
Reasons for Using Nonparametric Statistical Tests
Examples of Nonparametric Statistical Tests
15.7 Looking Ahead
15.8 Summary
15.9 Important Terms
15.10 Formulas Introduced in This Chapter
Degrees of Freedom for the Chi-Square Goodness-of-Fit Test
Expected Frequency for the Chi-Square Goodness-of-Fit Test
Chi-Square Statistic (χ^2)
Cramér's φ for the Chi-Square Goodness-of-Fit Test
Degrees of Freedom for the Chi-Square Test of Independence
Expected Frequency for the Chi-Square Test of Independence
Cramér's φ for the Chi-Square Test of Independence
15.11 Using SPSS
Chi-Square Goodness-of-Fit Test—Equal Hypothesized Proportions: The Personality Study (15.1)
Chi-Square Goodness-of-Fit Test—Unequal Hypothesized Proportions: The Personality Study (15.3)
Chi-Square Test of Independence: The Color Preference Study (15.4)
Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Learning Check 6

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Odd-Numbered Exercises

Appendix A: Tables

Appendix B: Review of Basic Mathematics

I. Symbols

II. Terms

III. Addition, Subtraction, Multiplication, Division

A. All Positive Numbers

B. Adding All Negative Numbers

C. Adding a Combination of Positive and Negative Numbers

D. Subtracting Negative Numbers

E. Multiplying Two Negative Numbers

F. Multiplying a Combination of Positive and Negative Numbers

G. Dividing Two Negative Numbers

H. Dividing One Positive Number and One Negative Number

IV. Order of Operations

V. Fractions, Decimals, Percentages

A. Fraction

B. Decimal

C. Percentage

VI. Solving Simple Equations With One Unknown (X)

A. X Has a Value Subtracted From It (Solve for X Using Addition)

B. X Has a Value Added to It (Solve for X Using Subtraction)

C. X Is Multiplied by a Value (Solve for X Using Division)

D. X Is Divided by a Value (Solve for X Using Multiplication)

VII. Solving More Complex Equations With One Unknown (X)

A. X Is Multiplied by a Value and Then Has a Value Either Added to It or Subtracted From It (Solve for X Using Either Subtraction or Addition and Then Division)

B. X Is Divided by a Value and Then Has a Value Either Added to It or Subtracted From It (Solve for X Using Either Addition or Subtraction and Then Multiplication)

C. X Has a Value Added to It or Subtracted From It and Is Then Multiplied by a Value (Solve for X Using Division and Then Either Subtraction or Addition)

D. X Has a Value Either Added to It or Subtracted From It and Is Then Divided by a Value (Solve for X Using Multiplication and Then Either Subtraction or

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Addition)
Glossary
References
Index

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

PREFACE

A Few (?) Words About Me

When I entered college, I thought my path would take me to law school; however, taking an Intro to Psych class made me realize I enjoy the challenge of trying to understand the human mind. The college I attended, U.C. Santa Cruz, was a fairly unconventional university, but somehow in the midst of a sea of humanistic psychologists I became attracted to the empirical and methodological aspects of psychology by taking classes from instructors such as David Harrington and Dane Archer, who showed me that statistics could appeal to students if taught with a gentle, guiding approach that addresses questions relevant to students' lives. After graduating from college, I was able to get a job as a research assistant for a human resource consulting firm. Despite my lack of work experience, I was hired primarily as a result of having taken statistics and research methods courses, which taught me that learning statistics has benefits both inside and outside of the classroom.

Several years later, I started grad school at U.C. Berkeley, where two events critical to this book took place. First, serving as a teaching assistant, I found I really enjoyed helping students, particularly in statistics and research methods classes often approached with fear and suspicion. Second, I took graduate classes from Geoff Keppel, who had developed his own unique system for analyzing experimental research designs. His lectures and books showed me that statistics can be taught in a systematic way that highlights similarities rather than differences between different research situations. Geoff managed to transform something as daunting-sounding as a "3 x 2 x 4 research design" into the mathematical equivalent of playing with wooden toy alphabet blocks labeled "A," "B," and "C."

After leaving Cal, I was hired by San Jose State, where I teach an introductory statistics course aimed at students with a wide range of backgrounds, ability, and motivation. When I went to select a textbook for the course, for the first time I looked carefully at the wide range of offerings. What I found striking (and still find striking) was that the majority of books focused on providing formulas and very small sets of data designed to demonstrate how to correctly calculate the correct numbers from these formulas. Little emphasis, however, was given to what these numbers meant or implied. Given my own experiences learning and using statistics, I thought a book was needed that focusses on how statistics may be used to answer questions and test ideas within the larger research process.

Goal of this Book

This textbook, aimed primarily at students majoring in the behavioral and social sciences, includes many of the same topics and chapters as the majority of existing introductory statistics books. However, what I believe distinguishes this book is its thematically-based approach that places statistics within the research process. The primary purpose of this

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

book is to not just teach students how to calculate statistics but how to interpret the results of statistical analyses in light of a study's research hypothesis, and to communicate one's results and interpretations to a broader audience. Hopefully, this book will not only help students understand the purpose and use of statistics, but it will also give them a greater appreciation of how research studies are conceived, conducted, and communicated.

The book is intended for students with a limited knowledge of mathematics, statistics, and research design and methodology. Some common obstacles faced by students are their math anxiety, fear of conducting arithmetic calculations, and an inability to see the relevance of the material to their daily lives. To address these concerns, this books attempts to judiciously use mathematical terminology, write in a conversational manner, and utilize actual research situations rather than contrived sets of numbers.

Organization of the Book

The 15 chapters of this book may be placed into three general categories. Chapters 1 to 4 are designed to introduce students to the research process and how data that have been collected may be organized, presented, and summarized:

Chapter 1 (Introduction to Statistics) defines the concept of statistics, explains why statistics are important to learn, and describes the steps in the research process that are the framework for the rest of the book.

Chapter 2 (Examining Data: Tables and Figures) discusses the importance of examining data that have been collected and describes how tables and figures may be created to represent data.

Chapter 3 (Measures of Central Tendency) presents measures of central tendency (the mode, median, and the mean) and compares them in terms of their relative strengths and weaknesses.

Chapter 4 (Measures of Variability) presents measures of variability (range, interquartile range, variance, standard deviation, index of qualitative variation) and the concept of parameters (population variance and standard deviation).

Chapters 5 through 10 discuss the process of conducting statistical analyses to test research questions and hypotheses, as well as issues and controversies regarding this process:

Chapter 5 (Normal Distributions) introduces theoretical distributions, starting with normal distributions, before moving on to the standard normal distribution (normal curve).

Chapter 6 (Probability and Introduction to Hypothesis Testing) presents the structure for later chapters by introducing, defining, and illustrating steps in hypothesis testing.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Chapter 7 (Testing One Sample Mean) tests one sample mean using the *z* -test and *t* -test and introduces sampling distributions, the central limit theorem, and the Student *t* distribution.

Chapter 8 (Estimating the Mean of a Population) introduces interval estimation by developing the confidence interval for the mean and discusses interval estimation as an alternative to null hypothesis significance testing (NHST).

Chapter 9 (Testing the Difference Between Two Means) covers the *t* -test for independent means for both equal and unequal sample size and introduces within-subjects research designs using the *t* -test for dependent means.

Chapter 10 (Errors in Hypothesis Testing, Statistical Power, and Effect Size) discusses Type I and Type II error and introduces measures of effect size as a way of drawing appropriate conclusions about an analysis.

Chapters 11 to 15 discuss different statistical procedures used in research situations that vary in the number of independent variables as well as how the variables have been measured:

Chapter 11 (One-way Analysis of Variance (ANOVA)) discusses the one-way ANOVA, the relationship between the *t* -statistic and the F-ratio, and planned and post-hoc comparisons.

Chapter 12 (Two-way Analysis of Variance (ANOVA)) introduces factorial research designs, the calculation of F-ratios for main effects and interaction effects, and illustrates the use of simple effects analyses to test research hypotheses.

Chapter 13 (Correlation) defines and illustrates the concept of correlation, discusses the Pearson and Spearman correlation coefficients, and makes the distinction between correlational statistics and correlational research methods.

Chapter 14 (Linear Regression and Multiple Correlation) introduces the concept of regression, discusses linear regression, and works through an example of multiple regression with two predictor variables.

Chapter 15 (Chi-Square) describes the analysis of categorical variables using the chi-square goodness-of-fit test and the chi-square test of independence before ending the book with a brief introduction to nonparametric statistics.

In most cases, the chapters in this book share a uniform format and structure, centered on the research process, based on examples of actual published research studies addressing topics relevant to students' lives. For each example, the research hypothesis is clearly identified; after a brief description of how the data in the study were collected, the calculation and interpretation of both descriptive and inferential statistics are presented,

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ending with the extent to which the results of the statistical analyses support the research hypothesis and how these findings may be communicated to a broader audience.

Pedagogical Features

The chapters in this textbook contain several reoccurring features designed to enhance students' comprehension of the material. These include the following:

Periodic learning checks within chapters. Short learning checks, in the form of conceptual review questions and computational exercises, are provided throughout each chapter.

Explicit presentation of formulas and calculations. This book provides both definitional formulas to highlight their relation to the concept being measured and computational formulas to facilitate calculations. In applying these formulas, all of the steps in the calculation of statistics such as the *t* -test and F-ratio are included to help students in completing exercise and homework assignments.

Illustration of the presentation and communication of analyses. Every chapter includes a discussion and illustration of how statistical analyses may be reported as well how they may be communicated to a larger audience in both written and visual form.

Highlighted definition of terms. Every term and concept introduced in this book is boldfaced, immediately followed by an explicit definition.

End of chapter summaries and exercises. The end of each chapter contains a summary of critical points, a list of terms and formulas introduced in the chapter, and a large number of exercises, many of which refer to published research studies.

Calculation of statistics using SPSS. Many chapters include end-of-chapter content that provides an annotated presentation of how statistics may be calculated using the SPSS software. Screen shots of data files, menu selections, and output files are included and annotated to locate and identify relevant aspects of the analysis.

What's New to the 3rd Edition

I'd like to start by expressing my most heartfelt appreciation to the instructors and reviewers of the first two editions of this book as their evaluation, insights, criticisms, and suggestions were absolutely vital to the changes made in this third edition. I'd like to particularly thank the instructors who've adopted the book who were kind enough to spontaneously send me their thoughts and suggestions. It is my sincere hope that all of you see the incorporation of your input in this edition.

There were two main goals in this edition. The first goal was to make the writing as brief and concise as possible, manifested in the goal of reducing the length of each chapter by 5% to

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

10%. This goal was achieved by simplifying the language and shortening or eliminating what I, the book's adopters, and the book's reviewers felt was unnecessary detail or discussion. The second goal was to increase and update the exercises at the end of many of the chapters, including the use of recent published research as the basis of these exercises. Another goal was to update the demonstration of the use of SPSS to analyze data using the most recent version (Version 31 at the time of this book's publication).

More specifically, in terms of content, the following changes were made to these chapters:

Chapter 1 (Introduction to Statistics): The published research studies used to illustrate concepts in the text, learning checks, and exercises were added to and updated using topics relevant to students (e.g., fear of missing out (FOMO), cryptocurrency, legalization of cannabis, Gen Z, COVID, fake news).

Chapter 6 (Probability and Introduction to Hypothesis Testing): The Super Bowl example used to introduce and illustrate the process of hypothesis testing was updated.

Chapter 8 (Estimating the Mean of a Population): The main research example, which involves data from an actual national salary survey, has been updated throughout this chapter with more current salary data.

As with the first two editions, any and all questions, concerns, suggestions, and complaints you have about the book are most welcome and can be sent to me at howard.tokunaga@sjsu.edu .

A Few Tips for Students

To you, the college student about to read this book as part of taking a statistics course: "Welcome!" and "Great job!" We welcome you because you're about to embark on a semester-long journey that we hope enhances your skills and widens your perspective; we congratulate you because it's a journey not everyone is willing to take. We say to our students that by taking just one statistics course they are ahead of 95% of the population in their knowledge of statistics.

We also tell our students that our primary goal for them is not to become a statistics genius but just to become comfortable with the process of analyzing and interpreting data. In that regard, we agree with a 2012 *Forbes* magazine article entitled *The Six Classes That Will Make Any College Grad Employable* that noted "a good statistics class will help the student become comfortable pushing numbers around and avoiding false interpretations."

At the present moment, we know our encouragement may be of little comfort because you might be anxious about having to learn statistics. Some of you might be anxious about having to learn statistical concepts and formulas; others might be anxious because the research process seems pretty complicated. Talking with students who have taken our courses over the years has helped us assemble the following advice:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Master the Material Presented in the Early Chapters

Chapters 1 through 6 discuss how research is typically conducted, how data are summarized and described, and the process by which researchers conduct statistical analyses to test their ideas. It is absolutely critical for you to have a firm understanding of these chapters as they lay the groundwork for later chapters that discuss a variety of statistical procedures used by researchers to test hypotheses.

What does it mean to master this material? First, *read the chapters both before and after they're discussed in class* . Reading the chapters beforehand helps you identify anything that's unclear and have your questions addressed by your instructor; re-reading the chapters afterward helps confirm your understanding of the material. Next, *be able to define and explain key conceptspresented in these chapters* . These concepts are highlighted in bold-faced letters, and it's important to learn them when they're first introduced because they'll appear throughout the remainder of the book. Next, *do the learning checks within the chapters and the exercises at the end of the chapters* , because they include exercises and review questions you can ask yourself to assess your understanding of the material. Most importantly, *do not miss class during the early part of the semester* ! It's been our experience that students who miss critical lectures at the beginning of the term often have difficulty keeping up as the semester continues.

Review High School Algebra

If you happen to move beyond the introductory statistics course for which you're reading this book, you'll find that statistics are often calculated using computers and statistical software rather than hand calculations. Consequently, some statistics textbooks no longer include mathematical formulas but instead have students use statistical software or Microsoft Excel. However, this book takes a different approach for two main reasons. First, we believe that in order to understand the results of statistical analyses, one must understand their underlying foundation. Learning statistics via computer software can lead to a "brain-dead" approach in which students are at the mercy (rather than control) of their computers. We refer to this as the "I'm only as smart as my printout" method of learning statistics, also known as "Because this is what my computer told me."

Second, this book contains mathematical formulas and calculations for one simple reason: The statistics in this book are not hard to calculate. To assess whether you're prepared to read this book, take the following test:

1. Do you know how to add, subtract, multiply, and divide?
2. Do you know how to calculate the square or square root of a number?
3. Do you understand the order of operations and parentheses within mathematical formulas?

If your answers to these questions are "Yes," congratulations! You possess the ability to conduct *every* mathematical calculation in this book because *none* of the formulas involve geometry, trigonometry, or calculus. However, if you're unsure of your mathematical ability, we suggest you review your high school algebra. Also, there's an appendix at the end of this

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

book that includes the mathematical concepts and operations needed to conduct the statistical analyses in this book. We highly recommend you read this appendix to review and assess your mathematical skills.

Learn to Use a Statistical Calculator

Although learning the statistics in this book doesn't require the use of computers or software, you will need a hand calculator. Calculators with statistical capabilities are typically identified as "statistical" or "scientific" calculators. Although there's a wide selection of brands and models, know that the calculators most appropriate for this book cost only $10 to $15. We do *not* recommend you use or purchase a more expensive calculator! We prefer these simpler calculators because students are able to do their calculations more quickly and with fewer errors. We do suggest you purchase a calculator similar to the one your instructor uses as they will be more familiar with its features and idiosyncrasies.

In selecting a calculator, look closely at the keyboard; you'll need a calculator whose keys have labels such as $\bar{X}$, σ, and ΣX (Chapters 3 and 4 discuss what these symbols represent). When you go calculator shopping, you'll find calculators that have keys with labels such as ΣY and XY. Although calculators with these Y symbols are only slightly more expensive than calculators with just the X symbols, we do *not* recommend their purchase because these features are not needed to perform the calculations in this book. In fact, we've found students with Y calculators have a more difficult time performing simpler calculations.

Ask Questions

Conducting research is the process of asking and answering questions. Accordingly, throughout this book issues are often framed in the form of questions, which we hope encourages you to ask questions as well. The first person you should direct any questions to is yourself. As you work your way through the chapters, get into the habit of asking yourself, "What does this mean?" A good test of your level of comprehension is whether you can explain the material to yourself in a meaningful way.

You should also be sure to work with your instructor to confirm your level of understanding and clarify unanswered questions. Instructors often say to students, "The only bad question is the one you don't ask." Don't be hesitant about asking questions in class. Instructors will tell you they use one student's question to assess an entire class's level of understanding. This is what we call the "pencil test"—when one student asks a question, we look to see how many other pencils get raised, which reflects the number of students who had the same question. Asking questions enhances the learning experience for yourself, your classmates, and your instructor.

Form a Study Group

No matter how much we encourage students to ask questions in class, it seems they hesitate for fear of drawing attention to themselves. As a result, we turn to another well-

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

traveled saying: “Misery loves company.” Forming a study group with other students will help you keep up with reading assignments, confirm your comprehension of the material, clarify any confusion you have, prepare for exams, and provide social support. We do suggest that your group contain at least one person who will raise the group’s questions and concerns to the instructor during class or office hours.

Practice, Practice, Practice

The best way to learn the topics covered by this book is to obtain as much practice as possible. Throughout the chapters are examples in which the calculations have been worked out for you; in the middle and end of each chapter are a number of exercises. The answers to some of the exercises are provided at the back of this book, and your instructor has access to the answers to the other exercises. Do the end-of-chapter exercises even if your instructor does not assign them as homework in order to identify any reoccurring mistakes you make. We find that the vast majority of errors students make are simple computational errors, committed when students do the calculations too quickly.

Also note that the publisher of this book, Sage Publications, has developed a webpage for this book that contains a huge amount of supplemental tools such as videos and quizzes for each chapter. Find it by going to the Sage Publishing website **collegepublishing.sagepub.com** , searching for my name (Tokunaga), and looking for the **Resources** tab.

Enough cheerleading and preaching – let’s get started!

WELCOME!

GREAT JOB!

Additional Teaching Resources

Visit **collegepublishing.sagepub.com** , and navigate to the Resources tab on your book’s page to find the teaching materials designed to accompany this textbook. On this site you will find an array of materials that will save you time and to help you keep students engaged, including the following:

Learning management system cartridges that easily integrate with your course management system so student test results and graded assignments seamlessly flow into your gradebook

Test banks , aligned to Bloom’s Taxonomy, that provide a diverse range of test items, including multiple choice, true/false, and essay questions

Lecture notes that provide an outline and the key concepts in each chapter to aid in lecture preparation

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

PowerPoint ® slides that offer a flexible, accessible, and customizable solution for creating multimedia lectures

Tables and Figures from the book to support lecture preparation and class discussions

Sample course syllabi , including suggested models for structuring your course

Answers to even-numbered questions from the text help facilitate grading

SPSS datasets for use with exercises from the text.

Multimedia content , including original SAGE videos featuring tutorials with author Howard T. Tokunaga that bring concepts to life and appeal to diverse learners

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ACKNOWLEDGMENTS

This book has traveled a long journey, with many people playing both transient and persevering roles in its development. First and foremost, I'd like to thank my kids, Meagan and Will, for providing me perspective and inspiration amid all of the twists and turns we faced while this book was being written. I'm so very blessed to have you as my "burden." Away from home, I'm grateful to my graduate students at San Jose State—working on this book has given me much empathy for their thesis-related travails. However, they should be aware that the completion of this book ends the unspoken agreement that I not ask, "So how's your thesis?" if they don't ask, "So how's your book?" A number of friends have provided most welcome relief after long days staring at a computer screen, with particular thanks going to Ellie, Eileen, Susan, Bonnie, Ann, Lisa, and Terry. I would like to thank Charles Linsmeier and David Shirley for their many helpful comments and advice on earlier versions of this book, Julie West for her help in developing chapter exercises, and Youn-Jeong Choi and Naomi Ran for their careful examination of the chapters and exercises.

I would like to express my appreciation to everyone at SAGE who has played a role in shepherding this book to completion. First and foremost, I must express my gratitude to my editor, Vicki Knight. It has been a joy and an honor to work with someone with such a truly exceptional combination of experience, knowledge, perspective, and humor. In production I was fortunate enough to work with others on the SAGE team who made the process a pleasure: Yvonne McDuffee, editorial assistant; Katie Guarino, assistant editor; Nicole Elliott, executive marketing manager; Candice Harman, cover designer; and the copyeditor, Michelle Ponce. Furthermore, I am grateful to the reviewers who took the time to read the chapters, raise questions, and provide constructive comments and suggestions to assist me in clarifying and strengthening the manuscript. The following are the reviewers for third edition:

Erin M. Fekete, University of Indianapolis

Brooke Magnus, Boston College

Loretta Neal McGregor; Arkansas State University

Michael T. Moore, Adelphi University

Megan M. Ringel, St. Edward's University

Bettina Viereck, University of Hartford

Janice Weaver, Community College of Allegheny County

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ABOUT THE AUTHOR

Howard T. Tokunaga is professor of Psychology at San Jose State University, where he serves as coordinator of the MS Program in Industrial/Organizational (I/O) Psychology and teaches undergraduate and graduate courses in statistics, research methods, and I/O psychology. He received his bachelor's degree in psychology at U.C. Santa Cruz and his PhD in psychology at U.C. Berkeley. In addition to his teaching, he has consulted with a number of public sector and private sector organizations on a wide variety of management and human resource issues. He is author of *Moving from IBM SPSS to R and RStudio: A Statistics Companion* , and coauthor (with G. Keppel) of *Introduction to Design and Analysis* : *A Student's Handbook* . In his spare time, he enjoys the outdoors, museums, live theater, and being as physically, socially, and politically active as possible.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CHAPTER 1 INTRODUCTION TO STATISTICS

Learning Objectives

1.1 Define the concept of statistics.

1.2 Describe different reasons for learning statistics.

1.3 Describe the main steps in the research process and the tasks and concepts within each step.

In introducing this book to you, we assume you're a college student taking what is perhaps your first course in statistics to fulfill a requirement for your major or a general education requirement. If so, you may be asking yourself two questions:

What is statistics?

Why learn statistics?

The ultimate goal of this book is to help you begin to answer these two questions.

1.1 What Is Statistics?

You may not know it, but you encounter a variety of "statistics" in your day-to-day activities: the typical cost of going to college, the yearly income of the average college graduate, the average price of a car, and so on. So what exactly is "statistics"?

The Merriam-Webster dictionary defines **statistics** as *a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data* . When most people think about statistics, they focus on the "analysis" aspect of this definition—that is to say, they focus on numbers that result from analyzing data. However, statistics not only is concerned about how data are *analyzed* but also how data are *collected* and how the results of statistical analyses are *interpreted* and *communicated* . The main purpose of this book is to introduce, describe, and illustrate the role of statistics within the larger research process.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1.2 Why Learn Statistics?

There are many reasons why you should learn statistics. First, not only do you currently encounter statistics in your daily activities, but *throughout your life you are affected by the results of research and statistical analyses* . Which college you attend is based in part on test scores developed by researchers. You may have to take a personality or intelligence test to get a job. The choices of drugs and medicines available to you are based on medical research. Learning about statistics will help you become a more informed and aware consumer of research and statistical analyses that affect many aspects of your life.

A second reason for learning statistics is that *you may be required to interpret the results of statistical analyses* . Many college courses require students to read academic research journal articles. Evaluating published research is complicated by the fact that different people studying the same topic may come up with diverse or even opposing conclusions. Understanding statistics and their role in the research process will help you decide whether conclusions drawn in research articles are appropriate and justified.

Another reason for learning statistics is that *you might be required to do your own research* . College courses sometimes have students design and conduct mini-research studies; undergraduate majors might encourage students to do senior honors theses; graduate programs often require masters' theses and doctoral dissertations. Learning to collect and analyze data will help you address your own questions in an objective, systematic manner.

A final reason for learning statistics is that it may *help you in your future career* . The website Careercast.com conducts an annual survey in which they evaluate jobs on four dimensions: environmental factors, income, employment outlook, and stress. Recently, their #1 rated job was "data scientist," and the #2 rated job was "statistician." They noted, "A statistician's skill set can be used to break down and analyze large quantities of data. The demand for these skills spans a variety of industries, including marketing, banking, government, sports, retail, and even healthcare."

It's generally a good idea for students to maintain a healthy level of curiosity or even skepticism in regard to their education. However, we find that when it comes to learning statistics, students' feelings may be characterized as fear or anxiety. We understand these feelings, and we hope the benefits of learning statistics will become clear to you and help overcome any concerns you have.

1.3 Introduction to the Stages of the Research Process

Much of scientific research involves asking questions. Throughout this book, we'll examine a broad range of questions regarding human attitudes and behavior that contemporary researchers have asked and attempted to answer. The following are research questions we'll address in this chapter to introduce the stages of the research process:

> Does appealing to one's fear of missing out (FOMO) affect the intention to invest in cryptocurrencies?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Do pro-cannabis messages on social media lead to cannabis use?

Do parents differ in their awareness of their children's online activities?

Do nonbinary college students have the same experiences as cis men or cis women?

How might you try to answer questions such as these? You could base your answers on your personal beliefs, or you could adopt the answers given to you by others. But rather than relying on subjective beliefs and feelings, researchers in many academic disciplines test their ideas using the scientific method. The **scientific method** is a method of investigation that uses the objective and systematic collection and analysis of empirical data to test theories and hypotheses.

At its simplest, this book will portray the scientific method as consisting of five main steps or phases:

developing a research hypothesis to be tested,

collecting data,

analyzing the data,

drawing a conclusion regarding the research hypothesis, and

communicating the findings of the study.

Accomplishing each of these five steps requires completing a number of tasks, as shown in Figure 1.1 . Because this sequence of steps serves as the model for the research process we'll discuss throughout this book, each step is briefly introduced in the following sections. It's important to understand that the research process depicted in Figure 1.1 represents an ideal way of doing research. The "real" way, as you may discover in your own efforts, is often anything but a smooth ride but rather is filled with false starts, dead ends, and wrong turns.

Developing a Research Hypothesis to Be Tested

The initial stage—and the first step—of the research process is to develop a research hypothesis to be tested. A **research hypothesis** may be defined as a statement regarding an expected or predicted relationship between variables. A **variable** is a property or characteristic of an object, event, or person that can take on different values. One example of a variable is "U.S. state," a variable with 50 possible values (Alabama, Arkansas, etc.).

Research hypotheses are usually developed through the completion of several tasks:

identifying a question or issue to be examined,

reviewing and evaluating relevant theories and research, and

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

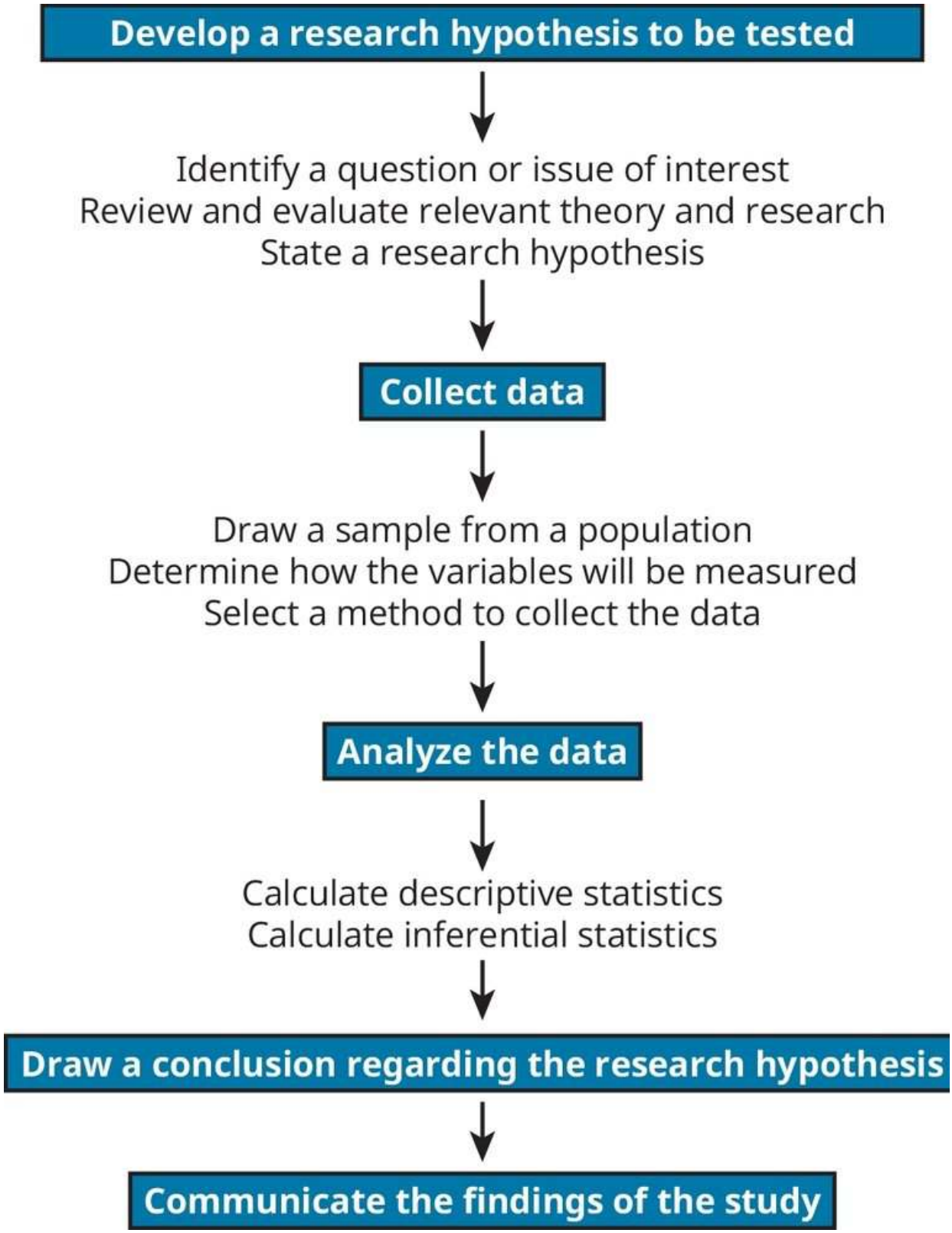
stating a research hypothesis.

Identifying a Question or Issue to Be Examined

Most research starts with a question posed by the researcher that often comes from the researcher's own ideas and daily observations. Although this may not seem terribly scientific, there's an advantage in using one's own experience as a starting point: People are generally much more motivated to explore issues that concern them personally. In teaching statistics, we advise students developing their own research projects to study something of interest to them. Conducting research can be tedious, difficult, and frustrating. At various points, you may ask yourself, "Why am I doing this?" Being able to provide a satisfactory answer to this question will help you overcome obstacles you encounter along the way.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

▶ Long Description: Figure 1.1 Steps in the Research Process Within the Scientific Method

Figure 1.1 Steps in the Research Process Within the Scientific Method

Reviewing and Evaluating Relevant Theories and Research

Beyond your own curiosity, research questions often arise from studying the theories, ideas, and research of others. A **theory** is a set of propositions used to describe or explain a phenomenon. The purpose of a theory is to summarize and explain specific facts using a logically consistent framework. Placing a question within a theoretical framework provides guidance and structure to research.

Reviewing and evaluating existing theories and research helps you decide whether it's worth the time and energy required to conduct a study. By seeing what others have done, you may conclude that the current way of thinking on a topic is incomplete or mistaken. By doing this review, you're able to ensure that the research you conduct adds to and improves upon an existing body of knowledge.

Stating a Research Hypothesis: Independent and Dependent Variables

Understanding and evaluating an existing literature not only helps develop a question of interest, but it may also lead to a predicted answer to that question. Within the scientific method, this answer is stated as a research hypothesis, defined earlier as a statement regarding an expected or predicted relationship between variables. Table 1.1 lists the research questions and research hypotheses for the studies listed at the beginning of this section. For example, the first research hypothesis states that "consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos."

One characteristic of research hypotheses such as those listed in Table 1.1 is that they identify the variables that are the focus of their research. As mentioned earlier, a variable is a property or characteristic with different values. Variables can be classified in several ways. In specifying a research hypothesis, researchers often speak in terms of "independent" and "dependent" variables. An **independent variable** may be defined as a variable manipulated or controlled by the researcher. A **dependent variable** , on the other hand, is a variable measured by the researcher that is expected to change or vary as a function of the independent variable. The independent variable is seen as the "cause" or "treatment," whereas the dependent variable is the "effect" or "outcome."

Researchers are interested in examining the effect of the independent variable on the dependent variable. Consider the first research hypothesis provided in Table 1.1 : "Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos." Here, the independent variable is the type of appeal consumers are exposed to: FOMO or non-FOMO. The dependent variable is the intention to invest in cryptos. In this study, the effect of the independent variable on the dependent variable is that

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

differences in the intention to invest in cryptos (the dependent variable) may “depend” upon which type of appeal (FOMO or non-FOMO) a consumer receives. Table 1.2 lists the independent and dependent variables for each of the research hypotheses in Table 1.1 .

In addition to identifying variables, a second characteristic of research hypotheses is that they specify the nature and direction of the relationship between variables. For example, the first research hypothesis in Table 1.1 states that “Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *higher* intention to invest in cryptos.” The word *higher* indicates the nature and direction of the relationship between the independent variable (type of appeal) and the dependent variable (intention to invest in cryptos). The direction of the relationship would not have been stated if the hypothesis had included the less specific phrase “will have a *different* intention,” which simply indicates that the two groups of consumers are not expected to be the same. Table 1.3 provides directional and nondirectional research hypotheses for the research studies from Table 1.1 .

Table 1.1 Examples of Research Questions and Hypotheses

Research Question	Research Hypothesis
Does appealing to one’s fear of missing out (FOMO) affect the intention to invest in cryptocurrencies? (Friederich et at., 2024)	Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.
Do pro-cannabis messages on social media lead to cannabis use? (Willoughby et al., 2024)	Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.
Do parents differ in their awareness of their children’s online activities? (Symons et al., 2017)	Mothers have more accurate knowledge of their children’s online activities than do fathers.
Do nonbinary college students have the same experiences as cis men or cis women? (Marx et al., 2024)	Nonbinary college students experience higher levels of victimization than do cis men or cis women.

The ability to form a directional research hypothesis is dependent on the state of the literature on the question of interest. If little or perhaps conflicting research has been conducted, researchers may not be able to form a directional hypothesis. For example, one study examining the relationship between exercise deprivation (not allowing people to exercise) and depression. reported, “We did not have a directional hypothesis, and when participants asked what we expected to find in this study, we replied: ‘We really are not sure since the results of earlier work on exercise deprivation are mixed’” (Mondin et al., 1996, p. 1200).

Table 1.2 Research Hypotheses and Their Independent and Dependent Variables

Research Hypothesis	Independent Variable	Dependent Variable
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher	Type of appeal	Intention to invest in cryptos

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Independent Variable	Dependent Variable
intention to invest in cryptos.		
Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.	Exposure to pro-cannabis messages	Intention to use cannabis
Mothers have more accurate knowledge of their children's online activities than do fathers.	Parent	Knowledge of children's online activities
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	Gender identity	Amount of victimization

Collecting Data

Once a research hypothesis has been formulated, researchers are ready to proceed to the second stage in the research process: collecting data. This step is seen as being composed of three tasks:

drawing a sample from a population,

determining how the variables will be measured, and

selecting a method by which to collect the data.

Table 1.3 Directional and Nondirectional Research Hypotheses

Directional Research Hypothesis	Nondirectional Research Hypothesis
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *higher* intention to invest in cryptos.	Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *different* intention to invest in cryptos.
Exposure to pro-cannabis messages on social media is related to *increased* intentions to use cannabis.	Exposure to pro-cannabis messages on social media is *related to* intentions to use cannabis.
Mothers have *more* accurate knowledge of their children's online activities than do fathers.	Mothers have *different* knowledge of their children's online activities than do fathers.
Nonbinary college students experience *higher* levels of victimization than do cis men or cis women.	Nonbinary college students experience *different* levels of victimization than do cis men or cis women.

Drawing a Sample From a Population

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The first step in collecting data is to identify the group of participants to which the research hypothesis applies. The group to which the results of a study may be applied or generalized is called a population. A **population** is the total number of possible units or elements that could potentially be included in a study. Although people often think the word "population" refers to everyone in the world, researchers typically define their populations more narrowly. For example, the population for the first research hypothesis in Table 1.1 could be "consumers," "U.S. consumers," "Georgia consumers," or "University of Georgia consumers." Researchers typically try to define their populations as broadly as possibly (e.g., "consumers" rather than "University of Georgia consumers") to maximize the applications or implications of their research.

Learning Check 1

Reviewing What You've Learned So Far



1. Review questions
    a. What are the five steps involved in the research process?
    b. Why is it useful to review and evaluate theories and research before conducting a study?
    c. What are the two main characteristics of research hypotheses?
    d. What is the difference between an independent variable and a dependent variable?
2. The following are several research hypotheses from published studies. For each research hypothesis, identify the independent and dependent variable.
    a. "College students will rate instructors who dress formally (i.e., business suit and tie) as having more expertise than instructors who dress casually (i.e., slacks and shirt)" (Sebastian & Bristow, 2008, p. 197).
    b. "It was expected tha. . . . greater amounts of television viewin. . . . would predict greate. . . . posttraumatic stress symptoms" (McLeish & Del Ben, 2008, p. 417).
    c. "Do borrowers who did not complete college have greater financial distress compared with . . . non-borrowers who did not complete college [or] borrowers who earned a degree?" (Jabbari et al., 2023, p. 6).
    d. "We expected that drivers would believe that texting while driving is more detrimental to their driving performance than browsing social media" (Hashash et al., 2019, p. 70).

It's extremely difficult to collect data from all members of a population. Imagine, for example, the time and money needed to collect information from every consumer in the United States! For this reason, researchers typically draw conclusions about populations based on information collected from a sample of the population. A **sample** is a subset or portion of a

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

population. Table 1.4 describes the samples used in the four studies introduced in Table 1.1 . As you can see from Table 1.4 , samples vary greatly in terms of their targeted population and size (the number of participants). The extent to which the results of a study can be generalized to the target population depends on the extent to which the sample is representative of the population.

Determining How Variables Will Be Measured: Levels of Measurement

The research hypotheses described in Table 1.1 involve variables such as type of appeal, intentions to use cannabis, and gender identity. To conduct a research study, the researcher must decide how to measure the variables stated in the research hypothesis. **Measurement** is the assignment of numbers or categories to objects or events according to rules.

For example, to measure the variable “height,” a researcher might use “number of inches from the ground in bare feet” as a form of measurement. To measure “success in college,” a student’s grade point average (GPA) may be obtained from school transcripts. “Self-esteem” might be measured by having people complete a questionnaire and be categorized as having either “low” or “high” self-esteem. As these examples demonstrate, different types of variables have different forms of measurement. Consequently, researchers have identified four distinct levels for measuring variables: nominal, ordinal, interval, and ratio.

The values of variables measured at the **nominal level of measurement** differ in category or type. The word *nominal* implies having to do with “names,” such that we use first and last names to distinguish between people. Gender identity is an example of a nominal variable, in that it consists of categories or types (such as cis male and cis female) rather than numeric values. In the first research hypothesis in Table 1.1 , the independent variable of “type of appeal” is measured at the nominal level with two categories: FOMO and non-FOMO.

Variables measured at the **ordinal level of measurement** have values that can be placed in an order relative to the other values. Rankings (such as finishing first, second, or third in a race), size (small, medium, large, or extra-large), and ratings (below average, average, or above average) are examples of ordinal variables. The values of ordinal variables indicate that one value represents relatively more or less of the variable than other values, but they don’t specify the precise difference between values. For example, although you can say that a “medium” drink is smaller than a “large” drink which is smaller than an “extra-large” drink, you can’t specify the exact difference in size between the three drinks.

Table 1.4 Research Hypotheses and Study Samples

Research Hypothesis	Study Sample
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.	“We recruited 150 participants from Amazon MTurk” (Friederich et at., 2024, p. 105).
Exposure to pro-cannabis messages on social media is	“Teens (N = 350) . . . needed to be between 13 and 17 years old and from Washington state” (Willoughby et

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Study Sample
related to increased intentions to use cannabis.	al., 2024, p. 185).
Mothers have more accurate knowledge of their children's online activities than do fathers.	"triadic data including reports from children aged 13 to 18, mothers and fathers . . . recruited in Flanders, the northern, Dutch speaking region of Belgium . . . a total of 357 valid triads was achieved" (Symons et al., 2017, p. 405).
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	"undergraduate college students aged 18–30 who were enrolled at 4-year institutions. This yielded a sample of 46,510 students" (Marx et al., 2024, p. 334).

The values of variables measured at the **interval level of measurement** are numbers equally spaced along a continuum. One example of an interval variable is the Fahrenheit scale of temperature. Here, a difference of 5 degrees has the same meaning anywhere along the scale; for example, the difference between 45°F and 50°F is the same as the difference between 65°F and 70°F. Many variables studied in the behavioral sciences (e.g., personality characteristics or attitudes) are considered to be measured at the interval level of measurement.

Variables at the **ratio level of measurement** are identical to interval variables, with one exception: Ratio scales possess what is known as a true zero point, for which the value of zero (0) represents the complete absence of the variable. Variables that describe a physical dimension (such as height, weight, distance, and time duration) typically have a true zero point. In the last research hypothesis in Table 1.1 , the variable "number of times victimized in last 12 months" is a ratio variable because it has a true zero point, where zero would indicate the complete absence of victimizations.

One advantage of ratio measurement versus interval is that ratio variables allow for a greater number of comparisons among values. Consider, for example, the value 6 for the ratio variable "inches." Not only is the difference between 4 and 6 inches the same as the difference between 6 and 8 inches (thereby involving addition and subtraction, which also can be used with interval variables), 6 inches is twice as long as 3 inches and half as long as 12 inches (involving multiplication and division). Multiplication and division comparisons cannot be made with interval variables. You cannot, for example, say that a temperature of 90°F is three times as much temperature as 30°F.

The boldfaced and italicized text in Table 1.5 illustrates how the four studies described in the earlier tables measured their independent and dependent variables. As you can see, the variables in these studies have been measured in a variety of ways at different levels of measurement.

Table 1.5 Research Hypotheses and Measurement of Variables

Research Hypothesis	Independent Variable	Dependent Variable

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Independent Variable	Dependent Variable
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.	Type of appeal **FOMO, non-FOMO (nominal)**	Intention to invest in cryptos **Likelihood of investing** **1-7 scale (1 = Highly unlikely to 7 = Highly likely) (ordinal)**
Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.	Exposure to pro-cannabis messages on social media **How often saw pro-cannabis messages (ratio)**	Intentions to use cannabis **Likelihood of using cannabis in the future (ratio)**
Mothers have more accurate knowledge of their children's online activities than do fathers.	Type of parent **Mother, Father (nominal)**	Accuracy of online activities **Accurate, somewhat accurate, inaccurate (nominal)**
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	Gender identity **Nonbinary, cis man, cis woman** **(nominal)**	Number of times victimized in last 12 months **(ratio)**

What difference does the level of measurement make?

A variable's level of measurement has important implications for researchers in that it influences how research hypotheses are stated and how data are analyzed. For example, imagine you study the variable "success in college." You could measure this variable at the *nominal* level of measurement by classifying students into one of two groups: successful or unsuccessful. Researchers typically use nominal variables when they are interested in studying differences between groups, such as, "Do successful and unsuccessful students differ in their study habits?"

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

You could instead measure success in college as an *ordinal* variable by ranking students from top to bottom, which enables a researcher to study relative differences with less concern for the precise size of these differences. For example, using ranked data, you could ask, “How similar are younger and older faculty members’ rankings of their students?”

To study success in college at the *interval* level, you could rate each student on a scale from 1 (low) to 100 (high). You could then use these ratings to address a research question such as, “Is there a relationship between faculty ratings of their students and students’ ratings of themselves?”

Finally, to measure success in college using the *ratio* level of measurement, you could use students’ salary after graduation (in dollars) as an indicator of success. Salary is a ratio variable because it contains a true zero point. In practice, ratio variables can be used to ask many of the same questions as those involving interval variables. For example, you could ask, “Is there a relationship between faculty members’ ratings of students and students’ salaries after graduation?”

As you see, how researchers measure their variables influences how they state their research questions and hypotheses. The level of measurement has another important implication: It helps determine the statistics researchers use to analyze their data. Certain statistics can be applied to variables measured at the interval or ratio levels of measurement, whereas others are appropriate for nominal or ordinal variables. As we move from chapter to chapter of this book, the level of measurement of the variables being analyzed will be noted.

Selecting a Method to Collect the Data: Experimental and Nonexperimental Research Methods

In addition to drawing a sample from a population and determining how the variables will be measured, the third step in collecting data is to determine the type of research method to use to collect the data on these variables. Research methods may be classified into two main types:

experimental research methods, and

nonexperimental research methods.

Experimental research methods.

Experimental research methods are methods designed to test causal relationships between variables—more specifically, whether changes in independent variables produce or cause changes in dependent variables. To make inferences about cause-effect relationships, researchers conducting an experiment must first eliminate other possible causes or explanations for changes in the dependent variable besides the independent variable. If it can be claimed that a variable other than the independent variable created the changes in the dependent variable, the research results are said to be confounded. A **confounding variable** is a variable related to an independent variable that provides an alternative explanation for the relationship between the independent and dependent variables.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

To illustrate the concept of confounding variables, consider the following question: Is there a relationship between a mother's ethnicity and the birth weight of her baby? Although research has shown differences in the birth weights of babies of different ethnicities, one study noted that some of this research did not eliminate the possibility that these differences may be partly due to differences between ethnic groups on factors associated with low birth weights, such as the mother's average age at the time of pregnancy and behaviors such as smoking and drinking (Ma, 2008). These other factors are considered confounding variables in that they provide alternative explanations for the relationship between ethnicity and babies' birth weights. Confounding variables make it difficult for researchers to draw firm conclusions regarding the relationship between the independent and dependent variable.

Learning Check 2

Reviewing What You've Learned So Far



1. Review questions
    a. What is the difference between a population and a sample?
    b. Within the research process, what is the relationship between populations and samples?
    c. What are the four levels of measurement? How do they differ?
2. For each of the following variables, name the scale of measurement (nominal, ordinal, interval, or ratio).
    a. Type of school attended (public, private)
    b. Probability of graduating college in 4 years (0% to 100%)
    c. Rating of television (good, better, best)
    d. Number of computers in home

Researchers minimize the influence of confounding variables in two main ways. First, researchers exercise experimental control by making the research setting (i.e., characteristics of the research participants, the instruments or measures administered, the methods used to collect data) the same for all participants. In the birth weight study, for example, the research design might control for the effect of parental smoking on birth weight by only including nonsmokers in the study sample, thereby excluding those who smoke.

Another way researchers exert control over the research setting is by including a condition known as a **control group** , a group of participants in an experiment not exposed to the independent variable. For example, if you conducted an experiment examining the effects of caffeine on one's health, one group (the experimental group) might be instructed to drink coffee while a second group (the control group) drinks decaffeinated coffee. By comparing the results of the two groups, researchers can then assess the impact of caffeinated coffee on the health of those who drink it.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Because researchers cannot possibly identify and control for all potential confounding variables, a second strategy used to minimize the influence of confounding variables is **random assignment** , which is assigning participants to each category of an independent variable in such a way that each participant has an equal chance of being assigned to each category. For example, a researcher could randomly assign a participant to the experimental or control condition by flipping a coin: heads for experimental, tails for control. The purpose of random assignment is to equalize or neutralize the effects of confounding variables by distributing them equally over the levels of the independent variable.

Nonexperimental (correlational) research methods.

Experimental research designs are one of the best tools researchers have for making causal inferences. However, the ability to make causal inferences requires a great deal of control that may not always be possible or desirable. For this reason, researchers often employ **nonexperimental research methods** , which are research methods designed to measure relationships between variables without having the ability to infer cause-effect relationships. These research methods are sometimes referred to as *correlational* research methods. It is important to understand that correlational research methods are different from correlational statistics we'll introduce in Chapter 13. Correlational research methods pertain to how data are *collected* ; correlational statistics pertain to how data are *analyzed* . Some of the most common types of nonexperimental research designs include quasi-experiments, survey research, observational research, and archival research.

Quasi-experimental research compares naturally formed or preexisting groups rather than randomly assigning participants to conditions. One example of this type of research was a study examining the effects of the COVID-19 pandemic on three groups of students: those living on campus, those living off campus but within commuting distance to school, and those living far away in another city, state, or country (Bisconer & McGill, 2023). This study is a quasi-experiment because the students were not randomly assigned to one of the three living conditions before collecting data.

Survey research methods obtain information directly from a group of people regarding their opinions, beliefs, or behavior. The goal of survey research, which can include using questionnaires and interviews, is to obtain information from a sample to represent or estimate the views of a larger population. One example of survey research examined college students' mental health by having them fill out a survey in which they reported their spirituality (the desire for truth, meaning, and significance of one's life) and their self-acceptance (knowledge and acceptance of one's self and limitations) (Nonis et al., 2024). Survey research is nonexperimental in that the variables being measured are not manipulated by the researcher.

Observational research is the systematic and objective observation of naturally occurring behavior or events. Observational research is often used to study phenomena that the researcher cannot or should not deliberately manipulate. For example, researchers studying interpersonal conflict among children would never force children to fight each other. One study resolved this issue by videotaping groups of adolescents discussing topics such as competition and peer pressure and then counting the number of times the adolescents teased or mocked each other (Connolly et al., 2015).

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Rather than observing or measuring behavior directly, **archival research** is the use of archives (records or documents of the activities of individuals, groups, or organizations) to examine research questions or hypotheses. One example of an archival research study studied the question of whether men are perceived to be more creative than women (Proudfoot et al., 2015). In this study, the 100 most viewed talks on TED.com were identified, and the researchers compared viewers' posted ratings of the talks given by men versus women.

Labeling variables in nonexperimental research.

One issue that often confuses students when learning about nonexperimental research is the practice of referring to the variables as the "independent variable" and the "dependent variable." For example, in the study described previously that involved biological sex and TED.com ratings, biological sex may be called the "independent variable." However, because biological sex isn't manipulated by the researcher, it would be inappropriate to say that biological sex "causes" differences in TED.com ratings. To avoid confusion, researchers sometimes refer to the independent variable and the dependent variable in a nonexperimental study as the **predictor variable** and **criterion variable**, respectively.

Analyzing the Data

Once data collection has been completed, the next step in the research process involves data analysis. This part of the research process addresses the primary topic of this book: statistics. Because this is the first chapter of this book, we won't describe analyzing data in detail. Instead, we introduce the notion that there are two main purposes of analyzing data: (1) to organize, summarize, and describe data that has been collected and (2) to test and draw conclusions about ideas and hypotheses. These two purposes are met by calculating two main types of statistics: descriptive statistics and inferential statistics.

Calculating Descriptive Statistics

Descriptive statistics are statistics used to summarize and describe a set of data for a variable. For example, you've probably heard of statistics used to describe or summarize aspects of our society, such as crime statistics and unemployment statistics. Using a research-related example, Caitlin Abar, a researcher at Brown University, conducted a study examining increases in students' alcohol-related behaviors after entering college (Abar, 2012). To measure students' level of alcohol use, she created a variable called "typical weekend drinking," which "was measured as the sum of drinks consumed on a typical Friday and Saturday within the past 30 days" (p. 22). One way to summarize students' responses to the typical weekend drinking variable (0, 1, 2, etc.) would be to calculate the mean, which is the mathematical average of a set of scores. The mean, described in Chapter 3, is an example of a descriptive statistic. The first part of this book describes a variety of descriptive statistics used by researchers to summarize data they've collected.

Calculating Inferential Statistics

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Besides summarizing and describing data, a second purpose of statistics is to test hypotheses. **Inferential statistics** are statistical procedures used to test hypotheses and draw conclusions from data collected during research studies. Using inferential statistics, researchers make inferences about what's believed to exist in populations based on data gathered from samples of the population.

As an example of an inferential statistic, let's return to the example of students' alcohol-related behaviors (Abar, 2012). This study was interested in seeing whether these behaviors were related to aspects of these students' relationships with their parents. One aspect was "alcohol communications," defined as "how often they discussed alcohol related topics with their parents" (p. 22). To test a hypothesis regarding the relationship between students' alcohol use and their alcohol communications, we could use an inferential statistic known as the Pearson correlation coefficient, which will be discussed in Chapter 13. The chapters in the last half of this book discuss a broad variety of inferential statistics, along with detailed instructions for analyzing and drawing conclusions from statistical data and presenting research findings.

Drawing a Conclusion Regarding the Research Hypothesis

Once statistical analyses have been completed, the next step is to interpret the results of the analyses as they relate to the study's research hypothesis. More specifically, do the findings of the analyses support or not support the research hypothesis? The word *support* in the previous sentence is very important. Students and researchers are sometimes tempted to say their findings "prove" their hypotheses are either "true" or "false." However, as we'll discuss in Chapter 5, because researchers typically don't collect data from the entire population, they can't know with 100% certainty whether their hypothesis has been proven to be true or false. A *huge* part of this book is learning how to draw appropriate conclusions from statistical analyses.

Communicating the Findings of the Study

Conducting research requires a variety of different skills: conceptual skills to develop research hypotheses, methodological skills to collect data, and mathematical skills to analyze these data. Another integral part of the research process is the communication skills needed to inform others about a study. Researchers must be not only skilled scientists but also effective writers.

Learning Check 3

Reviewing What You've Learned So Far



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1. Review questions
    a. What are the main differences between experimental and nonexperimental research methods?
    b. What are confounding variables? What do researchers do to minimize the effects of confounding variables?
    c. What are the main types of nonexperimental research methods?
    d. What are the relative strengths and weaknesses of experimental and nonexperimental research?
    e. What are the main purposes of descriptive and inferential statistics?

For many years, the American Psychological Association (APA), the professional association for psychologists, has provided researchers guidance on how to communicate the results of their research. In 1929, the APA published a seven-page article entitled, “Instructions in Regard to Preparation of Manuscript.” In contrast, the seventh edition of the *Publication Manual of the American Psychological Association* , published in 2020, consists of 428 pages! This increase in length highlights the challenge of communicating research effectively and appropriately.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1.4 Plan of the Book

The primary purpose of this chapter was to introduce statistics and place it within the larger research process. The remainder of this book discusses different statistics used by researchers to examine various questions of interest. Although it's critical for you to learn to correctly calculate statistics, it is equally important to understand the role of statistics within the research process and appreciate issues related to the use (and sometimes misuse) of statistics.

The first half of the book (Chapters 2 through 6) introduces conceptual and mathematical issues that are the foundation of statistical analyses. Chapters 2, 3, and 4 discuss how data may be examined, described, summarized, and presented in numeric, visual, and graphic form. Descriptive statistics are introduced and discussed in these chapters. Chapters 5 and 6 introduce characteristics of the inferential statistics used to test research hypotheses. The chapters in the second half of the book cover a number of different inferential statistics used with different types of variables.

In most cases, the chapters in this book share a uniform format and structure and are centered on the research process, using examples of actual published research studies that address questions with which you may be familiar. For each example, we'll clearly state the research hypothesis, briefly describe how the data in the study were collected, illustrate the calculation and interpretation of descriptive and inferential statistics, discuss the extent to which the statistical analyses support the research hypothesis, and show how to communicate one's research to a broader audience.

1.5 Looking Ahead

We began this chapter by defining statistics and providing reasons why learning about them may be of value to you. Next, we placed statistics within the stages of the larger research process, a process that will be the foundation of this textbook. As we've explained, the research process is centered on a research hypothesis, which is a predicted relationship between variables. Once a hypothesis has been stated, the next step is to collect data to be analyzed. However, before conducting statistical analyses on a set of data, there is preliminary work that must be completed. The next chapter will focus on methods used by researchers to examine data.

1.6 Summary

Statistics may be defined as a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data. As such, statistics not only is concerned about how data are analyzed but also recognizes the importance of understanding how data are collected and how the results of analyses are interpreted and communicated. There are a variety of reasons why students should learn statistics: Statistics are encountered in a wide variety of daily activities, students may be asked or required to read and interpret the results of statistical analyses in their courses, students may use statistics in conducting their own research, and statistics may help one's career.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Researchers conduct research using the **scientific method** of inquiry, a method of investigation that uses the objective and systematic collection and analysis of empirical data to test theories and hypotheses. The research process used within the scientific method of inquiry consists of five main steps: developing a research hypothesis to be tested, collecting data, analyzing the data, drawing a conclusion regarding the research hypothesis, and communicating the findings to the study.

A **research hypothesis** is a statement regarding an expected or predicted relationship between variables. Developing a research hypothesis involves identifying a question or issue to be examined, reviewing and evaluating relevant theories and research, and stating the research hypothesis to be tested in the study. A **theory** is a set of propositions used to describe or explain a phenomenon. A research hypothesis contains **variables** , which are properties or characteristics of some object, event, or person that can take on different values. More specifically, a research hypothesis states the nature and direction of a proposed relationship between an **independent variable** (a variable manipulated or controlled by the researcher) and a **dependent variable** (a variable measured by the researcher that is expected to change or vary as a function of the independent variable).

Collecting data involves drawing a sample from a population, determining how the variables will be measured, and selecting a method to collect the data. A **population** is the total number of possible units or elements that could potentially be included in a study; a **sample** is a subset or portion of a population. Variables can be measured at one of four levels of measurement: **nominal** (values differing in category or type), **ordinal** (values placed in an order relative to the other values), **interval** (values equally spaced along a numeric continuum), or **ratio** (values equally spaced along a numeric continuum with a true zero point).

There are two main types of methods used to collect data: **experimental research methods** , designed to test whether changes in independent variables produce or cause changes in dependent variables, and **nonexperimental (correlational) research methods** , designed to examine the relationship between variables without the ability to infer cause-effect relationships.

In experimental research, researchers are concerned about possible **confounding variables** , which are variables related to independent variables that provide an alternative explanation for the relationship between independent and dependent variables. To minimize the impact of confounding variables, a research study may exercise experimental control over various aspects of the situation, including the use of a **control group** , which is a group of participants not exposed to the independent variable the research is designed to test, or **random assignment** , which involves assigning participants to each category of an independent variable in such a way that each participant has an equal chance of being assigned to each category.

Examples of nonexperimental research methods are **survey research** , in which information is directly obtained from a group of people regarding their opinions, beliefs, or behavior; **observational research** , which involves the systematic and objective observation of naturally occurring events; and **archival research** , which uses archives (records or documents of the activities of individuals, groups, or organizations) to examine research questions or hypotheses. In nonexperimental research, the independent variable and

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

dependent variable may be referred to as the **predictor variable** and **criterion variable**, respectively.

Once data have been collected, the next step is to analyze them using two main types of statistics: **descriptive statistics**, which summarize and describe a set of data, and **inferential statistics**, which are statistical techniques used to test hypotheses and draw conclusions from data collected during research studies.

Once the statistical analyses have been completed, the results of the analyses are interpreted regarding whether they support or not support the study's research hypothesis. The final step in the research process is to communicate the study to a broader audience.

1.7 Important Terms

archival research

confounding variable

control group

criterion variable

dependent variable

descriptive statistics

experimental research methods

independent variable

inferential statistics

interval level of measurement

level of measurement (nominal, ordinal, interval, ratio)

measurement

nonexperimental research methods

observational research

population

predictor variable

quasi-experimental research

random assignment

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

research hypothesis

sample

scientific method

statistics

survey research

theory

variable

1.8 Exercises

1. Identify the independent and dependent variable for each of the following research questions and hypotheses from published articles.
    a. “We hypothesize that finance majors will embody narcissistic traits more than other business majors” (Wood et al., 2021, p. 924).
    b. “Men are expected to employ physical characteristics (particularly those that are directly related to sex) more often than women in selecting dating candidates” (Hetsroni, 2000, p. 91).
    c. “It is hypothesized that perceptions of emotional pain and support needs will be higher for the community of the Black victim than those of the officers’” (Kahn et al., 2024).
    d. “Gen Z students are more likely to believe that ‘Group projects bring out the worst in people’ than Millennials” (Schlee et al., 2020, p. 145).
    e. “We hypothesized to find . . . higher stress and anxiety post COVID in the Portuguese group in contrast with the British ” (Penalba-Sanchez et al., 2022, p. 322).
    f. “We hypothesized that consumers would think that a ‘sale price’ presentation would generate a greater monetary savings than an ‘everyday low price’ presentation” (Tom & Ruiz, 1997, p. 403).
2. The following are a number of research questions and hypotheses from actual published articles. For each hypothesis, identify the independent and dependent variable.
    a. “The use of color in a Yellow Pages advertisement will increase the perception of quality of the products for a particular business when compared with noncolor advertisements” (Lohse & Rosen, 2001, p. 75).
    b. “We hypothesized that parents who use more frequent corporal and verbal punishmen. . . . will report more problem behaviors in their children” (Brenner & Fox, 1998, p. 252).
    c. “It was hypothesized that adolescents with anorexia nervosa woul. . . . be more respectful when compared with peers with bulimia nervosa” (Pryor & Wiederman, 1998, p. 292).
    d. “The purpose of the present research was to assess brand name recognition as a function of humor in advertisements. . . . It was predicted that participants would recognize product brand names that had been presented with humorous

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

advertisements more often than brand names presented with nonhumorous advertisements" (Berg & Lippman, 2001, p. 197).

e. "The purpose of this study was to explore the relation between time spent in daycare and the quality of exploratory behaviors in 9-month-old infant. . . . it was hypothesized tha. . . . infants who spent greater amounts of time in center-based care would demonstrate more advanced exploratory behaviors than infants who did not spend as much time in center-based care" (Schuetze et al., 1999, p. 269).

f. "It was hypothesized that students would score higher on test items in which the narrative contains topics and elements that resonate with their daily experiences versus test items comprised of material that is unfamiliar" (Erdodi, 2012, p. 172).

3. Identify the independent and dependent variable for each of the following research questions and hypotheses from published articles.
    a. "It was hypothesized that the number of selfies posted . . . would be positively correlated with dimensions of narcissism" (Barry et al., 2017, p. 51).
    b. "We hypothesized that preschoolers' self-reported preferences for aggressive media . . . would be positively related to parent reports of children's aggressive behaviors" (Jamnik & DiLalla, 2018, p. 159).
    c. "We hypothesized that helicopter parenting would lead to higher levels of depression" (Darlow et al., 2017, p. 3).
    d. "We predicted that weight-related shame would be associated with greater binge eating" (Craven & Fekete, 2019, p. 45).
    e. "We hypothesize that men with strong gender identities will be more sensitive to manipulated threats to their masculinity compared with men with gender identities that are less central to their identity" (Harrison & Michelson, 2019, p. 65).
    f. "After exposure to fake news story with user comments identifying the news as fake, people are less likely to share the fake news than after exposure to the fake news story with user comments supportive of the content" (Colliander, 2019, p. 204).
4. Name the scale of measurement (nominal, ordinal, interval, ratio) for each of the following variables:
    a. The amount of time needed to react to a sound
    b. Gender
    c. Score on the Scholastic Aptitude Test (SAT)
    d. Political orientation (not at all conservative, conservative, very conservative)
    e. Political affiliation (Democrat, Republican, Independent)
5. Name the scale of measurement (nominal, ordinal, interval, ratio) for each of the following variables:
    a. One's age (in years)
    b. Size of soft drink (small, medium, large, extra large)
    c. Voting behavior (in favor vs. against)
    d. IQ score
    e. Parent (mother vs. father)
6. A faculty member wishes to assess the relationship between students' scores on the SAT and their performance in college.
    a. What is a possible research hypothesis in this situation?
    b. What are the independent and dependent variables?
    c. How could you measure the variable "performance in college" at each of the four levels of measurement?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

7. A researcher hypothesizes that drivers who use cellular phones will get into a greater number of traffic accidents than drivers who do not use these phones. For this example,
    a. What are the independent and dependent variables?
    b. How could you measure the dependent variable?
    c. How could you conduct this study using an experimental research method? How could you conduct this study using a nonexperimental method?

Answers to Learning Checks

Learning Check 1

1.

    a. IV: Instructor's dress; DV: Expertise

    b. IV: Amount of television viewing; DV: Posttraumatic stress symptoms

    c. IV: Student group; DV: Financial distress

    d. IV: Activity while driving; DV: Driving performance

Learning Check 2

2.

    a. Nominal

    b. Ratio

    c. Ordinal

    d. Ratio

Answers to Odd-Numbered Exercises

1.

    a. IV: Type of business major; DV: Narcissistic traits

    b. IV: Biological sex; DV: Employment of physical characteristics

    c. IV: Community; DV: Perceptions of emotional pain and support needs

    d. IV: Generation; DV: Beliefs about group projects

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

e. IV: Country; DV: Stress and anxiety

f. IV: Price presentation; DV: Perception of savings

3.

a. IV: Number of selfies posted; DV: Narcissism

b. IV: Preference for aggressive media; DV: Parent reports of children's aggressive behaviors

c. IV: Helicopter parenting; DV: Depression

d. IV: Weight-related shame; DV: Binge eating

e. IV: Strength of gender identity; DV: Sensitivity to threat to masculinity

f. IV: Type of user comments to news story; DV: Sharing of news story

5.

a. Ratio

b. Ordinal

c. Nominal

d. Interval

e. Nominal

7.

a. IV: Cellular phone use; DV: Number of traffic accidents

b. Example: Determine the number of traffic accidents each person has experienced in the past year.

c. Example of experimental: Use a driving simulation with the "driver" talking on a phone and measure the number of potential driving mistakes or accidents.

d. Example of nonexperimental: Take a survey of the number of accidents people have been in and whether or not they were using a phone during the accident.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CHAPTER 2 EXAMINING DATA: TABLES AND FIGURES

Learning Objectives

2.1 Describe the main purpose and the research hypothesis of the “lottery” study.

2.2 Describe different reasons for examining data.

2.3 Describe the steps used to create a frequency distribution table.

2.4 Describe the steps used to create a grouped frequency distribution table.

2.5 Explain the differences between bar charts, pie charts, histograms, and frequency polygons.

2.6 Describe the three aspects and types of distributions of data.

Chapter 1 described a process by which research may be conducted, which begins by stating a research hypothesis. The next step involves collecting data that will ultimately be used in statistical analyses designed to test the research hypothesis. However, researchers don’t conduct analyses until they’ve first examined the data to make sure appropriate conclusions can be drawn from the analyses. This chapter describes *why* researchers examine data and *how* data may be examined using tables, charts, and graphs. As with many of the chapters in this book, this discussion will be guided by the findings from a published research study.

2.1 An Example From the Research: Winning the Lottery

A friend of yours flips a coin and asks you to guess whether it’s “heads” or “tails.” You guess “tails,” but the answer is “heads.” Your friend flips the coin a second time, and again you predict “tails.” But again the answer is “heads.” On the third coin flip, you guess “tails” once more—but, once again, you’re wrong. As your friend prepares to flip the coin a fourth time, you tell yourself there’s little chance the coin could be “heads” again. But the reality is that the fourth coin flip is equally likely to be “heads” or “tails” because the outcome of the first three coin flips has absolutely no impact on the fourth flip. In other words, the outcome of a coin flip is a random event that can’t be controlled or predetermined. But do you really believe this?

The impact of random events can take on greater significance than the outcome of a coin flip. For example, government-run lotteries provide the opportunity to win huge amounts of money through the random selection of numbers. In 2001, Dr. Karen Hardoon, a researcher at McGill University in Canada, studied the thought processes people use when purchasing lottery tickets (Hardoon et al., 2001). As previous research had found gamblers believe they can control random events such as the rolling of dice or the spinning of a roulette wheel, Dr. Hardoon and her colleagues were interested in

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

determining “whether individuals with gambling problems perceive the purchase of lottery tickets in a similar manner as non-problem gamblers” (p. 752). On the basis of earlier findings, Dr. Hardoon hypothesized that problem gamblers are less likely to believe lottery outcomes are random than are nonproblem gamblers.

To test their research hypothesis, Dr. Hardoon and her associates located a group of gamblers and asked whether they had ever experienced problems as a result of their gambling. Each participant was classified into one of two groups: problem gamblers and nonproblem gamblers.

The participants in the study, which we’ll refer to as the lottery study, were shown four lottery tickets and asked, “If you were to buy a ticket to play in the lottery, which one would you select?” Each ticket contained six numbers selected from among the numbers 1 through 49; the four tickets were given a particular label. The *sequence* ticket had numbers in consecutive order (30-31-32-33-34-35); numbers on the *pattern* ticket increased by 5s (5-10-15-20-25-30); the *nonequilibrated* numbers (35-37-40-43-44-49) were clustered at the upper end of the 49 numbers, and the *random* ticket (7-8-23-34-36-42) shared none of the characteristics of the other three tickets. We’ll call the variable in this study “lottery ticket,” a variable with four categories: sequence, pattern, nonequilibrated, and random.

The researchers recorded which of the four tickets each participant chose. Table 2.1 lists the data for the lottery ticket variable for the nonproblem gamblers (the data for the problem gamblers are provided in the exercises at the end of this chapter). Looking at this table, we see that the nonproblem gamblers differed in their lottery ticket choices; however, it’s difficult to describe these differences given the unorganized nature of the collected data. Consequently, before conducting statistical analyses on their data, researchers typically organize and examine them. In the following sections, we will discuss why and how data are organized and examined.

2.2 Why Examine Data?

There are a variety of reasons why researchers examine data they’ve collected before analyzing the data. These reasons include the following:

To gain an initial sense of the data

To identify outliers

To evaluate research methodology

To determine whether data meet statistical criteria and assumptions

Each of these reasons is discussed in the following sections.

Table 2.1 The Lottery Ticket Choices of 22 Nonproblem Gamblers

Participant	Lottery Ticket
1	**Random**
2	**Random**
3	**Nonequilibrated**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Participant	Lottery Ticket
4	**Pattern**
5	**Random**
6	**Sequence**
7	**Random**
8	**Random**
9	**Pattern**
10	**Pattern**
11	**Random**
12	**Pattern**
13	**Random**
14	**Nonequilibrated**
15	**Random**
16	**Pattern**
17	**Nonequilibrated**
18	**Random**
19	**Random**
20	**Pattern**
21	**Random**
22	**Random**

Gaining an Initial Sense of the Data

Because researchers spend a great amount of time and energy designing their research studies, they're eager to analyze the data they've collected. Students taking courses in statistics, on the other hand, face a different challenge: They're given sets of data with which they have no prior experience. Although both researchers and students may be motivated to analyze their data as quickly as possible, it's important to first examine the data in order to gain an initial sense of them.

As an example, imagine you conduct a study regarding the fuel efficiency of automobiles and collect data on the miles per gallon (MPG) of different types of cars. You might expect the scores in your data to range from 10 to 50 MPG, with the majority between 25 and 35 MPG. Examining data helps researchers gain an initial sense of whether the data are in line with their expectations. Examining data also provides a way of confirming the results of later statistical analyses. We've found that students' errors in their calculations could have been avoided had they first looked at the set of data. For example, examining data will help you avoid making statements such as, "The average MPG in my sample was 127 MPG."

Identifying Outliers

Researchers also examine their data to identify **outliers** , defined as rare, extreme scores that lie outside of the range of the majority of scores in a set of data. Imagine, for example, that a group of 15 students is asked to take a 20-item test of mathematical calculations; each student's score is the number of test items answered correctly. In scoring the tests, a researcher finds that 14 students

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

correctly answered between 8 and 17 questions. However, one student provided the wrong answer to every question for a score of 0. In this case, the score of 0 would be considered an outlier.

If not identified, outliers may distort conclusions researchers draw about their sample and their data, particularly when the sample is relatively small. In the previous example, the single score of 0 might lead an observer to conclude the class as a whole performed more poorly on the test than they actually did.

Evaluating Research Methodology

Examining data is also useful in assessing the effectiveness of the methods used to collect the data. For the 20-item math test, for example, it might be reasonable to expect the majority of scores to fall in the middle of the 0 to 20 range. But what if every student answered at least 17 of the 20 questions correctly? You might conclude the test was too easy and that more difficult questions are needed to accurately assess students' math skills. By comparing the collected data for a variable with the range of possible values, researchers are able to evaluate and modify their measurement tools.

Determining Whether Data Meet Statistical Criteria and Assumptions

Researchers also examine their data to see whether they meet certain statistical criteria and assumptions. Many of the statistics in this book are based on assumptions regarding the shape and nature of a set of data that has been collected. For example, these procedures assume that in most data sets, the majority of scores will fall in the middle of the range of possible values, with a small amount of scores at the highest and lowest possible values (you may have heard this referred to as "the bell curve"). In the example of the 20-item math test, it might be assumed that most scores will be between 7 and 13, with smaller number of scores less than 7 or greater than 13. When assumptions such as this are not met, it's possible to draw inappropriate conclusions from statistical analyses that are conducted.

2.3 Examining Data Using Tables

One part of examining data that have been collected for a variable involves answering a series of questions:

> What are the values of the variable?
>
> How many participants in the sample have each value of the variable?
>
> What percentage of the sample has each value of the variable?

As it may be difficult to answer questions such as these just by looking at the data in Table 2.1 , researchers typically organize their data by creating a table known as a **frequency distribution table** , a table that summarizes the number and percentage of participants for the different values of a variable. This section discusses the steps involved in creating frequency distribution tables using these questions.

What Are the Values of the Variable?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The first step in organizing the data for a variable is to identify all of the possible values for the variable. Table 2.2 begins the construction of a frequency distribution table by identifying the four values for the lottery ticket variable: sequence, pattern, nonequilibrated, and random.

How Many Participants in the Sample Have Each Value of the Variable?

Once the values of the variable have been identified, the next step in organizing a set of data is to determine the **frequency** of each value, which is the number of participants in a sample corresponding to a value of a variable. These frequencies help indicate the nature and shape of the distribution of scores for a variable in a sample.

Table 2.2 Identifying the Values of the Lottery Ticket Variable

Number	Lottery Ticket
1	Sequence
2	Pattern
3	Nonequilibrated
4	Random

Table 2.3 Using the Tick Mark Method to Determine the Frequencies of the Lottery Ticket Variable

Lottery Ticket	Tick Marks	*f*
Sequence	|	1
Pattern	||||— |	6
Nonequilibrated	|||	3
Random	||||— ||||— ||	12
Total	||||— ||||— ||||—||||—||	22

One way to determine the frequencies for a variable is to sort the data into the different values and then count the number of participants having each value using a “tick mark method,” where a tick mark is placed next to a value of the variable each time that value appears in the sample. As shown in Table 2.3 , the tick marks are then counted to determine the frequency for each value. For example, the number of tick marks and therefore the frequency (f) for the sequence ticket is 1 because it was picked by one participant (Participant 6 in Table 2.1). Similarly, the tick marks for each of the other three types of tickets can be recorded and counted until all of the data in the sample are accounted for.

What Percentage of the Sample Has Each Value of the Variable?

Although it’s important to determine the frequency for each value of a variable, these frequencies by themselves are not always particularly informative. For example, how you interpret “There are 35 Democrats in my sample” is different when the sample consists of 40 voters rather than 400 voters.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Because the interpretation of frequencies depends on the size of the sample, it's useful to calculate the *percentage* of the sample having each value of a variable. These percentages can be calculated using the following formula:

$$\% = \frac{f}{total\,number\,of\,scores} * 100$$

where *f* is the frequency for a value of a variable.

Looking at Table 2.3 , 1 of the 22 nonproblem gamblers chose the sequence ticket. Therefore, the percentage of the sample choosing this ticket is

$$\begin{aligned} \% &= \frac{f}{total\,number\,of\,scores} * 100 \\ &= \frac{1}{22} * 100 = .05 * 100 \\ &= 5\% \end{aligned}$$

After calculating the percentage for each of the four tickets, Table 2.4 provides the frequency distribution table for the lottery ticket variable.

Once a frequency distribution table has been constructed for a variable, researchers can begin to draw conclusions about their sample. For the lottery ticket variable, the random ticket had the highest frequency and percentage (*f* = 12, 54%) compared to the pattern (*f* = 6, 27%), nonequilibrated (*f* = 3, 14%), and sequence (*f* = 1, 5%) tickets. As a result, the researchers in the lottery study made the following observation:

"The results of the present study indicated that, for the entire sample . . . the greatest percentage of tickets selected by the entire sampl. . . . were random tickets." (Hardoon et al., 2001, p. 760)

Table 2.4 Frequency Distribution Table for the Lottery Ticket Variable

Lottery Ticket	***f***	**%**
Sequence	1	**5%**
Pattern	6	**27%**
Nonequilibrated	3	**14%**
Random	12	**54%**
Total	22	**100%**

Learning Check 1

Reviewing What You've Learned So Far

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



1. Review questions
    a. What are some reasons for examining data before conducting statistical analyses?
    b. Why are outliers of concern to researchers?
    c. What questions do you address about a set of data for a variable in constructing frequency distribution tables?
    d. Why is it useful to calculate percentages in addition to frequencies for each value of a variable?
2. For each of the following situations, create a frequency distribution table.
    a. A college counselor asks a group of 116 seniors what their plans are after graduating from college; 71 say they are going to “work,” 34 are planning on going to “graduate school,” and 11 say they are “not sure” what their plans are.
    b. A high school instructor teaching a course finds that of the students she teaches, 69 are Freshmen, 18 are Sophomores, 12 are Juniors, and 9 are Seniors.
    c. A pollster stops 12 people and asks if they are either “in favor” or “against” a local proposition in Table 2.5 :

Table 2.5 Example of People’s Position

Person	Position
1	**In favor**
2	**In favor**
3	**Against**
4	**In favor**
5	**Against**
6	**Against**
7	**In favor**
8	**Against**
9	**In favor**
10	**Against**
11	**In favor**
12	**In favor**

    d. The company that makes M&Ms candy (www.mms.com) conducted a survey asking people which new color they would like to have added: purple, aqua, or pink. Table 2.6 are the votes of a hypothetical sample:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Table 2.6
Hypothetical Sample of New M&M Color

Person	Color
1	**Pink**
2	**Purple**
3	**Pink**
4	**Aqua**
5	**Purple**
6	**Purple**
7	**Purple**
8	**Aqua**
9	**Purple**
10	**Pink**
11	**Aqua**
12	**Pink**
13	**Pink**
14	**Purple**
15	**Aqua**
16	**Pink**
17	**Purple**
18	**Purple**
19	**Aqua**
20	**Purple**
21	**Purple**
22	**Pink**
23	**Pink**
24	**Purple**

2.4 Grouped Frequency Distribution Tables

The frequency distribution table in Table 2.4 was sufficient for presenting the frequencies for the lottery ticket variable. However, other situations may involve variables with a large number of possible values. Consider, for example, the variable grade point average (GPA). The values for GPA typically range from 0.00 to 4.00, with hundreds of possible values in between. Creating a frequency distribution table for GPA could result in an extremely large table, with many values having frequencies of zero.

To examine a variable with a large number of values, it is useful to construct a **grouped frequency distribution table** , a table that groups the values of a variable measured at the interval or ratio level

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

of measurement into a small number of intervals and then provides the frequency and percentage for each interval. To illustrate how grouped frequency distribution tables are constructed, Table 2.7 (a) provides a sample of 42 hypothetical GPAs. In Table 2.7 (b), each of these 42 GPAs has been placed into one of eight intervals that comprise a grouped frequency distribution table.

To illustrate how to read and interpret a grouped frequency distribution table, the lowest interval in Table 2.7 (b) consists of students with GPAs less than (<) 2.00. The frequency in this interval (*f* = 1) consists of one student: Student 29 (GPA = 1.96). The next interval consists of students with GPAs ranging from 2.00 to 2.29 and contains three students (4, 22, and 38). This grouping continues until all of the GPAs in the sample have been accounted for.

To create a grouped frequency distribution table, the real limits of each interval should be identified. The **real limits** of an interval are the values of the variable that are halfway between the top of one interval and the bottom of the next interval. For example, because of how values of GPA are rounded, the lowest value of GPA that would lead someone to be placed in the interval 2.30–2.59 is 2.295, which is halfway between the top of the 2.00–2.29 interval and the bottom of the 2.30–2.59 interval. The **real lower limit** is the smallest value of a variable that would be grouped into a particular interval. In this case, 2.295 represents the real lower limit for the 2.30–2.59 interval. On the other hand, the **real upper limit** is the largest value of a variable that would be grouped into a particular interval. For the 2.30–2.59 interval, the real upper limit is 2.594.

Once the grouped frequency table for the GPA example has been created (Table 2.7(b)), what preliminary conclusions might be drawn about these students? First, because the 2.90–3.19 interval has the greatest frequency (*f* = 10), you might conclude that the typical student in this sample has a GPA close to 3.00. Second, moving in both directions away from the 2.90–3.19 interval, the frequencies of the different intervals grow progressively smaller, with fewer and fewer students having either very high or very low GPAs. Table 2.8 provides a list of guidelines for creating grouped frequency distribution tables.

Cumulative Percentages

In addition to determining the percentage of the sample that has each individual value or interval for a variable, it is sometimes useful to combine these percentages. For the GPA example, you may be interested in knowing, “What percentage of the sample had GPAs less than 2.60?” To answer this question, the percentages in the “%” column of the frequency distribution table must be combined. This type of percentage is referred to as a **cumulative percentage** , defined as the percentage of a sample at or below a particular value of a variable.

Table 2.7a Grade Point Average (GPA), 42 Students: Individual GPAs

Student	GPA
1	**3.05**
2	**2.83**
3	**3.26**
4	**2.19**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Student	GPA
5	**3.52**
6	**3.34**
7	**3.02**
8	**2.97**
9	**3.71**
10	**2.50**
11	**2.77**
12	**3.10**
13	**3.06**
14	**3.83**
15	**2.39**
16	**3.16**
17	**3.92**
18	**2.37**
19	**3.65**
20	**3.70**
21	**3.00**
22	**2.26**
23	**2.65**
24	**3.40**
25	**2.87**
26	**2.92**
27	**2.89**
28	**3.41**
29	**1.96**
30	**3.77**
31	**3.20**
32	**2.86**
33	**2.94**
34	**2.71**
35	**3.43**
36	**3.95**
37	**3.32**
38	**2.28**
39	**3.08**
40	**3.25**
41	**2.40**
42	**3.22**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Table 2.7b Grade Point Average (GPA), 42 Students: Grouped Frequency Distribution Table

GPA	*f*	%
3.80+	3	7%
3.50−3.79	5	12%
3.20−3.49	9	21%
2.90−3.19	10	24%
2.60−2.89	7	17%
2.30−2.59	4	10%
2.00−2.29	3	7%
<2.00	1	2%
Total	42	100%

Table 2.8 Guidelines for Creating Grouped Frequency Distribution Tables

Guideline	Guideline Description
1. *Variables are grouped into approximately 10 intervals.*	The exact number of intervals depends on the data in each sample. For example, data that fall within a small range of values may be accurately represented with a relatively small number of intervals, whereas data extending across a wide range may require a larger number of intervals.
2. *The number of intervals should accurately represent the data.*	The intervals should represent the nature of the data as accurately as possible. For example, there should not be so large a number of intervals that many intervals have frequencies of zero (e.g., 3.00−3.05) or so few intervals that a large majority of the sample falls into only one or two intervals (e.g., 3.00−3.99).
3. *Intervals should be of equal size.*	To make the intervals comparable, they should be of equal size or width. For example, you would not want one interval of 2.01−2.50 (a width of .50) and another of 2.51−2.70 (a width of .20). One exception to this guideline is the lowest or highest interval (e.g., less than 2.00 or greater than 3.70), which is typically wider than the other intervals because it contains a relatively small frequency.
4. *Intervals should not overlap.*	The intervals should be distinct from each other, such that each score may be placed in only one interval. For example, intervals of 2.30− *2.60* and *2.60*− 2.90 are a problem because the GPA 2.60 is in both intervals leading to confusion regarding within which interval to place someone with a GPA of 2.60.

Table 2.9 includes the cumulative percentages for the GPA example. As an example of a cumulative percentage, the cumulative percentage for the 2.30–2.59 interval (19%) has been calculated by combining the percentages for three intervals: < 2.00, 2.00–2.29, and 2.30–2.59: 2% + 7% + 10% = 19%. As a result, we could conclude that 19% of this sample had GPAs less than 2.60. The accumulating of percentages continues until all (100%) of the scores in the sample are combined. As

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

shown in Table 2.9 , cumulative percentages are typically located in the final column of a frequency distribution table and labeled “Cum %.”

Table 2.9 Frequency Distribution Table With Cumulative Percentages for the GPA Variable

GPA	*f*	%	Cum %
3.80+	3	7%	**100%**
3.50−3.79	5	12%	**93%**
3.20−3.49	9	21%	**81%**
2.90−3.19	10	24%	**60%**
2.60−2.89	7	17%	**36%**
2.30−2.59	4	10%	**19%**
2.00−2.29	3	7%	**9%**
< 2.00	1	2%	**2%**
Total	42	100%	

2.5 Examining Data Using Figures

In addition to organizing data into frequency distribution tables, researchers also examine data visually using figures such as charts or graphs. There are four types of figures often used to visually portray data for variables: bar charts, pie charts, histograms, and frequency polygons. Which type of figure a researcher may use depends on the variable’s level of measurement (nominal, ordinal, interval, ratio). Data for variables measured at the nominal or ordinal level of measurement are typically displayed using bar charts or pie charts. The values of interval and ratio variables are graphed using histograms and frequency polygons. Each of these types of visual illustration is discussed in the following sections.

Displaying Nominal and Ordinal Variables: Bar Charts and Pie Charts

The values of variables measured at the *nominal* level of measurement differ in category or type; one example of a nominal variable is the lottery ticket variable discussed earlier, which consisted of four categories: sequence, pattern, nonequilibrated, and random. The values of variables measured at the *ordinal* level of measurement can be placed in an order relative to the other values; age (young, middle age, elderly) is one example of an ordinal variable.

One way to display nominal and ordinal variables is to use a bar chart. A **bar chart** is a figure that uses bars to represent the frequency or percentage of a sample corresponding to each value of a variable, with the different bars not touching each other. Figure 2.1 shows a bar chart for the data from the lottery ticket variable. In this figure, each bar corresponds to one of the four types of lottery tickets; the height of each bar corresponds to the frequency (*f*) of nonproblem gamblers who selected that type of ticket.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

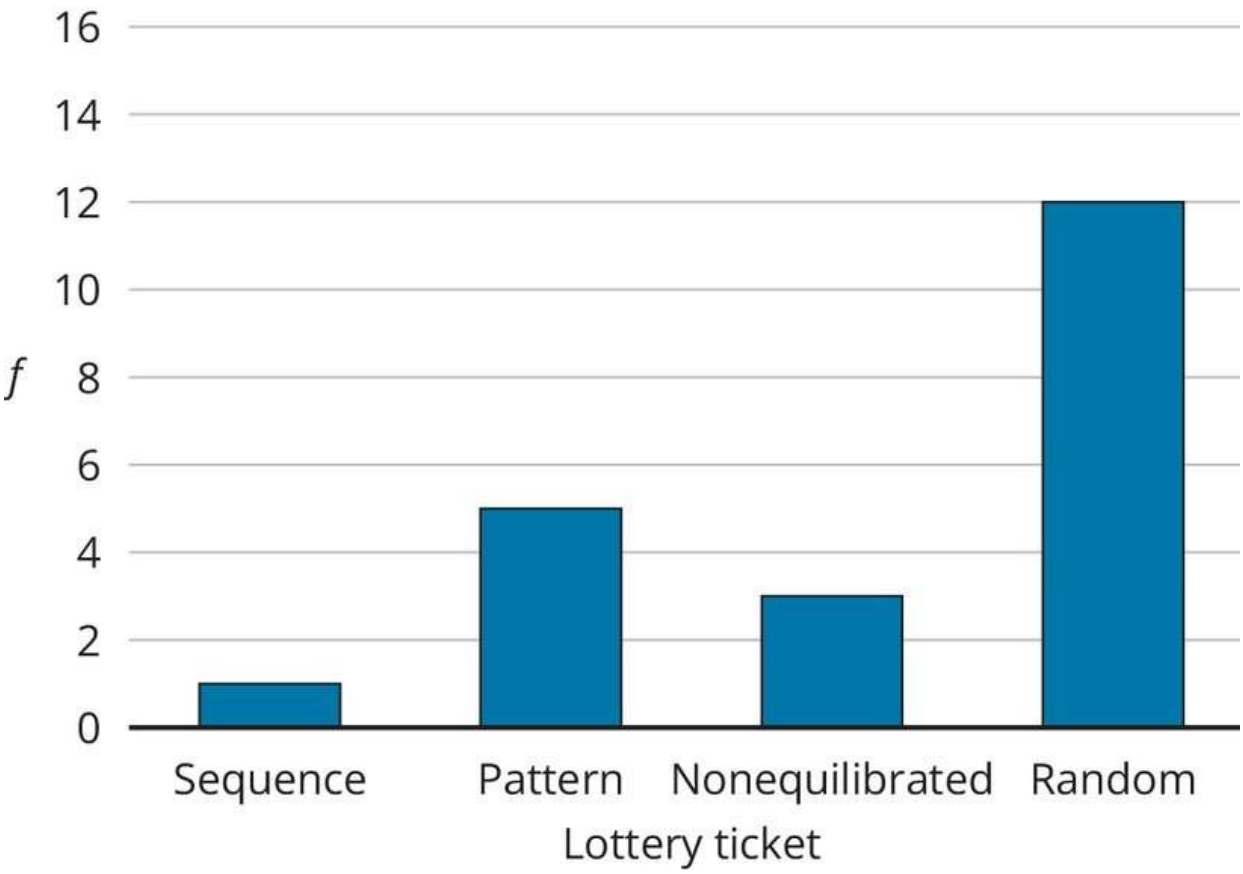


Figure 2.1 Bar Chart Showing the Frequencies of the Lottery Ticket Variable for the Nonproblem Gamblers

When creating a bar chart, the horizontal (X) axis contains the name of the variable and the values of the variable; the vertical (Y) axis contains the frequency (*f*) or percentage (%) for each value. Notice that the bars in the bar chart in Figure 2.1 do not touch each other; the gap between the bars indicates that the values of the variable represent distinct groups or categories that can't be placed along a numeric continuum.

A second type of figure used to display nominal or ordinal variables is a pie chart. A **pie chart** is a figure that uses a circle divided into proportions (slices) that represent the percentage of the sample for each value of a variable. Pie charts are created by dividing the 360 degrees of a circle into proportions corresponding to the percentage of each value.

Figure 2.2 shows a pie chart for the data for the lottery ticket variable. To illustrate how pie charts are constructed, the frequency distribution table in Table 2.4 shows that 5% of nonproblem gamblers chose the sequence ticket. Given that a pie chart consists of 360 degrees, 5% of the 360 degrees are needed to represent the sequence ticket. As 5% * 360 = 18, 18 degrees of the pie is assigned to the sequence ticket. Similarly, 97 degrees of the pie chart (27% of 360) are used to represent the pattern ticket. This process continues until all 360 degrees of the pie chart are accounted for.

The decision whether to use a bar chart or pie chart to illustrate a variable may depend on the number of values for the variable. For example, a pie chart for a variable that has 12 values may

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

result in many of the 12 “slices” being very thin and small, making them hard to see, meaning that a bar chart may be a better choice to illustrate this variable.

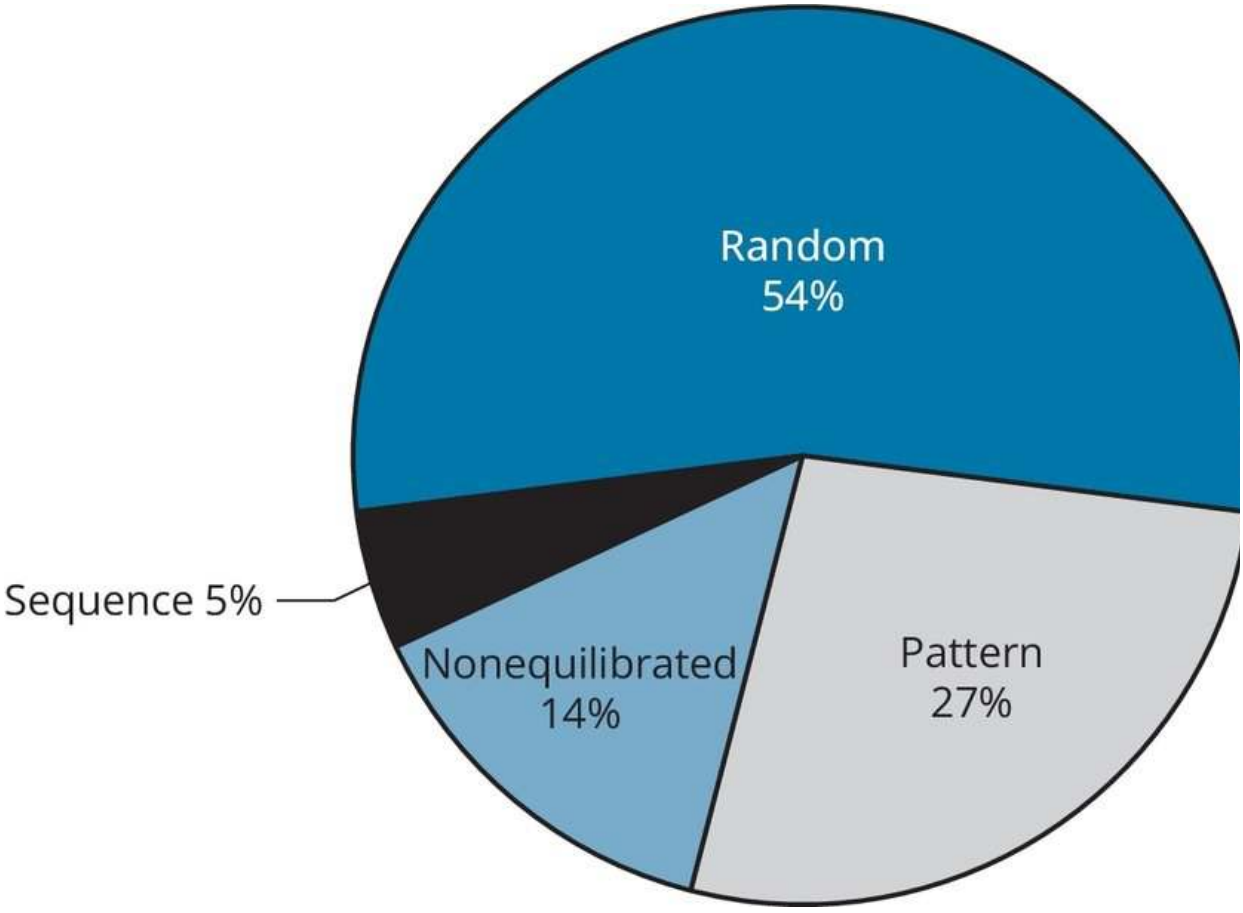


Figure 2.2 Pie Chart Showing the Percentages of the Lottery Ticket Variable for the Non−Problem Gamblers

Displaying Interval and Ratio Variables: Histograms and Frequency Polygons

Interval and ratio variables have values equally spaced along a numeric continuum and are identical to each other with one exception: Ratio scales possess a true zero point, for which the value of zero (0) represents the complete absence of the variable. The values of interval and ratio variables may be illustrated using histograms and frequency polygons, the choice of which depends on the number of values for the variable.

A **histogram** is a figure in which bars are used to represent the frequency of each value of a variable, with the different bars touching each other. The name and values of the variable are located on the X-axis, and the frequency or percentage for each value is on the Y-axis. Histograms are similar to bar charts, with one important difference: The bars in a histogram touch each other, indicating that the values of the variable are connected to each other along a numeric continuum.

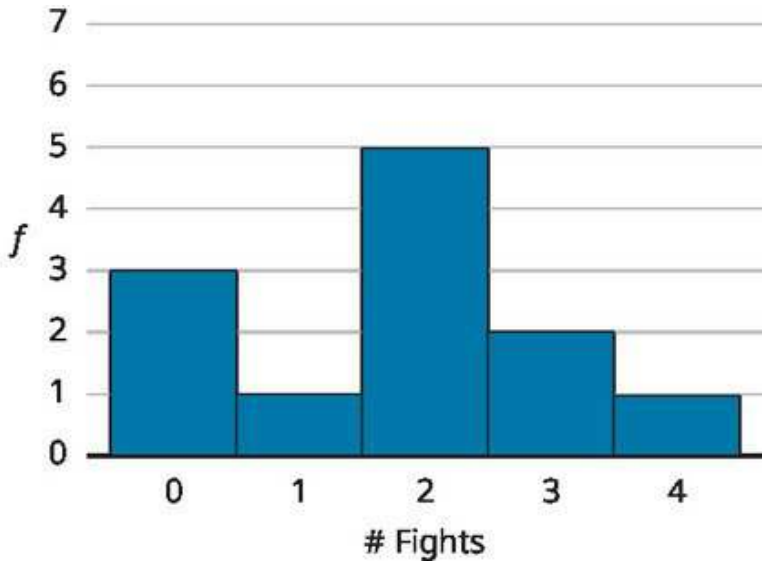
join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

As an example, imagine a teacher is interested in studying aggressive behavior in children and measures this variable by counting the number of fights each of her students is involved in over a 3-day period. “Number of fights” is a ratio variable rather than an interval variable because the value of 0 represents the complete absence of fights. Figure 2.3 provides a frequency distribution table and histogram for this variable.

A histogram is used when an interval or ratio variable consists of a relatively small number of values. But what if, for example, the number of fights could be anywhere from 0 to 20 rather than 0 to 5? In order to have the histogram fit on one page, the 21 bars would need to be very thin and crowded together, making the histogram confusing and difficult to read. In situations such as this, the values of the variable may be more clearly illustrated using a figure known as a **frequency polygon**, which is a line graph that uses data points to represent the frequency of each value of a variable, with straight lines connecting the data points. In a frequency polygon, the X-axis represents the name and values of the variable, while the Y-axis represents the frequency for each value.

# Fights	*f*
4	1
3	2
2	5
1	1
0	3
Total	12



▶ Long Description: Figure 2.3 Frequency Distribution Table and Histogram for the Number of Fights Variable

Figure 2.3 Frequency Distribution Table and Histogram for the Number of Fights Variable

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

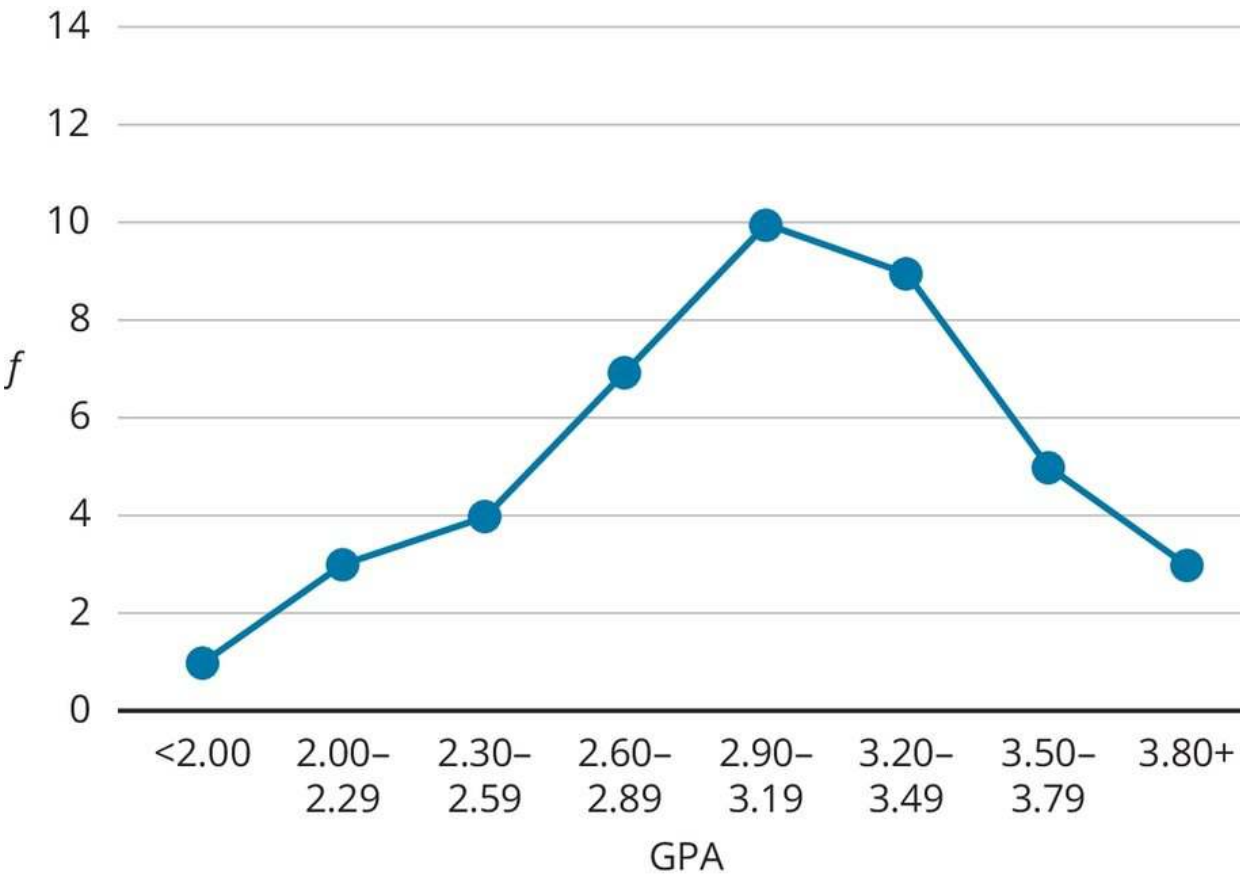


Figure 2.4 Frequency Polygon Showing the Frequency of Students in Each Interval of the Grade Point Average (GPA) Variable

Figure 2.4 shows a frequency polygon based on the grouped frequency distribution table of the earlier example of students' GPA (Table 2.7(b)). In this frequency polygon, the data points representing the frequencies of the eight GPA intervals are connected using straight, sharp lines. These lines indicate that the values of the variable can be connected together along a numeric continuum.

Although a frequency polygon like the one in Figure 2.4 may look jagged as it moves from one data point to the next, it provides a general sense of the shape of the distribution of scores for a variable. For example, looking at Figure 2.4 , you might conclude that the majority of GPAs in this sample are in the middle of the 1.00 to 4.00 range, with relatively low frequencies of students with either low or high GPAs.

Learning Check 2

Reviewing What You've Learned So Far

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



1. Review questions
    a. What is the difference between a frequency distribution table and a grouped frequency distribution table?
    b. For what levels of measurement would you create a pie chart, bar chart, histogram, or a frequency polygon?
    c. What is the difference between a bar chart and a histogram?
    d. Why would it be inappropriate to create a histogram or frequency polygon for a nominal or ordinal variable?
    e. Under what circumstances might you use a frequency polygon rather than a histogram to graph an interval or ratio variable?
2. For each of the following variables, identify whether you could use a bar chart, pie chart, histogram, and/or a frequency polygon to visually display the data.
    a. Miles per gallon (MPG)
    b. Living situation (on-campus, off-campus)
    c. Number of brothers and sisters
    d. Marital status (single, married, divorced)

Drawing Inappropriate Conclusions From Figures

A well-constructed figure can greatly facilitate the understanding of complex information. However, a poorly designed figure may not only be confusing, it can lead to inappropriate or misleading conclusions. Box 2.1 provides guidelines for constructing figures based on recommendations from the American Psychological Association (APA).

A simple example may demonstrate how altering the features of a figure can lead to different interpretations of the same set of data. Imagine a pollster assesses how voters with different political affiliations plan to vote on a ballot measure. She asks a sample of Democrats, Republicans, and Independents the question, “Do you plan on voting in favor of Measure A?” In examining her data, the pollster finds 56% of Democrats, 47% of Republicans, and 52% of Independents intend to vote “yes” on the measure.

Figure 2.5 graphs these three percentages using two bar charts; bar charts are used to display the data because the variable of interest (political affiliation) is a nominal variable (Democrat, Republican, or Independent). The bars don’t touch each other because the three political affiliations are distinct from each other and can’t be placed along a numeric continuum.

Although the two bar charts are similar, there is one important difference in how they are constructed. Following the guidelines in Box 2.1, the bar chart in Figure 2.5 (a) starts the Y-axis with the value 0% and ends with 100%. However, the bar chart in Figure 2.5 (b) starts with the value 45% and ends with 65%. By limiting the range of values of the percentages, the bar chart in Figure 2.5 (b) exaggerates differences in the heights of the three bars. As a result, someone looking at Figure 2.5 (b) may interpret the differences between the three political affiliations as being larger than someone looking at Figure 2.5 (a).

The saying “a picture is worth a thousand words” is particularly relevant for the presentation of statistical data. Because figures can provide a vivid representation of data, it’s the responsibility of researchers to portray their data in an accurate manner and to avoid creating images that may mislead or confuse.

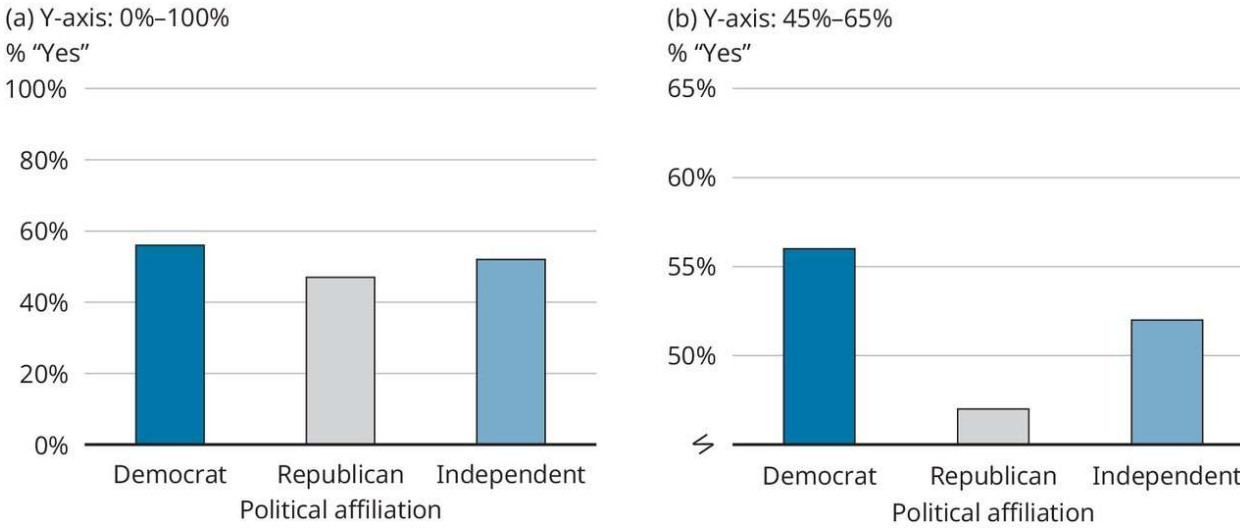
join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Box 2.1 Some APA Guidelines for Creating Figures

1. The figure should be formatted in a portrait rather than landscape page orientation (the reader should not turn the page sideways to read the table).
2. The figure is numbered using Arabic numbers (Figure 1, not Figure *I* or Figure *A*).
3. The figure includes a title that describes the contents of the figure. For example, it is not enough to name a figure simply " Figure 2.1 " or " Figure 2.1 . Bar chart."
4. Figures have a rectangular shape, with the height of the vertical (Y) axis two thirds to three fourths the length of the horizontal (X) axis.
5. The horizontal (X) axis contains the values of the variable; the vertical (Y) axis contains the frequency (*f*) or percentage (%) of the sample having each value of the variable.
6. The *X* - and *Y* -axes include descriptive labels written in both upper- and lowercase letters (as opposed to all capital letters).
7. The labels for the *X* - and *Y* -axes are short enough to fit on one line along the axes.
8. The label of the Y-axis is often placed parallel to the Y-axis (the letters face sideways rather than vertically) but can sometimes be seen on top of the y-axis for accessibility.
9. The Y-axis starts with the value zero (0) and is divided into equally spaced values or intervals.
10. The Y-axis is divided into 6 to 10 values (just enough values to accommodate bar length). The bars should neither be flattened along the bottom of the Y-axis nor bumped up against the top of the Y-axis.

Political affiliation	% "Yes"
Democrat	56%
Republican	47%
Independent	52%



▶ Long Description: Figure 2.5 Different Displays of the Same Data

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Figure 2.5 Different Displays of the Same Data

2.6 Examining Data: Describing Distributions

We've discussed how researchers construct tables and figures to organize their data. Looking at a table or figure enables researchers to describe the distribution of values or scores for a variable. Describing a distribution involves addressing questions such as, "Which values of the variable occur most often?," "How do the frequencies of the values of the variable change in relation to the most common values?," and "What is the amount of differences among the different values?"

Questions such as these pertain to three aspects of distributions: modality, symmetry, and variability. These three aspects of distributions are introduced in the following sections; they'll also be discussed in greater detail in the next two chapters of this book.

Modality

The first feature of a distribution of data for a variable is its modality. The **modality** of a distribution refers to the value or values of the variable that have the highest frequency or occur most often in a set of data. Using the lottery study as an example, identifying the modality of the lottery ticket variable involves answering the question, "Which of the four lottery tickets was chosen the most often?"

The modality of a distribution can be identified by looking at a frequency distribution table; the modality is the value(s) of the variable with the highest frequency (*f*) or percentage (%). The modality may also be determined by examining a figure and locating the tallest bar in a bar chart or histogram, the highest point in a frequency polygon, or the biggest "slice" in a pie chart. For the lottery ticket variable, the modality is the random ticket.

In identifying the modality of a distribution, it may be that one value or a limited number of values adjacent to one another clearly have the highest frequency. In such cases, the distribution is referred to as a unimodal distribution. A **unimodal distribution** is a distribution where one value of a variable occurs with the greatest frequency. Looking at the frequency distribution table (Table 2.4) or figure (Figure 2.1) for the lottery ticket variable, this distribution would be considered unimodal as one value, the random ticket, clearly occurred more often than the other three ticket choices.

In contrast to a unimodal distribution, it's possible that two distinct values may both have relatively high frequencies; if so, the distribution is referred to as a **bimodal distribution** , which is a distribution where two values clearly distinct from each other have the greatest frequency. A bimodal distribution sometimes resembles the two humps on a camel's back. As an example of a bimodal distribution, in April 2024 the Pew Research Center (www.pewresearch.org) asked a sample of 7,000 voters whether they felt Donald Trump had the mental fitness needed to be president, with five responses ranging from "extremely" to "not at all." The majority of voters chose one of the two extreme choices, either "extremely" (23%) or "not at all" (35%), with only 14% choosing the middle response of "somewhat."

And, in addition to unimodal and bimodal distributions, it's possible for a distribution to have more than two values with relatively equally high frequencies; this type of distribution is known as a **multimodal distribution** —a distribution where more than two values have the greatest frequency. In examining people's attitudes toward gun control, in May 2023, the news network CNN asked 1,227 people the question, "Do you think allowing gun owners to carry their guns in public places makes those places safer, less safe, or makes no difference in how safe a public place is?" They found people differed greatly in their responses, with 32% saying "makes public places safer," 36% saying

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

"makes public places less safe," and 32% saying "makes no difference," Figure 2.6 presents an example of a unimodal, bimodal, and multimodal distribution.

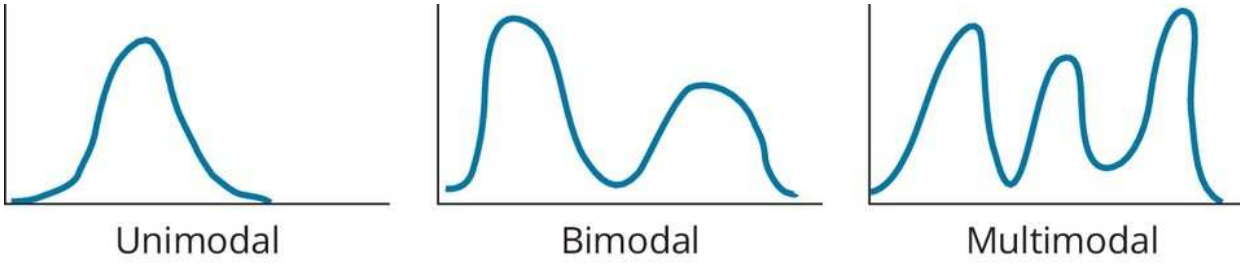


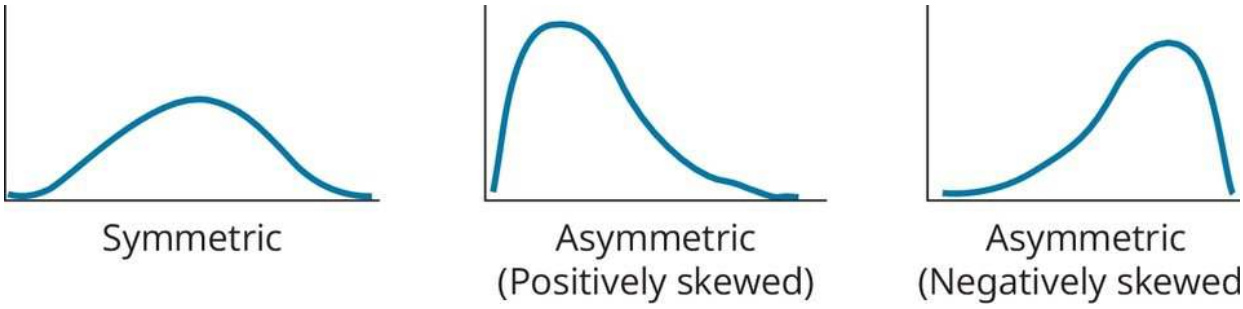
Figure 2.6 Distributions With Different Modalities (Unimodal, Bimodal, and Multimodal)

Symmetry

The second feature of a distribution of data is its symmetry. The **symmetry** of a distribution refers to how the frequencies of values of a variable change in relation to the most common or frequently occurring values. Describing the symmetry of a distribution means answering the question, "What is the shape of the distribution?"

The symmetry of a distribution can be determined by looking at the frequencies in a frequency distribution table or the bars or points in a figure. By doing so, a distribution may be described as being either symmetric or asymmetric. A **symmetric distribution** is one in which the frequencies change in a similar manner moving away in both directions from the most frequently occurring values. An example of a symmetric distribution is illustrated in Figure 2.7 . One way to determine whether a distribution is symmetric is to draw the distribution on a piece of paper and then fold the distribution vertically down the middle so that the two halves are on top of each other. A distribution is symmetric when the two halves are similar in shape such that they resemble mirror images of each other.

However, a distribution may also be asymmetric. An **asymmetric distribution** (also known as a *skewed* distribution) is one in which the frequencies change in a different manner moving away in both directions from the most frequently occurring values. In asymmetric distributions, the highest frequencies are located at one end of the distribution rather than in the middle, such that if you fold an asymmetric distribution down the middle, the shapes of the two halves do not look the same.



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Figure 2.7 Distributions With Different Symmetries (Symmetric, Positively Skewed, and Negatively Skewed)

In an asymmetric distribution, the frequencies take on the shape of a long tail as the values move toward the end of the distribution; scores in the tail of the distribution are often considered to be outliers. For example, in the 2022 Current Population Survey, the U.S. Census Bureau (www.census.gov) estimated that most Americans (63%) had annual incomes less than $50,000. However, a very small percentage (3%) were greater than $200,000; these higher incomes would be considered outliers.

Depending on the location of its tail, asymmetric distributions are referred to either as positively skewed or negatively skewed. A **positively skewed distribution** is one in which the higher frequencies are at the lower end of the distribution, with the tail on the upper (right) end of the distribution; a **negatively skewed distribution** is one in which the higher frequencies are at the upper end of the distribution, with the tail on the lower (left) end of the distribution. Figure 2.7 provides an illustration of a symmetric distribution and two asymmetric distributions (one positively skewed and one negatively skewed). Annual income, illustrated by the U.S. Census Bureau example mentioned earlier, is an example of a positively skewed distribution.

Variability

The third way in which distributions may be described is in terms of the amount or nature of differences among the different values for a variable. The **variability** of a distribution refers to the amount of differences in a distribution of data for a variable. The variability in a distribution is described by answering the question, “To what extent are the scores in the distribution different from each other?”

The amount of variability in a distribution can be estimated by looking at a frequency distribution table or figure. Imagine, for example, an instructor gives an 11-item quiz to her class and finds almost all of the students answer six, seven, or eight of the items correctly. In this situation, the distribution would resemble the shape of a tall mountain: thin and tall, with many of the scores residing in a few values of the variable. A distribution with a small amount of variability such as this one is referred to as a **peaked distribution** , defined as a distribution where much of the data are in a small number of values of a variable.

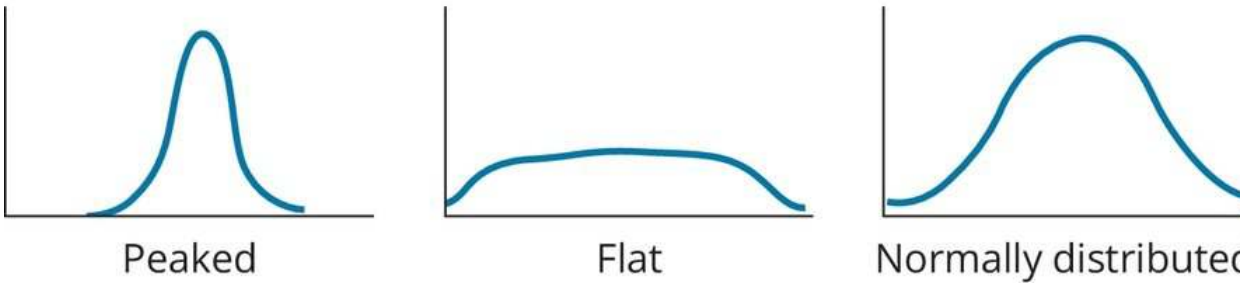


Figure 2.8 Distributions With Different Variability (Peaked, Flat, and Normally Distributed)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

On the other hand, what would the distribution look like if she had found her students' scores were distributed relatively evenly across the entire range of possible scores? In this situation, the distribution would not be tall and thin like a mountain but rather would be low and wide like a plateau. This type of distribution is referred to as a **flat distribution** , which is a distribution where the data are spread evenly across the values of a variable. The CNN gun control example described earlier is an example of a flat distribution.

In discussing modality, symmetry, and variability, distributions may have more than one mode, may be asymmetric (skewed), and may be peaked or flat. But what if a distribution has none of these qualities? A distribution that is unimodal, symmetric, and neither peaked nor flat may be described as "bell shaped." If you look at a bell, you'll notice that it is the highest in the middle, and as you move away from the top of a bell in both the left and right directions, the shape of the bell slopes downward in the same manner. A bell-shaped distribution for a variable is given a special name: a normally distributed variable.

A **normally distributed variable** is a distribution for a variable that is considered unimodal, symmetric, and neither peaked nor flat. Note that the word *normal* is a statistical concept and not a value judgment. In other words, asymmetric, peaked, or flat distributions are not considered "abnormal." The concept of normal distributions will be discussed in greater detail in Chapter 5. Figure 2.8 provides an illustration of three distributions with different amounts of variability: peaked, flat, and normally distributed.

Learning Check 3

Reviewing What You've Learned So Far



1. Review questions
    a. What are the three aspects used to describe distributions of variables?
    b. What is the difference between unimodal, bimodal, and multimodal distributions?
    c. What is the difference between symmetric and asymmetric (skewed) distributions?
    d. What is the difference between positively and negatively skewed distributions?
    e. What is the difference between peaked and flat distributions?
    f. What are the characteristics of a normally distributed variable?
2. For each of the following sets of data, determine whether the distribution of scores is unimodal, bimodal, or multimodal.
    a. 4, 2, 11, 6, 10, 2, 6, 13, 2, 10, 6, 10, 1
    b. 15, 17, 13, 18, 17, 14, 20, 18, 17, 14, 17
    c. 20, 10, 50, 10, 80, 30, 70, 10, 60, 10, 80, 0, 30, 80, 10, 100, 80
3. For each of the following sets of data, determine whether the distribution of scores is symmetric or asymmetric (skewed).
    a. 4, 5, 1, 4, 2, 4, 3, 5, 3, 4, 4, 3
    b. 220, 230, 250, 230, 200, 240, 210, 230, 270, 220, 240, 230, 240, 230
    c. 3, 1, 9, 2, 3, 6, 2, 7, 3, 5, 2, 6, 3, 2, 2
4. For each of the following sets of data, determine whether the distribution of scores is peaked or flat.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

a. 16, 10, 19, 12, 13, 15, 14, 17, 15, 20, 11, 13, 19, 16, 15
b. 12, 15, 12, 11, 13, 11, 12, 12, 14, 10, 12, 11, 16, 12, 13, 12, 11
c. 14, 12, 16, 14, 13, 15, 12, 14, 14, 15, 14, 14, 13, 14

2.7 Looking Ahead

This chapter has discussed why and how researchers examine data. Examining data by creating tables and figures requires a number of different skills, including organizational skills (to prepare a set of data for statistical analysis), communication skills (to present data in tables or figures), and conceptual skills (to describe a distribution). The next two chapters focus on mathematical skills related to the statistical analysis of data, describing how researchers describe distributions numerically rather than visually. Given that this is a statistics textbook, we will focus on calculating statistics to describe and analyze data. However, and we cannot say this strongly enough, it is *extremely* important to examine a set of data before conducting statistical analyses as this will prevent errors both in your calculations of statistics as well as your interpretation of the results of these calculations.

2.8 Summary

Before conducting statistical analyses, researchers typically examine their data to gain an initial sense of the data, detect data coding or data entry errors, identify ***outliers*** (rare, extreme scores that lie outside of the range of the majority of scores in a set of data), evaluate research methodology, and determine whether the data meet statistical criteria and assumptions.

One common way of examining variables involves creating tables known as ***frequency distribution tables*** (tables that summarize the number and percentage of participants that express the different values for a variable) and ***grouped frequency distribution tables*** (tables that group the values of a variable measured at the interval or ratio level of measurement into a small number of intervals and then provide the frequency and percentage within each interval). To create a grouped frequency table, the ***real limits*** of each interval (the values of the variable that fall halfway between the top of one interval and the bottom of the next interval) need to be identified.

Data for a variable may also be examined by creating figures such as bar charts, pie charts, histograms, and frequency polygons. ***Bar charts*** and ***pie charts*** are used for variables measured at the nominal or ordinal level of measurement; a bar chart uses bars to represent the frequency or percentage of a sample corresponding to each value of a variable, with the different bars not touching each other; a pie chart is a circle divided into proportions that represent the percentage of the sample corresponding to each value of a variable. ***Histograms*** and ***frequency polygons*** are used for variables measured at the interval or ratio level of measurement; a histogram is a figure in which bars are used to represent the frequency of each value of a variable, with the different bars touching each other; a frequency polygon is a line graph that uses data points to represent the frequency of each value of a variable, with lines connecting the data points.

From examining their data, researchers can describe the distribution of values or scores for a variable. There are three main aspects of distributions: modality, symmetry, and variability.

The ***modality*** of a distribution is the value(s) of the variable that have the highest frequencies or occur most often in a set of data. ***Unimodal distributions*** are distributions where one value occurs with the greatest frequency, ***bimodal distributions*** have two values with the greatest frequency, and ***multimodal distributions*** have more than two values with the greatest frequencies.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The ***symmetry*** of a distribution refers to how the frequencies of values of a variable change in relation to the most common or frequently occurring values of the variable. In ***symmetric distributions*** , the frequencies change in a similar manner, moving away in both directions from the most common values of the variable. In ***asymmetric (skewed) distributions*** , the frequencies change in a different manner, moving away in both directions from the most common values of the variable.

The ***variability*** of a distribution refers to the amount of differences in a distribution of data. In ***peaked distributions*** , much of the data are in a few values of a variable; in ***flat distributions*** , the data are spread evenly across the values of a variable. A ***normally distributed variable*** has a symmetric distribution that is neither peaked nor flat.

2.9 Important Terms

asymmetric (skewed) distribution

bar chart

bimodal distribution

cumulative percentage

flat distribution

frequency (*f*)

frequency distribution table

frequency polygon

grouped frequency distribution table

histogram

modality

multimodal distribution

negatively skewed distribution

normally distributed variable

outlier

peaked distribution

pie chart

positively skewed distribution

real limits

real lower limit

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

real upper limit

symmetric distribution

symmetry

unimodal distribution

variability

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

2.10 Formula Introduced in This Chapter

Percentage (%)

$$\% = \frac{f}{total\ number\ of\ scores} * 100$$

2.11 Using IBM SPSS Software *

Creating Data Files: Defining Variables and Entering Data

* IBM SPSS software. SPSS is a registered trademark of International Business Machines Corporation.

1. Open a new data file (See Figure 2.9).

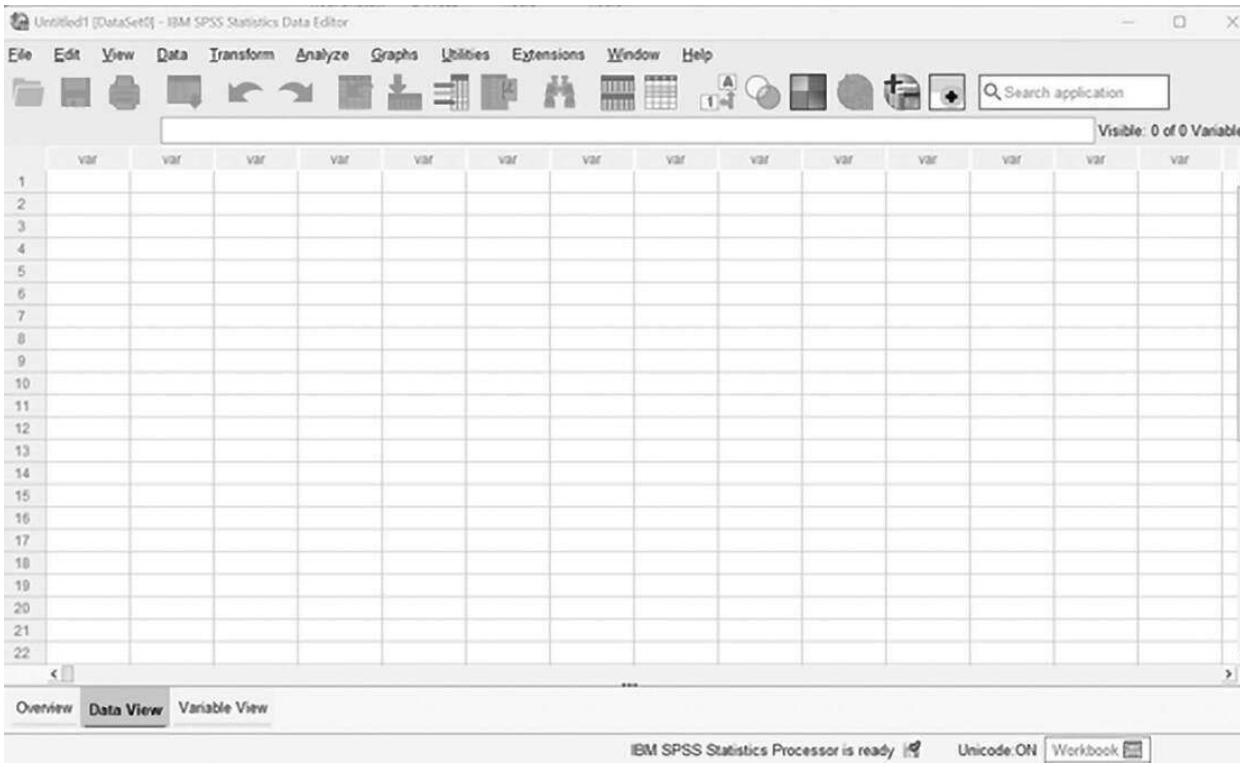


Figure 2.9 Data View

2. Define and name variable(s) (See Figure 2.10).

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

How? (1) Click **Variable View** tab, (2) click below **Name** , and (3) type name of variable in the cell.

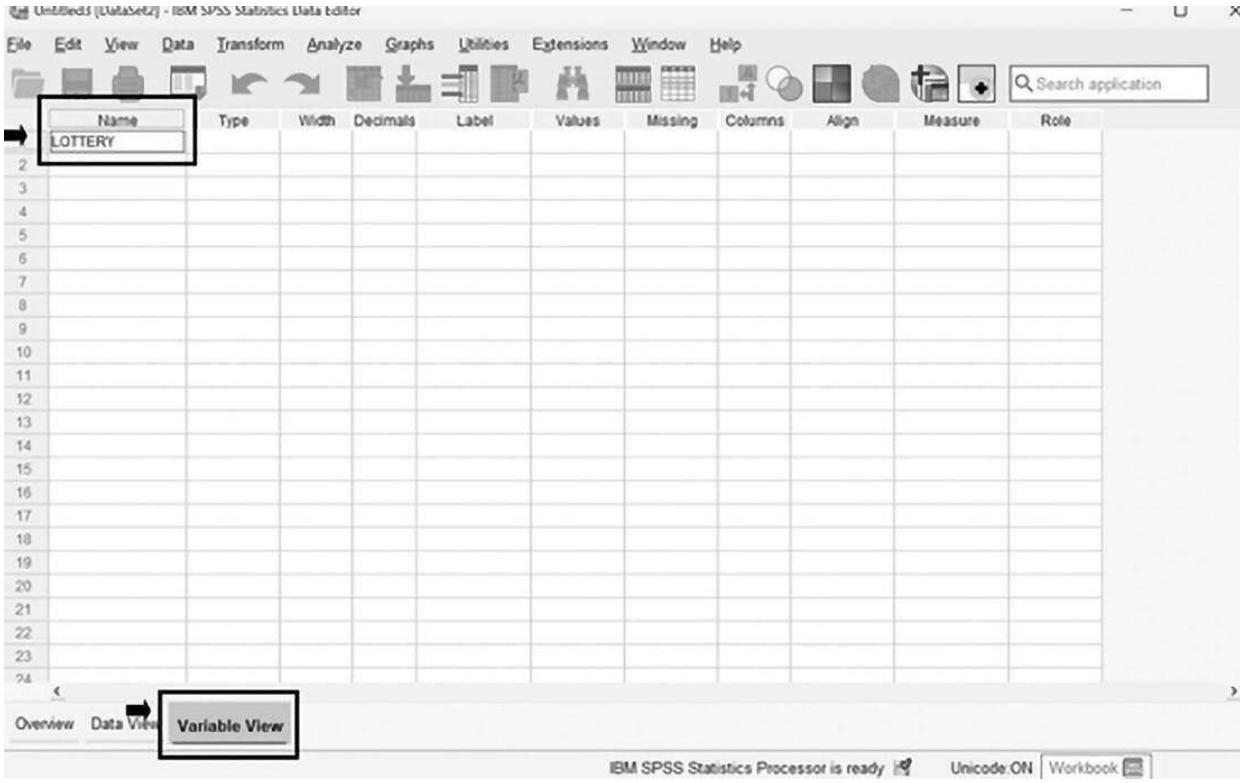


Figure 2.10 Variable View

3. Determine number of decimals for your variable(s) (See Figure 2.11).

How? Click desired number of decimal places in the **Decimals** box.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

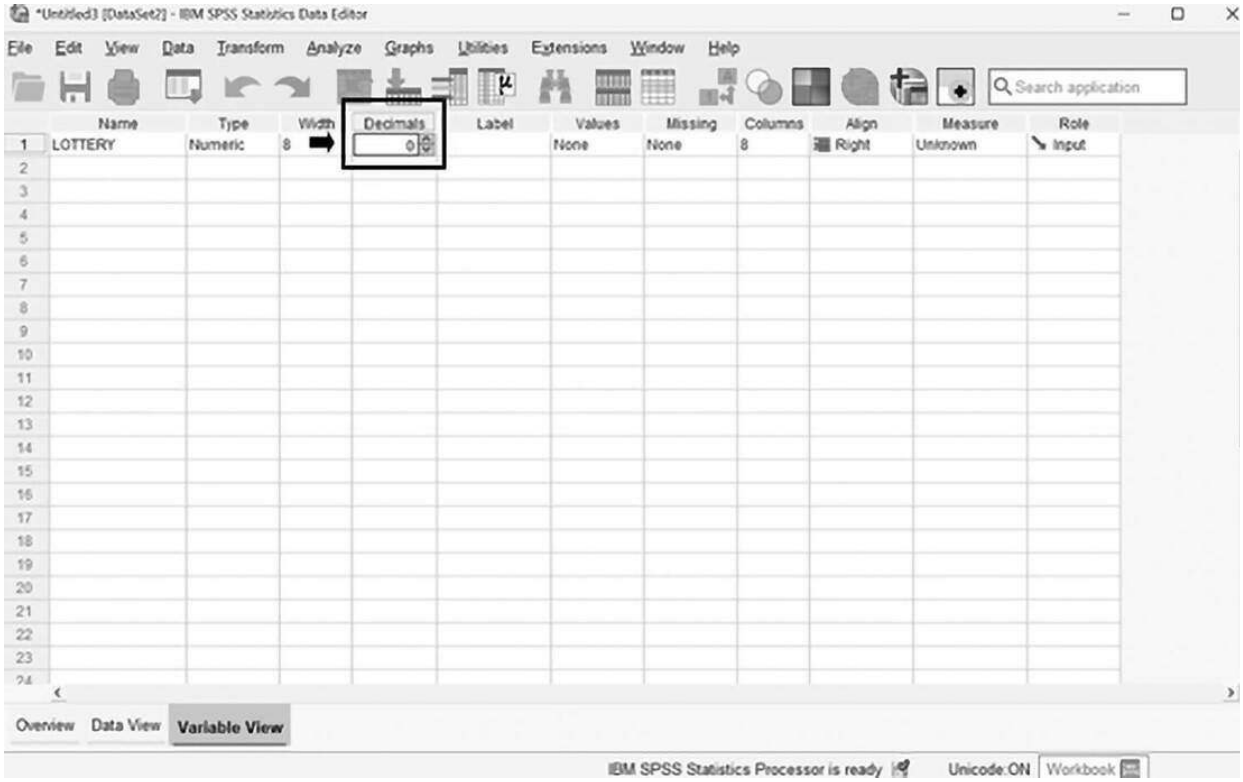


Figure 2.11 Decimals Box

4. Provide labels for your variable(s) (See Figure 2.12).

How? (1) Click below **Label** , and (2) type the label for the variable in the cell.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

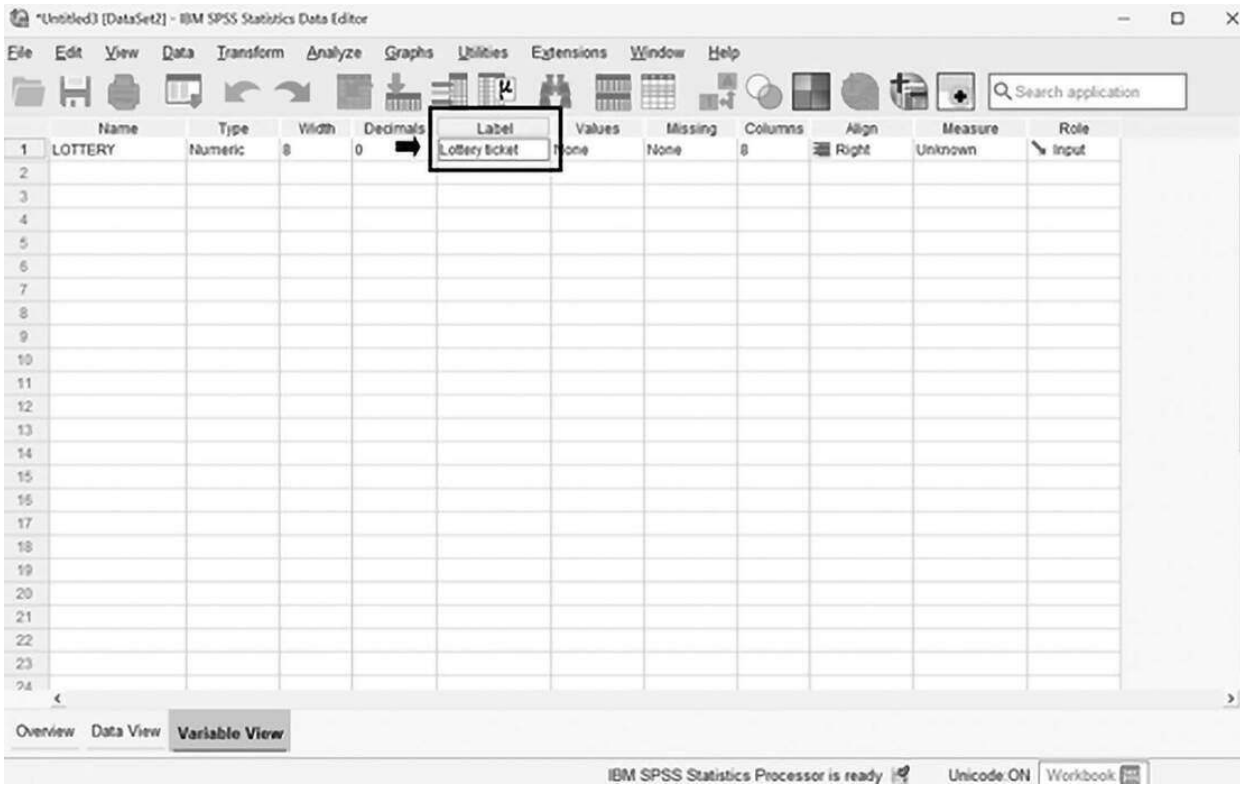


Figure 2.12 Label Box

5. Provide labels for values of nominal (categorical) variable(s) (See Figure 2.13).

How? (1) Click below **Values** , and (2) type labels for coded values of the variable.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

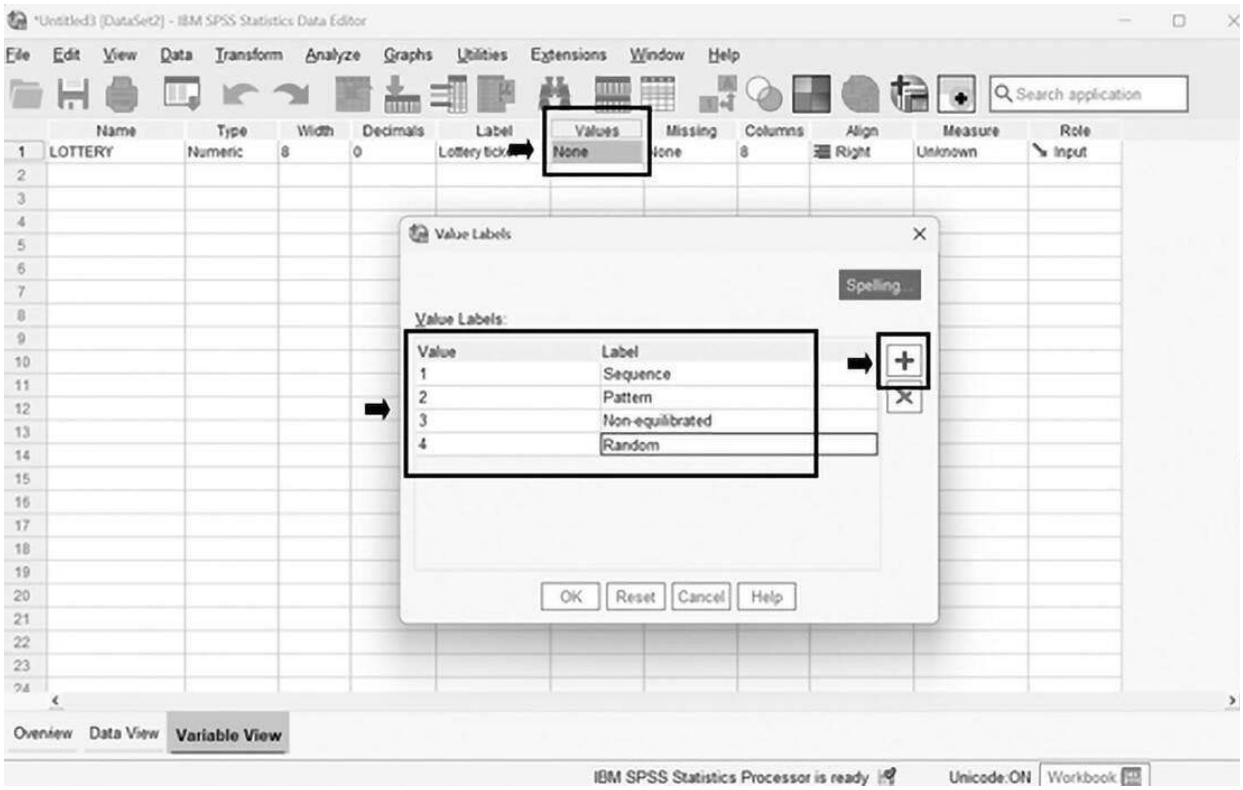


Figure 2.13 Nominal Variables

6. Enter data (See Figure 2.14).

How? (1) Click **Data View** tab and (2) enter data for the variable in the cells.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

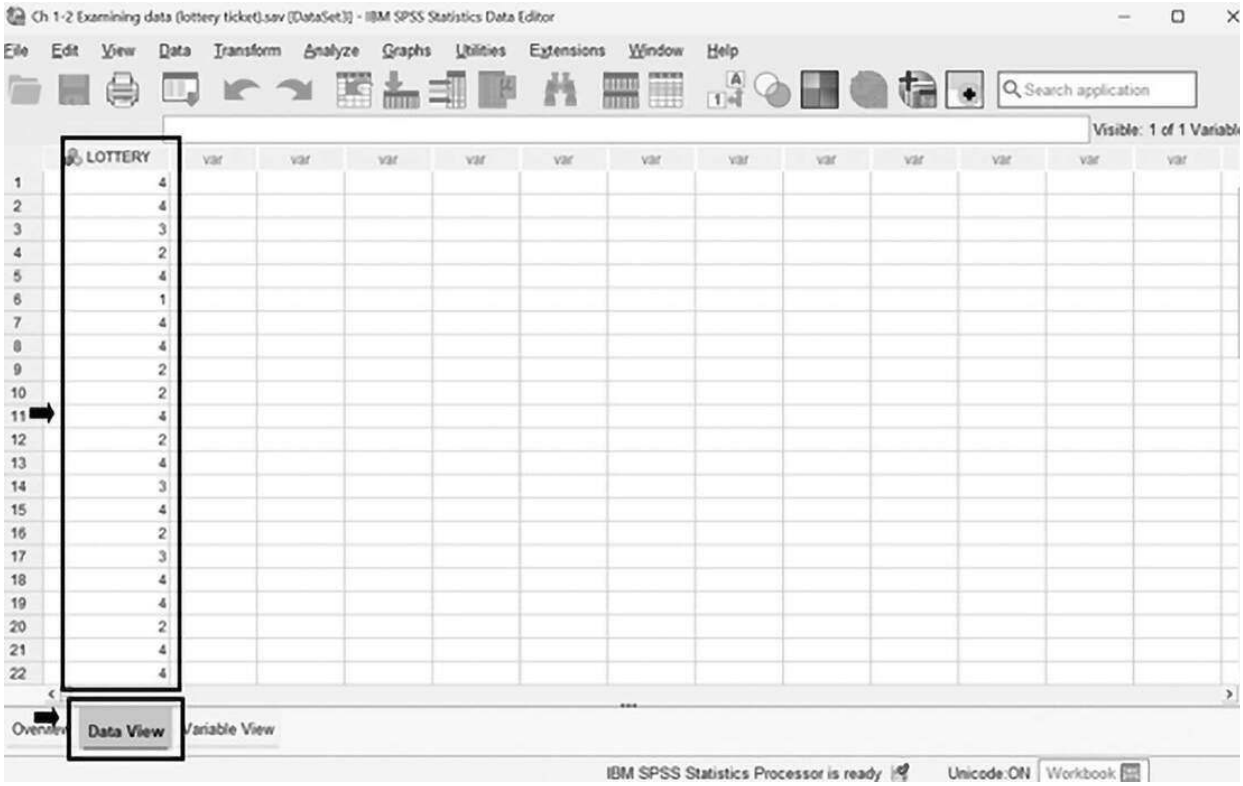


Figure 2.14 Enter Data View

Examining Data: Frequency Distribution Tables and Figures: The Lottery Study (2.1)

1. Select the frequency distribution procedure within SPSS (See Figure 2.15).

   How? (1) Click **Analyze** menu, (2) click **Descriptive Statistics** , and (3) click **Frequencies** .

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

To Jennifer , my kids Meagan and Eika , and my parents, Katsumi and Grayce Tokunaga.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

As a global academic publisher, Sage is driven by the belief that research and education are critical in shaping society. Our mission is building bridges to knowledge—supporting the development of ideas into scholarship that is certified, taught, and applied in the real world.

Sage's founder, Sara Miller McCune, transferred control of the company to an independent trust, which guarantees our independence indefinitely. This enables us to support an equitable academic future over the long term by building lasting relationships, championing diverse perspectives, and co-creating social and behavioral science resources that transform teaching and learning.

▶ Long Description:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

FUNDAMENTAL STATISTICS FOR THE SOCIAL AND BEHAVIORAL SCIENCES

Third Edition

Howard T. Tokunaga

San Jose State University



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

eISBN: 978-1-0718-6160-8

Copyright © 2027 by Sage.

All rights reserved. Except as permitted by U.S. copyright law, no part of this work may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without permission in writing from the publisher.

All third party trademarks referenced or depicted herein are included solely for the purpose of illustration and are the property of their respective owners. Reference to these trademarks in no way indicates any relationship with, or endorsement by, the trademark owner.



FOR INFORMATION:

2455 Teller Road, Thousand Oaks, California 91320, E-mail: orders@sagepub.com

3rd floor, HYLO, 103-105 Bunhill Row, London, EC1Y 8LZ, United Kingdom

10th Floor, Emaar Capital Tower 2, MG Road Sikanderpur, Sector 26, Gurugram, Haryana – 122002, India

18 Cross Street #10-10/11/12, China Square Central, Singapore 048423

Library of Congress Control Number: 2026943277

ISBN: 978-1-0718-6144-8

Printed in the United States of America

This book is printed on acid-free paper.

26 27 28 29 30 10 9 8 7 6 5 4 3 2 1

Acquisitions Editor: Anna Villarruel

Content Development Editor: Kenzie Offley

Editorial Assistant: Daria Terry

Production Editor: Syeda Aina Rahat Ali

Copy Editor: Michelle Ponce

Typesetter: Lumina Datamatics

Cover Designer: Candice Harman

Marketing Manager: Victoria Velasquez

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CONTENTS

Preface
Acknowledgments
About the Author
Chapter 1 Introduction to Statistics
Chapter 2 Examining Data: Tables and Figures
Chapter 3 Measures of Central Tendency
Chapter 4 Measures of Variability
Chapter 5 Normal Distributions
Chapter 6 Probability and Introduction to Hypothesis Testing
Chapter 7 Testing One Sample Mean
Chapter 8 Estimating the Mean of a Population
Chapter 9 Testing the Difference Between Two Means
Chapter 10 Errors in Hypothesis Testing, Statistical Power, and Effect Size
Chapter 11 One-Way Analysis of Variance (ANOVA)
Chapter 12 Two-Way Analysis of Variance (ANOVA)
Chapter 13 Correlation
Chapter 14 Linear Regression and Multiple Correlation
Chapter 15 Chi-Square
Appendix A: Tables
Appendix B: Review of Basic Mathematics
Glossary
References
Index

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

DETAILED CONTENTS

Preface

Acknowledgments

About the Author

Chapter 1 Introduction to Statistics

- 1.1 What Is Statistics?
- 1.2 Why Learn Statistics?
- 1.3 Introduction to the Stages of the Research Process
  - Developing a Research Hypothesis to Be Tested
    - Identifying a Question or Issue to Be Examined
    - Reviewing and Evaluating Relevant Theories and Research
    - Stating a Research Hypothesis: Independent and Dependent Variables
  - Collecting Data
    - Drawing a Sample From a Population
    - Determining How Variables Will Be Measured: Levels of Measurement
    - Selecting a Method to Collect the Data: Experimental and Nonexperimental Research Methods
  - Analyzing the Data
    - Calculating Descriptive Statistics
    - Calculating Inferential Statistics
  - Drawing a Conclusion Regarding the Research Hypothesis
  - Communicating the Findings of the Study
- 1.4 Plan of the Book
- 1.5 Looking Ahead
- 1.6 Summary
- 1.7 Important Terms
- 1.8 Exercises

Answers to Learning Checks

Learning Check 1

Learning Check 2

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Odd-Numbered Exercises

Chapter 2 Examining Data: Tables and Figures

2.1 An Example From the Research: Winning the Lottery

2.2 Why Examine Data?

Gaining an Initial Sense of the Data

Identifying Outliers

Evaluating Research Methodology

Determining Whether Data Meet Statistical Criteria and Assumptions

2.3 Examining Data Using Tables

What Are the Values of the Variable?

How Many Participants in the Sample Have Each Value of the Variable?

What Percentage of the Sample Has Each Value of the Variable?

2.4 Grouped Frequency Distribution Tables

Cumulative Percentages

2.5 Examining Data Using Figures

Displaying Nominal and Ordinal Variables: Bar Charts and Pie Charts

Displaying Interval and Ratio Variables: Histograms and Frequency Polygons

Drawing Inappropriate Conclusions From Figures

2.6 Examining Data: Describing Distributions

Modality

Symmetry

Variability

2.7 Looking Ahead

2.8 Summary

2.9 Important Terms

2.10 Formula Introduced in This Chapter

Percentage (%)

2.11 Using IBM SPSS Software *

Creating Data Files: Defining Variables and Entering Data

Examining Data: Frequency Distribution Tables and Figures: The Lottery Study (2.1)

2.12 Exercises

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 3 Measures of Central Tendency
3.1 An Example From the Research: The 10% Myth
3.2 Understanding Central Tendency
3.3 The Mode
Determining the Mode
3.4 The Median
Calculating the Median: Odd Number of Scores
Calculating the Median: Even Number of Scores
Calculating the Median When Multiple Scores Have the Median Value
3.5 The Mean
Calculating the Mean
Reporting Calculated Numbers
Rounding Calculated Numbers
Calculating the Mean From Frequency Distribution Tables
The Mean as a Balancing Point in a Distribution
The Sample Mean vs. the Population Mean: Statistics vs. Parameters
3.6 Comparison of the Mode, Median, and Mean
Strengths and Weaknesses of the Mode
Strengths and Weaknesses of the Median
Strengths and Weaknesses of the Mean
Choosing a Particular Measure of Central Tendency
3.7 Measures of Central Tendency: Drawing Conclusions
3.8 Looking Ahead
3.9 Summary
3.10 Important Terms
3.11 Formulas Introduced in This Chapter
Median (Odd Number of Scores)
Median (Even Number of Scores)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Mean
Mean (Frequency Distribution Table)
Mean (Balancing Point in a Distribution)
3.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 4 Measures of Variability
4.1 An Example From the Research: How Many Sometimes" in an Always"?
4.2 Understanding Variability
4.3 The Range
Strengths and Weaknesses of the Range
4.4 The Interquartile Range
Strengths and Weaknesses of the Interquartile Range
4.5 The Variance (s^2)
Definitional Formula for the Variance
Computational Formula for the Variance
Why Not Use the Absolute Value of the Deviation in Calculating the Variance?
Why Divide by N − 1 Rather Than N in Calculating the Variance?
4.6 The Standard Deviation (s)
Definitional Formula for the Standard Deviation
Computational Formula for the Standard Deviation
Eyeball Estimating" the Mean and Standard Deviation
4.7 Measures of Variability for Populations
The Population Variance (σ^2)
The Population Standard Deviation (σ)
Measures of Variability for Samples vs. Populations
4.8 Measures of Variability for Nominal Variables
The Index of Qualitative Variation
4.9 Measures of Variability: Drawing Conclusions

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

4.10 Looking Ahead
4.11 Summary
4.12 Important Terms
4.13 Formulas Introduced in This Chapter
Range
Interquartile Range
Variance (s^2) (Definitional Formula)
Variance (s^2) (Computational Formula)
Standard Deviation (s) (Definitional Formula)
Standard Deviation (s) (Computational Formula)
Population Variance (σ^2) (Definitional Formula)
Population Standard Deviation (σ) (Definitional Formula)
Index of Qualitative Variation (IQV)
4.14 Using SPSS
Calculating Measures of Central Tendency and Variability: The Frequency Rating Study (4.1)
4.15 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 5 Normal Distributions
5.1 Normal Distributions
What Are Normal Distributions?
Why Are Normal Distributions Important?
Is There Only One Normal Distribution?
5.2 The Standard Normal Distribution
What Is the Standard Normal Distribution?
Interpreting z-Scores: Distance From the Mean
Interpreting z-Scores: Relative Position Within the Distribution
Finding the Area Between the Mean and a z-Score

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Finding the Area Less or Greater Than a z-Score
Finding the Area Between Two z-Scores
5.3 Applying z-Scores to Normal Distributions
Transforming Normal Distributions Into the Standard Normal Distribution
Understanding the z-Score Formula
Example: Transforming SAT Scores Into z-Scores
Interpreting z-Scores: Distance From the Mean
Interpreting z-Scores: Relative Position Within Distribution
Finding the Area Between the Mean and a Score
Finding the Area Less or Greater Than a Score
Finding the Area Between Two Scores
Transforming Scores Into z-Scores: Drawing Conclusions
5.4 Standardizing Frequency Distributions
Example: Measuring Politicians' Behavior
Standardizing a Frequency Distribution
Interpreting Standardized Scores
Using Standardized Scores to Compare or Combine Variables
Characteristics of Standardized Distributions
When Can You Apply the Normal Curve Table to a Standardized Distribution?
5.5 Looking Ahead
5.6 Summary
5.7 Important Terms
5.8 Formulas Introduced in This Chapter
z-Score
Standardized Score
5.9 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 6 Probability and Introduction to Hypothesis Testing

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

6.1 A Brief Introduction to Probability
- What Is Probability?
- Why Is Probability Important to Researchers?
- Applying Probability to Normal Distributions
- Applying Probability to Binomial Distributions

6.2 Example: Making Heads or Tails of the Super Bowl

6.3 Introduction to Hypothesis Testing
- State the Null and Alternative Hypotheses (H_0 and H_1)
  - The Null Hypothesis (H_0)
  - The Alternative Hypothesis (H_1)
- Make a Decision About the Null Hypothesis
  - Defining a Low" Probability of a Statistic: Setting Alpha (α)
  - Identifying the Values of the Statistic With a Low Probability
  - Stating a Decision Rule
  - Calculating a Value of a Statistic
  - Making the Decision Whether to Reject the Null Hypothesis
- Draw a Conclusion From the Analysis
- Relate the Result of the Analysis to the Research Hypothesis
- Summary of Steps in Hypothesis Testing

6.4 Issues Related to Hypothesis Testing: An Introduction
- The Issue of Proof" in Hypothesis Testing
- Errors in Decision Making
- Factors Influencing the Decision About the Null Hypothesis
  - Sample Size
  - Alpha (α)
  - Directionality of the Alternative Hypothesis

6.5 Looking Ahead

6.6 Summary

6.7 Important Terms

6.8 Formula Introduced in This Chapter

Probability

6.9 Exercises

Answers to Learning Checks

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 7 Testing One Sample Mean
7.1 An Example From the Research: Do You Read Me?
7.2 The Sampling Distribution of the Mean
Characteristics of the Sampling Distribution of the Mean
7.3 Inferential Statistics: Testing One Sample Mean (σ Known)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: z-Test for One Mean
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the z-Test for One Mean
Summary
7.4 A Second Example From the Research: Unique Invulnerability
7.5 Introduction to the t-Distribution
Characteristics of the Student t-Distribution
7.6 Inferential Statistics: Testing One Sample Mean (σ Not Known)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: t-Test for One Mean
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the t-Test for One Mean

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Summary
7.7 Factors Affecting the Decision About the Null Hypothesis
Sample Size
Alpha (α)
Directionality of the Alternative Hypothesis
7.8 Looking Ahead
7.9 Summary
7.10 Important Terms
7.11 Formulas Introduced in This Chapter
z-Test for One Mean
Population Standard Error of the Mean $\left(\mathbf{i}t\sigma_{\mathbf{i}t\overline{X}}\right)$
Degrees of Freedom (df) for One Sample
t-Test for One Mean
Standard Error of the Mean $\left(\mathbf{i}ts_{\mathbf{i}t\overline{X}}\right)$
7.12 Using SPSS
Testing One Sample Mean (σ Not Known): The Unique Invulnerability Study (7.6)
7.13 Exercises
Answers to Learning Checks
Learning Check 2
Learning Check 3
Learning Check 5
Learning Check 6
Answers to Odd-Numbered Exercises
Chapter 8 Estimating the Mean of a Population
8.1 An Example From the Research: Salary Survey
8.2 Introduction to the Confidence Interval for the Mean
Using the Sampling Distribution of the Mean to Create the Confidence Interval for the Mean
8.3 The Confidence Interval for the Mean (σ Not Known)
State the Desired Level of Confidence
Calculate the Confidence Interval and Confidence Limits
Calculate the Standard Error of the Mean $\left(s_{\overline{X}}\right)$

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Calculate the Degrees of Freedom (df) and Identify the Critical Value of t

Calculate the Confidence Interval and Confidence Limits

Draw a Conclusion About the Confidence Interval

Probability of an Interval or the Probability of a Population Mean?

Summary

8.4 The Confidence Interval for the Mean (σ Known)

State the Desired Level of Confidence

Calculate the Confidence Interval and Confidence Limits

Calculate the Population Standard Error of the Mean $\left(\sigma_{\overline{X}}\right)$

Identify the Critical Value of z

Calculate the Confidence Interval and Confidence Limits

Draw a Conclusion About the Confidence Interval

Summary

8.5 Factors Affecting the Width of the Confidence Interval for the Mean

Sample Size

Estimating the Sample Size Needed for a Desired Confidence Interval

Level of Confidence

Choosing Between Different Levels of Confidence: Probability Versus Precision

8.6 Interval Estimation and Hypothesis Testing

Concerns About Hypothesis Testing

The Interpretation of Statistical Significance and Nonsignificance

The Influence of Sample Size on the Decision About the Null Hypothesis

Interval Estimation as an Alternative to Hypothesis Testing

8.7 Looking Ahead

8.8 Summary

8.9 Important Terms

8.10 Formulas Introduced in This Chapter

Confidence Interval for the Mean (σ Not Known)

Confidence Interval for the Mean (σ Known)

Estimated Sample Size for the Confidence Interval for the Mean

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

8.11 Using SPSS
Calculating the confidence interval for the mean (The salary survey study (8.1))
8.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 9 Testing the Difference Between Two Means
9.1 An Example From the Research: You Can Just Wait
9.2 The Sampling Distribution of the Difference
Characteristics of the Sampling Distribution of the Difference
9.3 Inferential Statistics: Testing the Difference Between Two Sample Means
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: t-Test for Independent Means
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the t-Test for Independent Means
Summary
9.4 Inferential Statistics: Testing the Difference Between Two Sample Means (Unequal Sample Sizes)
An Example From the Research: Kids' Motor Skills and Fitness
Inferential Statistics: Testing the Difference Between Two Sample Means (Unequal Sample Sizes)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Relate the Result of the Analysis to the Research Hypothesis

Assumptions of the t-Test for Independent Means (Unequal Sample Sizes)

Summary

9.5 Inferential Statistics: Testing the Difference Between Paired Means

Research Situations Appropriate for Within-Subjects Research Designs

An Example From the Research: Web-Based Family Interventions

Inferential Statistics: Testing the Difference Between Paired Means

Calculate the Difference Between the Paired Scores

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Draw a Conclusion From the Analysis

Relate the Result of the Analysis to the Research Hypothesis

Assumptions of the t-Test for Dependent Means

Summary

9.6 Looking Ahead

9.7 Summary

9.8 Important Terms

9.9 Formulas Introduced in This Chapter

Degrees of Freedom (df), Difference Between Two Sample Means

t-Test for Independent Means

Standard Error of the Difference $\left(s_{\overline{X}_1-\overline{X}_2}\right)$

Standard Error of the Difference (Unequal Sample Sizes)

Degrees of Freedom (df), Difference Between Paired Means

t-Test for Dependent Means

Standard Error of Difference Scores (s D)

9.10 Using SPSS

Testing the Difference Between Two Sample Means: The Parking Lot Study (9.1)

Testing the Difference Between Paired Means: The Web-Based Intervention Study (9.5)

9.11 Exercises

Answers to Learning Checks

Learning Check 1

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Answers to Odd-Numbered Exercises
Chapter 10 Errors in Hypothesis Testing, Statistical Power, and Effect Size
10.1 Hypothesis Testing vs. Criminal Trials
10.2 An Example From the Research: Truth or Consequences
Correct Decisions and Mistakes in Criminal Trials
10.3 Two Errors in Hypothesis Testing: Type I and Type II Error
Type I Error: The Risk in Rejecting the Null Hypothesis
What Is the Probability of Making a Type I Error?
What Is the Probability of Not Making a Type I Error?
Why Is Type I Error a Concern?
Type II Error: The Risk in Not Rejecting the Null Hypothesis
What Is the Probability of Making a Type II Error?
What Is the Probability of Not Making a Type II Error?
Why Is Type II Error a Concern?
Why Does Type II Error Occur?
Type I and Type II Error: Summary
10.4 Controlling Type I and Type II Error
Controlling Type I Error
Concerns About Controlling Type I Error
Controlling Type II Error
Increasing Sample Size
Raising Alpha (α)
Using a Directional Alternative Hypothesis (H1)
Increasing Between-Group Variability
Decreasing Within-Group Variability
Controlling Type I and Type II Error: Summary
10.5 Measures of Effect Size
Interpreting Inferential Statistics: The Influence of Sample Size
The Inappropriateness of “Highly Significant”

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Effect Size as Variance Accounted for

Measure of Effect Size for the Difference Between Two Sample Means: r^2

Interpreting Measures of Effect Size

Presenting Measures of Effect Size

Reasons for Calculating Measures of Effect Size

A Second Measure of Effect Size for the Difference Between Two Sample Means: Cohen's d

10.6 Looking Ahead

10.7 Summary

10.8 Important Terms

10.9 Formulas Introduced in This Chapter

Probability of Type I Error

Probability of Type II error

r Squared (r^2)

Cohen's d

10.10 Exercises

Answers to Learning Checks

Learning Check 1

Learning Check 2

Learning Check 3

Learning Check 4

Answers to Odd-Numbered Exercises

Chapter 11 One-Way Analysis of Variance (ANOVA)

11.1 An Example From the Research: It's Your Move

11.2 Introduction to Analysis of Variance (ANOVA)

Review: Testing the Difference Between the Means of Two Samples

Understanding Between-Group and Within-Group Variability

Relating Between-Group and Within-Group Variability

Introduction to Analysis of Variance (ANOVA) and the F-Ratio

Characteristics of the F-Ratio Distribution

11.3 Inferential Statistics: One-Way Analysis of Variance (ANOVA)

State the Null and Alternative Hypotheses (H_0 and H_1)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Value, and State a Decision Rule
Calculate a Statistic: F-Ratio for the One-Way ANOVA
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (R^2)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Summary
11.4 A Second Example: The Parking Lot Study Revisited
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Value, and State a Decision Rule
Calculate a Statistic: F-Ratio for the One-Way ANOVA
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (R^2)
Draw a Conclusion From the Analysis
The Relationship Between the t-Test and the F-Ratio
11.5 Analytical Comparisons Within the One-Way ANOVA
Planned Versus Unplanned Comparisons
Between-Group and Within-Group Variance in Analytical Comparisons
Conducting Planned Comparisons
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Conducting Unplanned Comparisons
Concerns Regarding Unplanned Comparisons
Methods of Controlling Familywise Error

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Controlling Familywise Error: The Tukey Test

11.6 Looking Ahead

11.7 Summary

11.8 Important Terms

11.9 Formulas Introduced in This Chapter

Degrees of Freedom for Between-Group Variance, One-Way ANOVA (df_{BG})

Degrees of Freedom for Within-Group Variance, One-Way ANOVA (df_{WG})

Between-Group Variance, One-Way ANOVA (MS_{BG})

Total Sample Mean, One-Way ANOVA $\left(\overline{X}_T\right)$

Within-Group Variance, One-Way ANOVA (MS_{WG})

F-Ratio, One-Way ANOVA (F)

Measure of Effect Size, One-Way ANOVA (R^2)

Relationship Between t-Test and F-Ratio for the One-Way ANOVA

F-Ratio, Analytical Comparison (F_{comp})

Between-Group Variance, Analytical Comparison $\left(MS_{\mathrm{BG}_{\mathrm{comp}}}\right)$

Measure of Effect Size, Analytical Comparison $\left(R^2_{\mathrm{comp}}\right)$

Critical Value, Tukey Test for Unplanned Comparisons (F_T)

11.10 Using SPSS

One-Way Analysis of Variance (ANOVA): The Chess Study (11.1)

Analytical Comparisons: Planned Comparisons—Explanation vs. Prediction (11.5)

11.11 Exercises

Answers to Learning Checks

Learning Check 2

Learning Check 3

Learning Check 4

Answers to Odd-Numbered Exercises

Chapter 12 Two-Way Analysis of Variance (ANOVA)

12.1 An Example From the Research: Vote—or Else!

12.2 Introduction to Factorial Research Designs

Testing Interaction Effects

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

An Example of Interaction Effect

An Example of No Interaction Effect

Testing Main Effects

The Relationship Between Main Effects and Interaction Effects

Factorial Designs vs. Single-Factor Designs: Advantages and Disadvantages

12.3 The Two-Factor (A × B) Research Design

Notational System for Components of the Two-Factor Research Design

Analyzing the Two-Factor Research Design: Descriptive Statistics

Notational System for Means in the Two-Factor Research Design

Examining the Descriptive Statistics

12.4 Introduction to Analysis of Variance (ANOVA) for the Two-Factor Research Design

F-Ratios in the Two-Factor Research Design

12.5 Inferential Statistics: Two-Way Analysis of Variance (ANOVA)

State the Null and Alternative Hypotheses (H_0 and H_1)

Main Effect of Factor A

Main Effect of Factor B

A × B Interaction Effect

Make Decisions About the Null Hypotheses

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Values, and State Decision Rules

Calculate Statistics: F-Ratios for the Two-Way ANOVA

Make Decisions Whether to Reject the Null Hypotheses

Determine the Levels of Significance

Calculate Measures of Effect Size (R^2)

Draw Conclusions From the Analyses

Relate the Results of the Analyses to the Research Hypothesis

12.6 Investigating a Significant A × B Interaction Effect: Analysis of Simple Effects

12.7 Looking Ahead

12.8 Summary

12.9 Important Terms

12.10 Formulas Introduced in This Chapter

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Degrees of Freedom for Main Effect of Factor A, Two-Way ANOVA (df_A)
Degrees of Freedom for Main Effect of Factor B, Two-Way ANOVA (df_B)
Degrees of Freedom for A × B Interaction Effect, Two-Way ANOVA ($df_{A\times B}$)
Degrees of Freedom for Within-Group Variance, Two-Way ANOVA (dfWG)
Between-Group Variance for Main Effect of Factor A, Two-Way ANOVA (MS_A)
Between-Group Variance for Main Effect of Factor B, Two-Way ANOVA (MS_B)
Sums of Squares for Between-Group Variance, Two-Way ANOVA (SS_{BG})
Between-Group Variance for A × B Interaction, Two-Way ANOVA ($MS_{A\times B}$)
Within-Group Variance, Two-Way ANOVA (MS_{WG})
F-Ratio for Main Effect of Factor A, Two-Way ANOVA (F_A)
F-Ratio for Main Effect of Factor B, Two-Way ANOVA (F_B)
F-Ratio for A × B Interaction, Two-Way ANOVA ($F_{A\times B}$)
Measure of Effect Size for Main Effect of Factor A, Two-Way ANOVA(R^2_A)
Measure of Effect Size for Main Effect of Factor B, Two-Way ANOVA(R^2_B)
Measure of Effect Size for A × B Interaction, Two-Way ANOVA(R^2_{AB})
12.11 Using SPSS
Two-Way Analysis of Variance (ANOVA): The Voting Message Study (12.1)
12.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 13 Correlation
13.1 An Example From the Research: It's Good for You!
13.2 Introduction to the Concept of Correlation
Describing the Relationship Between Variables
Nature of the Relationship
Direction of the Relationship
Strength of the Relationship

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Measuring the Relationship Between Variables: Correlational Statistics
Introduction to the Pearson Correlation Coefficient (r)
Calculating Correlational Statistics: The Role of Variance and Covariance
13.3 Inferential Statistics: Pearson Correlation Coefficient
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: Pearson Correlation Coefficient (r)
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (r^2)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the Pearson Correlation
A Computational Formula for the Pearson r
Represent the Covariance Between X and Y (SPXY)
Represent the Variance of Variable X (SSX)
Represent the Variance of Variable Y (SSY)
Calculate the Pearson Correlation Coefficient (r)
13.4 Correlating Two Sets of Ranks: The Spearman Rank-Order Correlation
Introduction to Ranked Variables
An Example From the Research: Do Students and Teachers Think Alike?
Inferential Statistics: The Spearman Rank-Order Correlation Coefficient
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
13.5 Correlational Statistics vs. Correlational Research
Does ANOVA = Causality?
“Correlation Is Not Causality”—What Does That Mean?
13.6 Looking Ahead

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

13.7 Summary
13.8 Important Terms
13.9 Formulas Introduced in This Chapter
Degrees of Freedom for Pearson Correlation (df)
Sum of Products (Definitional Formula) (SPXY)
Sum of Squares for the X Variable (Definitional Formula) (SSX)
Sum of Squares for the Y Variable (Definitional Formula) (SSY)
Pearson Correlation (Definitional Formula) (r)
Measure of Effect Size for the Pearson Correlation Coefficient (r^2)
Sum of Products (Computational Formula) (SPXY)
Sum of Squares for the X Variable (Computational Formula) (SSX)
Sum of Squares for the Y Variable (Computational Formula) (SSY)
Pearson Correlation (Computational Formula) (r)
Spearman Rank-Order Correlation (rs)
Measure of Effect Size for the Spearman Rank-Order Correlation $\left(r_s^2\right)$
13.10 Using SPSS
Pearson Correlation: The Picky Eater Study (13.1)
Spearman Rank-Order Correlation: The Business Student Study (13.4)
13.11 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Answers to Odd-Numbered Exercises
Chapter 14 Linear Regression and Multiple Correlation
14.1 Predicting One Variable From Another: Linear Regression
The Linear Regression Equation
Calculating the Linear Regression Equation
Calculate the Slope of the Equation (b)
Calculate the Y-Intercept of the Equation (a)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Report the Linear Regression Equation

Drawing the Linear Regression Equation

A Second Example: Put Down Your Pencils

Calculating the Linear Regression Equation

Drawing the Linear Regression Equation

14.2 Correlation With Two or More Predictors: Introduction to Multiple Correlation and Regression

The Limitations of Using Only One Predictor

Introduction to Multiple Correlation

Multiple Correlation With Two Predictors: An Example

Measuring the Relationship Between Two Predictors and a Criterion: The Multiple Correlation (R)

Testing the Multiple Correlation for Statistical Significance

Predicting a Criterion From Two Predictors: The Multiple Regression Equation

Conclusion

14.3 Looking Ahead

14.4 Summary

14.5 Important Terms

14.6 Formulas Introduced in This Chapter

Linear Regression Equation

Estimate of Slope (b)

Estimate of Y-Intercept (a)

Multiple Correlation Coefficient (R)

F-Ratio for the Multiple Correlation Coefficient (F)

Multiple Regression Equation

Regression Coefficient

Estimate of Y-Intercept (a) (Multiple Regression)

14.7 Using SPSS

Linear Regression: The Picky Eater Study (14.1)

Multiple Regression: The College Admission Example (14.2)

14.8 Exercises

Answers to Learning Checks

Learning Check 1

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 2

Answers to Odd-Numbered Exercises

Chapter 15 Chi-Square

15.1 An Example From the Research (One Categorical Variable): Are You My Type?

15.2 Introduction to the Chi-Square Statistic

Observed and Expected Frequencies: The Basis of the Chi-Square Statistic

Characteristics of the Distribution of Chi-Square Statistics

15.3 Inferential Statistic: Chi-Square Goodness-of-Fit Test

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Value, and State a Decision Rule

Calculate a Statistic: Chi-Square (χ^2)

Make a Decision Whether to Reject the Null Hypothesis

Determine the Level of Significance

Calculate a Measure of Effect Size (Cramér's φ)

Draw a Conclusion From the Analysis

Relate the Result of the Analysis to the Research Hypothesis

Assumptions Underlying the Chi-Square Statistic

Conducting the Chi-Square Goodness-of-Fit Test With Unequal Hypothesized Proportions

Calculate Expected Frequencies (f_e)

Calculate the Chi-Square Statistic (χ^2)

15.4 An Example From the Research (Two Categorical Variables): Seeing Red

15.5 Inferential Statistic: Chi-Square Test of Independence

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Value, and State a Decision Rule

Calculate a Statistic: Chi-square (χ^2)

Make a Decision Whether to Reject the Null Hypothesis

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Determine the Level of Significance
Calculate a Measure of Effect Size (Cramér's φ)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
15.6 Parametric and Nonparametric Statistical Tests
Parametric vs. Nonparametric Statistical Tests
Reasons for Using Nonparametric Statistical Tests
Examples of Nonparametric Statistical Tests
15.7 Looking Ahead
15.8 Summary
15.9 Important Terms
15.10 Formulas Introduced in This Chapter
Degrees of Freedom for the Chi-Square Goodness-of-Fit Test
Expected Frequency for the Chi-Square Goodness-of-Fit Test
Chi-Square Statistic (χ^2)
Cramér's φ for the Chi-Square Goodness-of-Fit Test
Degrees of Freedom for the Chi-Square Test of Independence
Expected Frequency for the Chi-Square Test of Independence
Cramér's φ for the Chi-Square Test of Independence
15.11 Using SPSS
Chi-Square Goodness-of-Fit Test—Equal Hypothesized Proportions: The Personality Study (15.1)
Chi-Square Goodness-of-Fit Test—Unequal Hypothesized Proportions: The Personality Study (15.3)
Chi-Square Test of Independence: The Color Preference Study (15.4)
Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Learning Check 6

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Odd-Numbered Exercises

Appendix A: Tables

Appendix B: Review of Basic Mathematics

I. Symbols

II. Terms

III. Addition, Subtraction, Multiplication, Division

A. All Positive Numbers

B. Adding All Negative Numbers

C. Adding a Combination of Positive and Negative Numbers

D. Subtracting Negative Numbers

E. Multiplying Two Negative Numbers

F. Multiplying a Combination of Positive and Negative Numbers

G. Dividing Two Negative Numbers

H. Dividing One Positive Number and One Negative Number

IV. Order of Operations

V. Fractions, Decimals, Percentages

A. Fraction

B. Decimal

C. Percentage

VI. Solving Simple Equations With One Unknown (X)

A. X Has a Value Subtracted From It (Solve for X Using Addition)

B. X Has a Value Added to It (Solve for X Using Subtraction)

C. X Is Multiplied by a Value (Solve for X Using Division)

D. X Is Divided by a Value (Solve for X Using Multiplication)

VII. Solving More Complex Equations With One Unknown (X)

A. X Is Multiplied by a Value and Then Has a Value Either Added to It or Subtracted From It (Solve for X Using Either Subtraction or Addition and Then Division)

B. X Is Divided by a Value and Then Has a Value Either Added to It or Subtracted From It (Solve for X Using Either Addition or Subtraction and Then Multiplication)

C. X Has a Value Added to It or Subtracted From It and Is Then Multiplied by a Value (Solve for X Using Division and Then Either Subtraction or Addition)

D. X Has a Value Either Added to It or Subtracted From It and Is Then Divided by a Value (Solve for X Using Multiplication and Then Either Subtraction or

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Addition)
Glossary
References
Index

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

PREFACE

A Few (?) Words About Me

When I entered college, I thought my path would take me to law school; however, taking an Intro to Psych class made me realize I enjoy the challenge of trying to understand the human mind. The college I attended, U.C. Santa Cruz, was a fairly unconventional university, but somehow in the midst of a sea of humanistic psychologists I became attracted to the empirical and methodological aspects of psychology by taking classes from instructors such as David Harrington and Dane Archer, who showed me that statistics could appeal to students if taught with a gentle, guiding approach that addresses questions relevant to students' lives. After graduating from college, I was able to get a job as a research assistant for a human resource consulting firm. Despite my lack of work experience, I was hired primarily as a result of having taken statistics and research methods courses, which taught me that learning statistics has benefits both inside and outside of the classroom.

Several years later, I started grad school at U.C. Berkeley, where two events critical to this book took place. First, serving as a teaching assistant, I found I really enjoyed helping students, particularly in statistics and research methods classes often approached with fear and suspicion. Second, I took graduate classes from Geoff Keppel, who had developed his own unique system for analyzing experimental research designs. His lectures and books showed me that statistics can be taught in a systematic way that highlights similarities rather than differences between different research situations. Geoff managed to transform something as daunting-sounding as a "3 x 2 x 4 research design" into the mathematical equivalent of playing with wooden toy alphabet blocks labeled "A," "B," and "C."

After leaving Cal, I was hired by San Jose State, where I teach an introductory statistics course aimed at students with a wide range of backgrounds, ability, and motivation. When I went to select a textbook for the course, for the first time I looked carefully at the wide range of offerings. What I found striking (and still find striking) was that the majority of books focused on providing formulas and very small sets of data designed to demonstrate how to correctly calculate the correct numbers from these formulas. Little emphasis, however, was given to what these numbers meant or implied. Given my own experiences learning and using statistics, I thought a book was needed that focusses on how statistics may be used to answer questions and test ideas within the larger research process.

Goal of this Book

This textbook, aimed primarily at students majoring in the behavioral and social sciences, includes many of the same topics and chapters as the majority of existing introductory statistics books. However, what I believe distinguishes this book is its thematically-based approach that places statistics within the research process. The primary purpose of this

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

book is to not just teach students how to calculate statistics but how to interpret the results of statistical analyses in light of a study's research hypothesis, and to communicate one's results and interpretations to a broader audience. Hopefully, this book will not only help students understand the purpose and use of statistics, but it will also give them a greater appreciation of how research studies are conceived, conducted, and communicated.

The book is intended for students with a limited knowledge of mathematics, statistics, and research design and methodology. Some common obstacles faced by students are their math anxiety, fear of conducting arithmetic calculations, and an inability to see the relevance of the material to their daily lives. To address these concerns, this books attempts to judiciously use mathematical terminology, write in a conversational manner, and utilize actual research situations rather than contrived sets of numbers.

Organization of the Book

The 15 chapters of this book may be placed into three general categories. Chapters 1 to 4 are designed to introduce students to the research process and how data that have been collected may be organized, presented, and summarized:

Chapter 1 (Introduction to Statistics) defines the concept of statistics, explains why statistics are important to learn, and describes the steps in the research process that are the framework for the rest of the book.

Chapter 2 (Examining Data: Tables and Figures) discusses the importance of examining data that have been collected and describes how tables and figures may be created to represent data.

Chapter 3 (Measures of Central Tendency) presents measures of central tendency (the mode, median, and the mean) and compares them in terms of their relative strengths and weaknesses.

Chapter 4 (Measures of Variability) presents measures of variability (range, interquartile range, variance, standard deviation, index of qualitative variation) and the concept of parameters (population variance and standard deviation).

Chapters 5 through 10 discuss the process of conducting statistical analyses to test research questions and hypotheses, as well as issues and controversies regarding this process:

Chapter 5 (Normal Distributions) introduces theoretical distributions, starting with normal distributions, before moving on to the standard normal distribution (normal curve).

Chapter 6 (Probability and Introduction to Hypothesis Testing) presents the structure for later chapters by introducing, defining, and illustrating steps in hypothesis testing.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Chapter 7 (Testing One Sample Mean) tests one sample mean using the *z* -test and *t* -test and introduces sampling distributions, the central limit theorem, and the Student *t* distribution.

Chapter 8 (Estimating the Mean of a Population) introduces interval estimation by developing the confidence interval for the mean and discusses interval estimation as an alternative to null hypothesis significance testing (NHST).

Chapter 9 (Testing the Difference Between Two Means) covers the *t* -test for independent means for both equal and unequal sample size and introduces within-subjects research designs using the *t* -test for dependent means.

Chapter 10 (Errors in Hypothesis Testing, Statistical Power, and Effect Size) discusses Type I and Type II error and introduces measures of effect size as a way of drawing appropriate conclusions about an analysis.

Chapters 11 to 15 discuss different statistical procedures used in research situations that vary in the number of independent variables as well as how the variables have been measured:

Chapter 11 (One-way Analysis of Variance (ANOVA)) discusses the one-way ANOVA, the relationship between the *t* -statistic and the F-ratio, and planned and post-hoc comparisons.

Chapter 12 (Two-way Analysis of Variance (ANOVA)) introduces factorial research designs, the calculation of F-ratios for main effects and interaction effects, and illustrates the use of simple effects analyses to test research hypotheses.

Chapter 13 (Correlation) defines and illustrates the concept of correlation, discusses the Pearson and Spearman correlation coefficients, and makes the distinction between correlational statistics and correlational research methods.

Chapter 14 (Linear Regression and Multiple Correlation) introduces the concept of regression, discusses linear regression, and works through an example of multiple regression with two predictor variables.

Chapter 15 (Chi-Square) describes the analysis of categorical variables using the chi-square goodness-of-fit test and the chi-square test of independence before ending the book with a brief introduction to nonparametric statistics.

In most cases, the chapters in this book share a uniform format and structure, centered on the research process, based on examples of actual published research studies addressing topics relevant to students' lives. For each example, the research hypothesis is clearly identified; after a brief description of how the data in the study were collected, the calculation and interpretation of both descriptive and inferential statistics are presented,

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ending with the extent to which the results of the statistical analyses support the research hypothesis and how these findings may be communicated to a broader audience.

Pedagogical Features

The chapters in this textbook contain several reoccurring features designed to enhance students' comprehension of the material. These include the following:

Periodic learning checks within chapters. Short learning checks, in the form of conceptual review questions and computational exercises, are provided throughout each chapter.

Explicit presentation of formulas and calculations. This book provides both definitional formulas to highlight their relation to the concept being measured and computational formulas to facilitate calculations. In applying these formulas, all of the steps in the calculation of statistics such as the *t* -test and F-ratio are included to help students in completing exercise and homework assignments.

Illustration of the presentation and communication of analyses. Every chapter includes a discussion and illustration of how statistical analyses may be reported as well how they may be communicated to a larger audience in both written and visual form.

Highlighted definition of terms. Every term and concept introduced in this book is boldfaced, immediately followed by an explicit definition.

End of chapter summaries and exercises. The end of each chapter contains a summary of critical points, a list of terms and formulas introduced in the chapter, and a large number of exercises, many of which refer to published research studies.

Calculation of statistics using SPSS. Many chapters include end-of-chapter content that provides an annotated presentation of how statistics may be calculated using the SPSS software. Screen shots of data files, menu selections, and output files are included and annotated to locate and identify relevant aspects of the analysis.

What's New to the 3rd Edition

I'd like to start by expressing my most heartfelt appreciation to the instructors and reviewers of the first two editions of this book as their evaluation, insights, criticisms, and suggestions were absolutely vital to the changes made in this third edition. I'd like to particularly thank the instructors who've adopted the book who were kind enough to spontaneously send me their thoughts and suggestions. It is my sincere hope that all of you see the incorporation of your input in this edition.

There were two main goals in this edition. The first goal was to make the writing as brief and concise as possible, manifested in the goal of reducing the length of each chapter by 5% to

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

10%. This goal was achieved by simplifying the language and shortening or eliminating what I, the book's adopters, and the book's reviewers felt was unnecessary detail or discussion. The second goal was to increase and update the exercises at the end of many of the chapters, including the use of recent published research as the basis of these exercises. Another goal was to update the demonstration of the use of SPSS to analyze data using the most recent version (Version 31 at the time of this book's publication).

More specifically, in terms of content, the following changes were made to these chapters:

Chapter 1 (Introduction to Statistics): The published research studies used to illustrate concepts in the text, learning checks, and exercises were added to and updated using topics relevant to students (e.g., fear of missing out (FOMO), cryptocurrency, legalization of cannabis, Gen Z, COVID, fake news).

Chapter 6 (Probability and Introduction to Hypothesis Testing): The Super Bowl example used to introduce and illustrate the process of hypothesis testing was updated.

Chapter 8 (Estimating the Mean of a Population): The main research example, which involves data from an actual national salary survey, has been updated throughout this chapter with more current salary data.

As with the first two editions, any and all questions, concerns, suggestions, and complaints you have about the book are most welcome and can be sent to me at howard.tokunaga@sjsu.edu .

A Few Tips for Students

To you, the college student about to read this book as part of taking a statistics course: "Welcome!" and "Great job!" We welcome you because you're about to embark on a semester-long journey that we hope enhances your skills and widens your perspective; we congratulate you because it's a journey not everyone is willing to take. We say to our students that by taking just one statistics course they are ahead of 95% of the population in their knowledge of statistics.

We also tell our students that our primary goal for them is not to become a statistics genius but just to become comfortable with the process of analyzing and interpreting data. In that regard, we agree with a 2012 *Forbes* magazine article entitled *The Six Classes That Will Make Any College Grad Employable* that noted "a good statistics class will help the student become comfortable pushing numbers around and avoiding false interpretations."

At the present moment, we know our encouragement may be of little comfort because you might be anxious about having to learn statistics. Some of you might be anxious about having to learn statistical concepts and formulas; others might be anxious because the research process seems pretty complicated. Talking with students who have taken our courses over the years has helped us assemble the following advice:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Master the Material Presented in the Early Chapters

Chapters 1 through 6 discuss how research is typically conducted, how data are summarized and described, and the process by which researchers conduct statistical analyses to test their ideas. It is absolutely critical for you to have a firm understanding of these chapters as they lay the groundwork for later chapters that discuss a variety of statistical procedures used by researchers to test hypotheses.

What does it mean to master this material? First, *read the chapters both before and after they're discussed in class* . Reading the chapters beforehand helps you identify anything that's unclear and have your questions addressed by your instructor; re-reading the chapters afterward helps confirm your understanding of the material. Next, *be able to define and explain key conceptspresented in these chapters* . These concepts are highlighted in bold-faced letters, and it's important to learn them when they're first introduced because they'll appear throughout the remainder of the book. Next, *do the learning checks within the chapters and the exercises at the end of the chapters* , because they include exercises and review questions you can ask yourself to assess your understanding of the material. Most importantly, *do not miss class during the early part of the semester* ! It's been our experience that students who miss critical lectures at the beginning of the term often have difficulty keeping up as the semester continues.

Review High School Algebra

If you happen to move beyond the introductory statistics course for which you're reading this book, you'll find that statistics are often calculated using computers and statistical software rather than hand calculations. Consequently, some statistics textbooks no longer include mathematical formulas but instead have students use statistical software or Microsoft Excel. However, this book takes a different approach for two main reasons. First, we believe that in order to understand the results of statistical analyses, one must understand their underlying foundation. Learning statistics via computer software can lead to a "brain-dead" approach in which students are at the mercy (rather than control) of their computers. We refer to this as the "I'm only as smart as my printout" method of learning statistics, also known as "Because this is what my computer told me."

Second, this book contains mathematical formulas and calculations for one simple reason: The statistics in this book are not hard to calculate. To assess whether you're prepared to read this book, take the following test:

1. Do you know how to add, subtract, multiply, and divide?
2. Do you know how to calculate the square or square root of a number?
3. Do you understand the order of operations and parentheses within mathematical formulas?

If your answers to these questions are "Yes," congratulations! You possess the ability to conduct *every* mathematical calculation in this book because *none* of the formulas involve geometry, trigonometry, or calculus. However, if you're unsure of your mathematical ability, we suggest you review your high school algebra. Also, there's an appendix at the end of this

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

book that includes the mathematical concepts and operations needed to conduct the statistical analyses in this book. We highly recommend you read this appendix to review and assess your mathematical skills.

Learn to Use a Statistical Calculator

Although learning the statistics in this book doesn't require the use of computers or software, you will need a hand calculator. Calculators with statistical capabilities are typically identified as "statistical" or "scientific" calculators. Although there's a wide selection of brands and models, know that the calculators most appropriate for this book cost only $10 to $15. We do *not* recommend you use or purchase a more expensive calculator! We prefer these simpler calculators because students are able to do their calculations more quickly and with fewer errors. We do suggest you purchase a calculator similar to the one your instructor uses as they will be more familiar with its features and idiosyncrasies.

In selecting a calculator, look closely at the keyboard; you'll need a calculator whose keys have labels such as $\bar{X}$, σ, and ΣX (Chapters 3 and 4 discuss what these symbols represent). When you go calculator shopping, you'll find calculators that have keys with labels such as ΣY and XY. Although calculators with these Y symbols are only slightly more expensive than calculators with just the X symbols, we do *not* recommend their purchase because these features are not needed to perform the calculations in this book. In fact, we've found students with Y calculators have a more difficult time performing simpler calculations.

Ask Questions

Conducting research is the process of asking and answering questions. Accordingly, throughout this book issues are often framed in the form of questions, which we hope encourages you to ask questions as well. The first person you should direct any questions to is yourself. As you work your way through the chapters, get into the habit of asking yourself, "What does this mean?" A good test of your level of comprehension is whether you can explain the material to yourself in a meaningful way.

You should also be sure to work with your instructor to confirm your level of understanding and clarify unanswered questions. Instructors often say to students, "The only bad question is the one you don't ask." Don't be hesitant about asking questions in class. Instructors will tell you they use one student's question to assess an entire class's level of understanding. This is what we call the "pencil test"—when one student asks a question, we look to see how many other pencils get raised, which reflects the number of students who had the same question. Asking questions enhances the learning experience for yourself, your classmates, and your instructor.

Form a Study Group

No matter how much we encourage students to ask questions in class, it seems they hesitate for fear of drawing attention to themselves. As a result, we turn to another well-

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

traveled saying: “Misery loves company.” Forming a study group with other students will help you keep up with reading assignments, confirm your comprehension of the material, clarify any confusion you have, prepare for exams, and provide social support. We do suggest that your group contain at least one person who will raise the group’s questions and concerns to the instructor during class or office hours.

Practice, Practice, Practice

The best way to learn the topics covered by this book is to obtain as much practice as possible. Throughout the chapters are examples in which the calculations have been worked out for you; in the middle and end of each chapter are a number of exercises. The answers to some of the exercises are provided at the back of this book, and your instructor has access to the answers to the other exercises. Do the end-of-chapter exercises even if your instructor does not assign them as homework in order to identify any reoccurring mistakes you make. We find that the vast majority of errors students make are simple computational errors, committed when students do the calculations too quickly.

Also note that the publisher of this book, Sage Publications, has developed a webpage for this book that contains a huge amount of supplemental tools such as videos and quizzes for each chapter. Find it by going to the Sage Publishing website **collegepublishing.sagepub.com** , searching for my name (Tokunaga), and looking for the **Resources** tab.

Enough cheerleading and preaching – let’s get started!

WELCOME!

GREAT JOB!

Additional Teaching Resources

Visit **collegepublishing.sagepub.com** , and navigate to the Resources tab on your book’s page to find the teaching materials designed to accompany this textbook. On this site you will find an array of materials that will save you time and to help you keep students engaged, including the following:

Learning management system cartridges that easily integrate with your course management system so student test results and graded assignments seamlessly flow into your gradebook

Test banks , aligned to Bloom’s Taxonomy, that provide a diverse range of test items, including multiple choice, true/false, and essay questions

Lecture notes that provide an outline and the key concepts in each chapter to aid in lecture preparation

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

PowerPoint ® slides that offer a flexible, accessible, and customizable solution for creating multimedia lectures

Tables and Figures from the book to support lecture preparation and class discussions

Sample course syllabi , including suggested models for structuring your course

Answers to even-numbered questions from the text help facilitate grading

SPSS datasets for use with exercises from the text.

Multimedia content , including original SAGE videos featuring tutorials with author Howard T. Tokunaga that bring concepts to life and appeal to diverse learners

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ACKNOWLEDGMENTS

This book has traveled a long journey, with many people playing both transient and persevering roles in its development. First and foremost, I'd like to thank my kids, Meagan and Will, for providing me perspective and inspiration amid all of the twists and turns we faced while this book was being written. I'm so very blessed to have you as my "burden." Away from home, I'm grateful to my graduate students at San Jose State—working on this book has given me much empathy for their thesis-related travails. However, they should be aware that the completion of this book ends the unspoken agreement that I not ask, "So how's your thesis?" if they don't ask, "So how's your book?" A number of friends have provided most welcome relief after long days staring at a computer screen, with particular thanks going to Ellie, Eileen, Susan, Bonnie, Ann, Lisa, and Terry. I would like to thank Charles Linsmeier and David Shirley for their many helpful comments and advice on earlier versions of this book, Julie West for her help in developing chapter exercises, and Youn-Jeong Choi and Naomi Ran for their careful examination of the chapters and exercises.

I would like to express my appreciation to everyone at SAGE who has played a role in shepherding this book to completion. First and foremost, I must express my gratitude to my editor, Vicki Knight. It has been a joy and an honor to work with someone with such a truly exceptional combination of experience, knowledge, perspective, and humor. In production I was fortunate enough to work with others on the SAGE team who made the process a pleasure: Yvonne McDuffee, editorial assistant; Katie Guarino, assistant editor; Nicole Elliott, executive marketing manager; Candice Harman, cover designer; and the copyeditor, Michelle Ponce. Furthermore, I am grateful to the reviewers who took the time to read the chapters, raise questions, and provide constructive comments and suggestions to assist me in clarifying and strengthening the manuscript. The following are the reviewers for third edition:

Erin M. Fekete, University of Indianapolis

Brooke Magnus, Boston College

Loretta Neal McGregor; Arkansas State University

Michael T. Moore, Adelphi University

Megan M. Ringel, St. Edward's University

Bettina Viereck, University of Hartford

Janice Weaver, Community College of Allegheny County

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ABOUT THE AUTHOR

Howard T. Tokunaga is professor of Psychology at San Jose State University, where he serves as coordinator of the MS Program in Industrial/Organizational (I/O) Psychology and teaches undergraduate and graduate courses in statistics, research methods, and I/O psychology. He received his bachelor's degree in psychology at U.C. Santa Cruz and his PhD in psychology at U.C. Berkeley. In addition to his teaching, he has consulted with a number of public sector and private sector organizations on a wide variety of management and human resource issues. He is author of *Moving from IBM SPSS to R and RStudio: A Statistics Companion* , and coauthor (with G. Keppel) of *Introduction to Design and Analysis* : *A Student's Handbook* . In his spare time, he enjoys the outdoors, museums, live theater, and being as physically, socially, and politically active as possible.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CHAPTER 1 INTRODUCTION TO STATISTICS

Learning Objectives

1.1 Define the concept of statistics.

1.2 Describe different reasons for learning statistics.

1.3 Describe the main steps in the research process and the tasks and concepts within each step.

In introducing this book to you, we assume you're a college student taking what is perhaps your first course in statistics to fulfill a requirement for your major or a general education requirement. If so, you may be asking yourself two questions:

What is statistics?

Why learn statistics?

The ultimate goal of this book is to help you begin to answer these two questions.

1.1 What Is Statistics?

You may not know it, but you encounter a variety of "statistics" in your day-to-day activities: the typical cost of going to college, the yearly income of the average college graduate, the average price of a car, and so on. So what exactly is "statistics"?

The Merriam-Webster dictionary defines **statistics** as *a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data* . When most people think about statistics, they focus on the "analysis" aspect of this definition—that is to say, they focus on numbers that result from analyzing data. However, statistics not only is concerned about how data are *analyzed* but also how data are *collected* and how the results of statistical analyses are *interpreted* and *communicated* . The main purpose of this book is to introduce, describe, and illustrate the role of statistics within the larger research process.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1.2 Why Learn Statistics?

There are many reasons why you should learn statistics. First, not only do you currently encounter statistics in your daily activities, but *throughout your life you are affected by the results of research and statistical analyses* . Which college you attend is based in part on test scores developed by researchers. You may have to take a personality or intelligence test to get a job. The choices of drugs and medicines available to you are based on medical research. Learning about statistics will help you become a more informed and aware consumer of research and statistical analyses that affect many aspects of your life.

A second reason for learning statistics is that *you may be required to interpret the results of statistical analyses* . Many college courses require students to read academic research journal articles. Evaluating published research is complicated by the fact that different people studying the same topic may come up with diverse or even opposing conclusions. Understanding statistics and their role in the research process will help you decide whether conclusions drawn in research articles are appropriate and justified.

Another reason for learning statistics is that *you might be required to do your own research* . College courses sometimes have students design and conduct mini-research studies; undergraduate majors might encourage students to do senior honors theses; graduate programs often require masters' theses and doctoral dissertations. Learning to collect and analyze data will help you address your own questions in an objective, systematic manner.

A final reason for learning statistics is that it may *help you in your future career* . The website Careercast.com conducts an annual survey in which they evaluate jobs on four dimensions: environmental factors, income, employment outlook, and stress. Recently, their #1 rated job was "data scientist," and the #2 rated job was "statistician." They noted, "A statistician's skill set can be used to break down and analyze large quantities of data. The demand for these skills spans a variety of industries, including marketing, banking, government, sports, retail, and even healthcare."

It's generally a good idea for students to maintain a healthy level of curiosity or even skepticism in regard to their education. However, we find that when it comes to learning statistics, students' feelings may be characterized as fear or anxiety. We understand these feelings, and we hope the benefits of learning statistics will become clear to you and help overcome any concerns you have.

1.3 Introduction to the Stages of the Research Process

Much of scientific research involves asking questions. Throughout this book, we'll examine a broad range of questions regarding human attitudes and behavior that contemporary researchers have asked and attempted to answer. The following are research questions we'll address in this chapter to introduce the stages of the research process:

> Does appealing to one's fear of missing out (FOMO) affect the intention to invest in cryptocurrencies?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Do pro-cannabis messages on social media lead to cannabis use?

Do parents differ in their awareness of their children's online activities?

Do nonbinary college students have the same experiences as cis men or cis women?

How might you try to answer questions such as these? You could base your answers on your personal beliefs, or you could adopt the answers given to you by others. But rather than relying on subjective beliefs and feelings, researchers in many academic disciplines test their ideas using the scientific method. The **scientific method** is a method of investigation that uses the objective and systematic collection and analysis of empirical data to test theories and hypotheses.

At its simplest, this book will portray the scientific method as consisting of five main steps or phases:

developing a research hypothesis to be tested,

collecting data,

analyzing the data,

drawing a conclusion regarding the research hypothesis, and

communicating the findings of the study.

Accomplishing each of these five steps requires completing a number of tasks, as shown in Figure 1.1 . Because this sequence of steps serves as the model for the research process we'll discuss throughout this book, each step is briefly introduced in the following sections. It's important to understand that the research process depicted in Figure 1.1 represents an ideal way of doing research. The "real" way, as you may discover in your own efforts, is often anything but a smooth ride but rather is filled with false starts, dead ends, and wrong turns.

Developing a Research Hypothesis to Be Tested

The initial stage—and the first step—of the research process is to develop a research hypothesis to be tested. A **research hypothesis** may be defined as a statement regarding an expected or predicted relationship between variables. A **variable** is a property or characteristic of an object, event, or person that can take on different values. One example of a variable is "U.S. state," a variable with 50 possible values (Alabama, Arkansas, etc.).

Research hypotheses are usually developed through the completion of several tasks:

identifying a question or issue to be examined,

reviewing and evaluating relevant theories and research, and

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

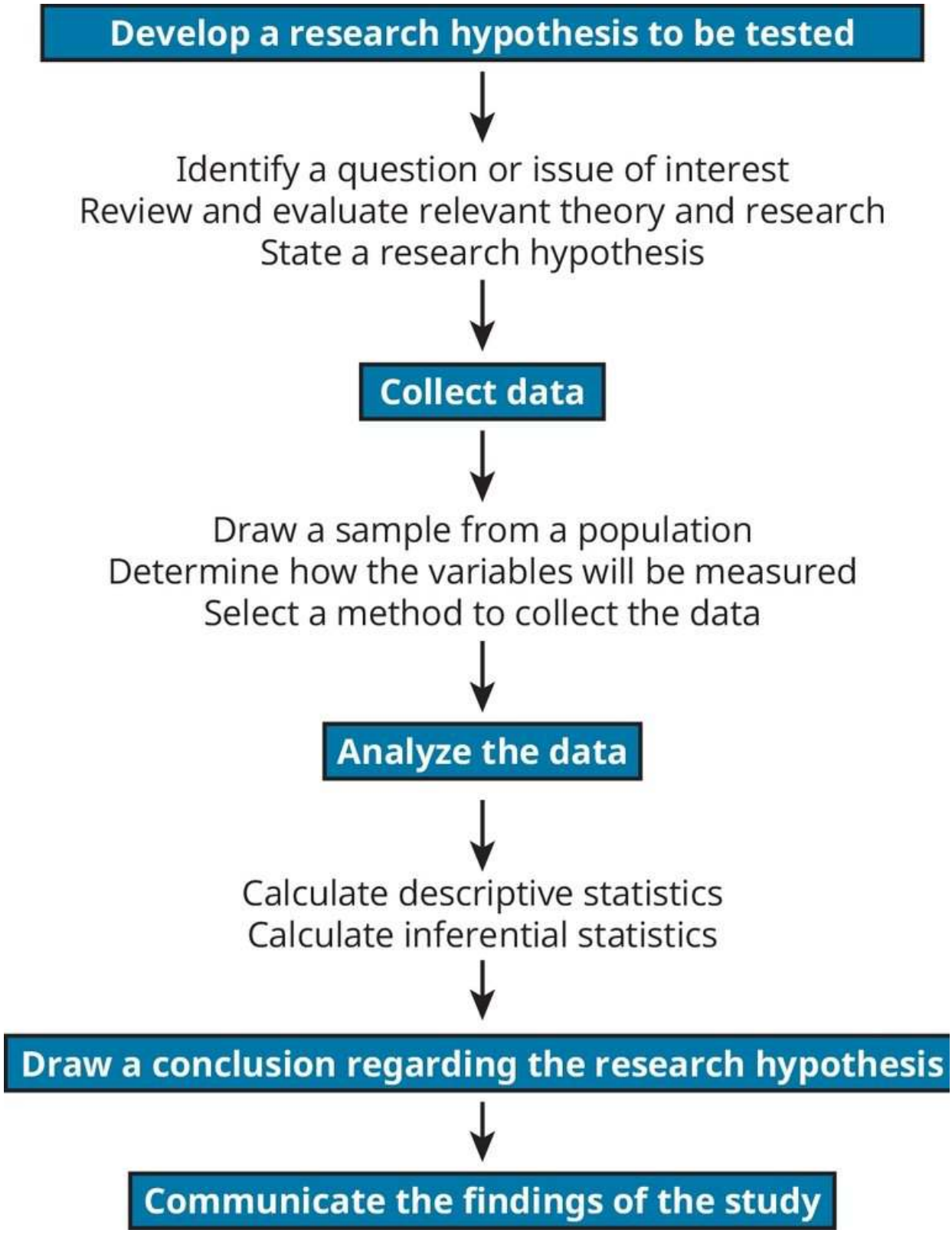
stating a research hypothesis.

Identifying a Question or Issue to Be Examined

Most research starts with a question posed by the researcher that often comes from the researcher's own ideas and daily observations. Although this may not seem terribly scientific, there's an advantage in using one's own experience as a starting point: People are generally much more motivated to explore issues that concern them personally. In teaching statistics, we advise students developing their own research projects to study something of interest to them. Conducting research can be tedious, difficult, and frustrating. At various points, you may ask yourself, "Why am I doing this?" Being able to provide a satisfactory answer to this question will help you overcome obstacles you encounter along the way.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

▶ Long Description: Figure 1.1 Steps in the Research Process Within the Scientific Method

Figure 1.1 Steps in the Research Process Within the Scientific Method

Reviewing and Evaluating Relevant Theories and Research

Beyond your own curiosity, research questions often arise from studying the theories, ideas, and research of others. A **theory** is a set of propositions used to describe or explain a phenomenon. The purpose of a theory is to summarize and explain specific facts using a logically consistent framework. Placing a question within a theoretical framework provides guidance and structure to research.

Reviewing and evaluating existing theories and research helps you decide whether it's worth the time and energy required to conduct a study. By seeing what others have done, you may conclude that the current way of thinking on a topic is incomplete or mistaken. By doing this review, you're able to ensure that the research you conduct adds to and improves upon an existing body of knowledge.

Stating a Research Hypothesis: Independent and Dependent Variables

Understanding and evaluating an existing literature not only helps develop a question of interest, but it may also lead to a predicted answer to that question. Within the scientific method, this answer is stated as a research hypothesis, defined earlier as a statement regarding an expected or predicted relationship between variables. Table 1.1 lists the research questions and research hypotheses for the studies listed at the beginning of this section. For example, the first research hypothesis states that "consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos."

One characteristic of research hypotheses such as those listed in Table 1.1 is that they identify the variables that are the focus of their research. As mentioned earlier, a variable is a property or characteristic with different values. Variables can be classified in several ways. In specifying a research hypothesis, researchers often speak in terms of "independent" and "dependent" variables. An **independent variable** may be defined as a variable manipulated or controlled by the researcher. A **dependent variable** , on the other hand, is a variable measured by the researcher that is expected to change or vary as a function of the independent variable. The independent variable is seen as the "cause" or "treatment," whereas the dependent variable is the "effect" or "outcome."

Researchers are interested in examining the effect of the independent variable on the dependent variable. Consider the first research hypothesis provided in Table 1.1 : "Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos." Here, the independent variable is the type of appeal consumers are exposed to: FOMO or non-FOMO. The dependent variable is the intention to invest in cryptos. In this study, the effect of the independent variable on the dependent variable is that

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

differences in the intention to invest in cryptos (the dependent variable) may “depend” upon which type of appeal (FOMO or non-FOMO) a consumer receives. Table 1.2 lists the independent and dependent variables for each of the research hypotheses in Table 1.1 .

In addition to identifying variables, a second characteristic of research hypotheses is that they specify the nature and direction of the relationship between variables. For example, the first research hypothesis in Table 1.1 states that “Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *higher* intention to invest in cryptos.” The word *higher* indicates the nature and direction of the relationship between the independent variable (type of appeal) and the dependent variable (intention to invest in cryptos). The direction of the relationship would not have been stated if the hypothesis had included the less specific phrase “will have a *different* intention,” which simply indicates that the two groups of consumers are not expected to be the same. Table 1.3 provides directional and nondirectional research hypotheses for the research studies from Table 1.1 .

Table 1.1 Examples of Research Questions and Hypotheses

Research Question	Research Hypothesis
Does appealing to one’s fear of missing out (FOMO) affect the intention to invest in cryptocurrencies? (Friederich et at., 2024)	Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.
Do pro-cannabis messages on social media lead to cannabis use? (Willoughby et al., 2024)	Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.
Do parents differ in their awareness of their children’s online activities? (Symons et al., 2017)	Mothers have more accurate knowledge of their children’s online activities than do fathers.
Do nonbinary college students have the same experiences as cis men or cis women? (Marx et al., 2024)	Nonbinary college students experience higher levels of victimization than do cis men or cis women.

The ability to form a directional research hypothesis is dependent on the state of the literature on the question of interest. If little or perhaps conflicting research has been conducted, researchers may not be able to form a directional hypothesis. For example, one study examining the relationship between exercise deprivation (not allowing people to exercise) and depression. reported, “We did not have a directional hypothesis, and when participants asked what we expected to find in this study, we replied: ‘We really are not sure since the results of earlier work on exercise deprivation are mixed’” (Mondin et al., 1996, p. 1200).

Table 1.2 Research Hypotheses and Their Independent and Dependent Variables

Research Hypothesis	Independent Variable	Dependent Variable
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher	Type of appeal	Intention to invest in cryptos

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Independent Variable	Dependent Variable
intention to invest in cryptos.		
Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.	Exposure to pro-cannabis messages	Intention to use cannabis
Mothers have more accurate knowledge of their children's online activities than do fathers.	Parent	Knowledge of children's online activities
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	Gender identity	Amount of victimization

Collecting Data

Once a research hypothesis has been formulated, researchers are ready to proceed to the second stage in the research process: collecting data. This step is seen as being composed of three tasks:

drawing a sample from a population,

determining how the variables will be measured, and

selecting a method by which to collect the data.

Table 1.3 Directional and Nondirectional Research Hypotheses

Directional Research Hypothesis	Nondirectional Research Hypothesis
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *higher* intention to invest in cryptos.	Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *different* intention to invest in cryptos.
Exposure to pro-cannabis messages on social media is related to *increased* intentions to use cannabis.	Exposure to pro-cannabis messages on social media is *related to* intentions to use cannabis.
Mothers have *more* accurate knowledge of their children's online activities than do fathers.	Mothers have *different* knowledge of their children's online activities than do fathers.
Nonbinary college students experience *higher* levels of victimization than do cis men or cis women.	Nonbinary college students experience *different* levels of victimization than do cis men or cis women.

Drawing a Sample From a Population

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The first step in collecting data is to identify the group of participants to which the research hypothesis applies. The group to which the results of a study may be applied or generalized is called a population. A **population** is the total number of possible units or elements that could potentially be included in a study. Although people often think the word "population" refers to everyone in the world, researchers typically define their populations more narrowly. For example, the population for the first research hypothesis in Table 1.1 could be "consumers," "U.S. consumers," "Georgia consumers," or "University of Georgia consumers." Researchers typically try to define their populations as broadly as possibly (e.g., "consumers" rather than "University of Georgia consumers") to maximize the applications or implications of their research.

Learning Check 1

Reviewing What You've Learned So Far



1. Review questions
    a. What are the five steps involved in the research process?
    b. Why is it useful to review and evaluate theories and research before conducting a study?
    c. What are the two main characteristics of research hypotheses?
    d. What is the difference between an independent variable and a dependent variable?
2. The following are several research hypotheses from published studies. For each research hypothesis, identify the independent and dependent variable.
    a. "College students will rate instructors who dress formally (i.e., business suit and tie) as having more expertise than instructors who dress casually (i.e., slacks and shirt)" (Sebastian & Bristow, 2008, p. 197).
    b. "It was expected tha. . . . greater amounts of television viewin. . . . would predict greate. . . . posttraumatic stress symptoms" (McLeish & Del Ben, 2008, p. 417).
    c. "Do borrowers who did not complete college have greater financial distress compared with . . . non-borrowers who did not complete college [or] borrowers who earned a degree?" (Jabbari et al., 2023, p. 6).
    d. "We expected that drivers would believe that texting while driving is more detrimental to their driving performance than browsing social media" (Hashash et al., 2019, p. 70).

It's extremely difficult to collect data from all members of a population. Imagine, for example, the time and money needed to collect information from every consumer in the United States! For this reason, researchers typically draw conclusions about populations based on information collected from a sample of the population. A **sample** is a subset or portion of a

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

population. Table 1.4 describes the samples used in the four studies introduced in Table 1.1 . As you can see from Table 1.4 , samples vary greatly in terms of their targeted population and size (the number of participants). The extent to which the results of a study can be generalized to the target population depends on the extent to which the sample is representative of the population.

Determining How Variables Will Be Measured: Levels of Measurement

The research hypotheses described in Table 1.1 involve variables such as type of appeal, intentions to use cannabis, and gender identity. To conduct a research study, the researcher must decide how to measure the variables stated in the research hypothesis. **Measurement** is the assignment of numbers or categories to objects or events according to rules.

For example, to measure the variable “height,” a researcher might use “number of inches from the ground in bare feet” as a form of measurement. To measure “success in college,” a student’s grade point average (GPA) may be obtained from school transcripts. “Self-esteem” might be measured by having people complete a questionnaire and be categorized as having either “low” or “high” self-esteem. As these examples demonstrate, different types of variables have different forms of measurement. Consequently, researchers have identified four distinct levels for measuring variables: nominal, ordinal, interval, and ratio.

The values of variables measured at the **nominal level of measurement** differ in category or type. The word *nominal* implies having to do with “names,” such that we use first and last names to distinguish between people. Gender identity is an example of a nominal variable, in that it consists of categories or types (such as cis male and cis female) rather than numeric values. In the first research hypothesis in Table 1.1 , the independent variable of “type of appeal” is measured at the nominal level with two categories: FOMO and non-FOMO.

Variables measured at the **ordinal level of measurement** have values that can be placed in an order relative to the other values. Rankings (such as finishing first, second, or third in a race), size (small, medium, large, or extra-large), and ratings (below average, average, or above average) are examples of ordinal variables. The values of ordinal variables indicate that one value represents relatively more or less of the variable than other values, but they don’t specify the precise difference between values. For example, although you can say that a “medium” drink is smaller than a “large” drink which is smaller than an “extra-large” drink, you can’t specify the exact difference in size between the three drinks.

Table 1.4 Research Hypotheses and Study Samples

Research Hypothesis	Study Sample
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.	“We recruited 150 participants from Amazon MTurk” (Friederich et at., 2024, p. 105).
Exposure to pro-cannabis messages on social media is	“Teens (N = 350) . . . needed to be between 13 and 17 years old and from Washington state” (Willoughby et

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Study Sample
related to increased intentions to use cannabis.	al., 2024, p. 185).
Mothers have more accurate knowledge of their children's online activities than do fathers.	"triadic data including reports from children aged 13 to 18, mothers and fathers . . . recruited in Flanders, the northern, Dutch speaking region of Belgium . . . a total of 357 valid triads was achieved" (Symons et al., 2017, p. 405).
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	"undergraduate college students aged 18–30 who were enrolled at 4-year institutions. This yielded a sample of 46,510 students" (Marx et al., 2024, p. 334).

The values of variables measured at the **interval level of measurement** are numbers equally spaced along a continuum. One example of an interval variable is the Fahrenheit scale of temperature. Here, a difference of 5 degrees has the same meaning anywhere along the scale; for example, the difference between 45°F and 50°F is the same as the difference between 65°F and 70°F. Many variables studied in the behavioral sciences (e.g., personality characteristics or attitudes) are considered to be measured at the interval level of measurement.

Variables at the **ratio level of measurement** are identical to interval variables, with one exception: Ratio scales possess what is known as a true zero point, for which the value of zero (0) represents the complete absence of the variable. Variables that describe a physical dimension (such as height, weight, distance, and time duration) typically have a true zero point. In the last research hypothesis in Table 1.1 , the variable "number of times victimized in last 12 months" is a ratio variable because it has a true zero point, where zero would indicate the complete absence of victimizations.

One advantage of ratio measurement versus interval is that ratio variables allow for a greater number of comparisons among values. Consider, for example, the value 6 for the ratio variable "inches." Not only is the difference between 4 and 6 inches the same as the difference between 6 and 8 inches (thereby involving addition and subtraction, which also can be used with interval variables), 6 inches is twice as long as 3 inches and half as long as 12 inches (involving multiplication and division). Multiplication and division comparisons cannot be made with interval variables. You cannot, for example, say that a temperature of 90°F is three times as much temperature as 30°F.

The boldfaced and italicized text in Table 1.5 illustrates how the four studies described in the earlier tables measured their independent and dependent variables. As you can see, the variables in these studies have been measured in a variety of ways at different levels of measurement.

Table 1.5 Research Hypotheses and Measurement of Variables

Research Hypothesis	Independent Variable	Dependent Variable

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Independent Variable	Dependent Variable
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.	Type of appeal **FOMO, non-FOMO (nominal)**	Intention to invest in cryptos **Likelihood of investing** **1-7 scale (1 = Highly unlikely to 7 = Highly likely) (ordinal)**
Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.	Exposure to pro-cannabis messages on social media **How often saw pro-cannabis messages (ratio)**	Intentions to use cannabis **Likelihood of using cannabis in the future (ratio)**
Mothers have more accurate knowledge of their children's online activities than do fathers.	Type of parent **Mother, Father (nominal)**	Accuracy of online activities **Accurate, somewhat accurate, inaccurate (nominal)**
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	Gender identity **Nonbinary, cis man, cis woman** **(nominal)**	Number of times victimized in last 12 months **(ratio)**

What difference does the level of measurement make?

A variable's level of measurement has important implications for researchers in that it influences how research hypotheses are stated and how data are analyzed. For example, imagine you study the variable "success in college." You could measure this variable at the *nominal* level of measurement by classifying students into one of two groups: successful or unsuccessful. Researchers typically use nominal variables when they are interested in studying differences between groups, such as, "Do successful and unsuccessful students differ in their study habits?"

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

You could instead measure success in college as an *ordinal* variable by ranking students from top to bottom, which enables a researcher to study relative differences with less concern for the precise size of these differences. For example, using ranked data, you could ask, “How similar are younger and older faculty members’ rankings of their students?”

To study success in college at the *interval* level, you could rate each student on a scale from 1 (low) to 100 (high). You could then use these ratings to address a research question such as, “Is there a relationship between faculty ratings of their students and students’ ratings of themselves?”

Finally, to measure success in college using the *ratio* level of measurement, you could use students’ salary after graduation (in dollars) as an indicator of success. Salary is a ratio variable because it contains a true zero point. In practice, ratio variables can be used to ask many of the same questions as those involving interval variables. For example, you could ask, “Is there a relationship between faculty members’ ratings of students and students’ salaries after graduation?”

As you see, how researchers measure their variables influences how they state their research questions and hypotheses. The level of measurement has another important implication: It helps determine the statistics researchers use to analyze their data. Certain statistics can be applied to variables measured at the interval or ratio levels of measurement, whereas others are appropriate for nominal or ordinal variables. As we move from chapter to chapter of this book, the level of measurement of the variables being analyzed will be noted.

Selecting a Method to Collect the Data: Experimental and Nonexperimental Research Methods

In addition to drawing a sample from a population and determining how the variables will be measured, the third step in collecting data is to determine the type of research method to use to collect the data on these variables. Research methods may be classified into two main types:

experimental research methods, and

nonexperimental research methods.

Experimental research methods.

Experimental research methods are methods designed to test causal relationships between variables—more specifically, whether changes in independent variables produce or cause changes in dependent variables. To make inferences about cause-effect relationships, researchers conducting an experiment must first eliminate other possible causes or explanations for changes in the dependent variable besides the independent variable. If it can be claimed that a variable other than the independent variable created the changes in the dependent variable, the research results are said to be confounded. A **confounding variable** is a variable related to an independent variable that provides an alternative explanation for the relationship between the independent and dependent variables.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

To illustrate the concept of confounding variables, consider the following question: Is there a relationship between a mother's ethnicity and the birth weight of her baby? Although research has shown differences in the birth weights of babies of different ethnicities, one study noted that some of this research did not eliminate the possibility that these differences may be partly due to differences between ethnic groups on factors associated with low birth weights, such as the mother's average age at the time of pregnancy and behaviors such as smoking and drinking (Ma, 2008). These other factors are considered confounding variables in that they provide alternative explanations for the relationship between ethnicity and babies' birth weights. Confounding variables make it difficult for researchers to draw firm conclusions regarding the relationship between the independent and dependent variable.

Learning Check 2

Reviewing What You've Learned So Far



1. Review questions
    a. What is the difference between a population and a sample?
    b. Within the research process, what is the relationship between populations and samples?
    c. What are the four levels of measurement? How do they differ?
2. For each of the following variables, name the scale of measurement (nominal, ordinal, interval, or ratio).
    a. Type of school attended (public, private)
    b. Probability of graduating college in 4 years (0% to 100%)
    c. Rating of television (good, better, best)
    d. Number of computers in home

Researchers minimize the influence of confounding variables in two main ways. First, researchers exercise experimental control by making the research setting (i.e., characteristics of the research participants, the instruments or measures administered, the methods used to collect data) the same for all participants. In the birth weight study, for example, the research design might control for the effect of parental smoking on birth weight by only including nonsmokers in the study sample, thereby excluding those who smoke.

Another way researchers exert control over the research setting is by including a condition known as a **control group** , a group of participants in an experiment not exposed to the independent variable. For example, if you conducted an experiment examining the effects of caffeine on one's health, one group (the experimental group) might be instructed to drink coffee while a second group (the control group) drinks decaffeinated coffee. By comparing the results of the two groups, researchers can then assess the impact of caffeinated coffee on the health of those who drink it.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Because researchers cannot possibly identify and control for all potential confounding variables, a second strategy used to minimize the influence of confounding variables is **random assignment** , which is assigning participants to each category of an independent variable in such a way that each participant has an equal chance of being assigned to each category. For example, a researcher could randomly assign a participant to the experimental or control condition by flipping a coin: heads for experimental, tails for control. The purpose of random assignment is to equalize or neutralize the effects of confounding variables by distributing them equally over the levels of the independent variable.

Nonexperimental (correlational) research methods.

Experimental research designs are one of the best tools researchers have for making causal inferences. However, the ability to make causal inferences requires a great deal of control that may not always be possible or desirable. For this reason, researchers often employ **nonexperimental research methods** , which are research methods designed to measure relationships between variables without having the ability to infer cause-effect relationships. These research methods are sometimes referred to as *correlational* research methods. It is important to understand that correlational research methods are different from correlational statistics we'll introduce in Chapter 13. Correlational research methods pertain to how data are *collected* ; correlational statistics pertain to how data are *analyzed* . Some of the most common types of nonexperimental research designs include quasi-experiments, survey research, observational research, and archival research.

Quasi-experimental research compares naturally formed or preexisting groups rather than randomly assigning participants to conditions. One example of this type of research was a study examining the effects of the COVID-19 pandemic on three groups of students: those living on campus, those living off campus but within commuting distance to school, and those living far away in another city, state, or country (Bisconer & McGill, 2023). This study is a quasi-experiment because the students were not randomly assigned to one of the three living conditions before collecting data.

Survey research methods obtain information directly from a group of people regarding their opinions, beliefs, or behavior. The goal of survey research, which can include using questionnaires and interviews, is to obtain information from a sample to represent or estimate the views of a larger population. One example of survey research examined college students' mental health by having them fill out a survey in which they reported their spirituality (the desire for truth, meaning, and significance of one's life) and their self-acceptance (knowledge and acceptance of one's self and limitations) (Nonis et al., 2024). Survey research is nonexperimental in that the variables being measured are not manipulated by the researcher.

Observational research is the systematic and objective observation of naturally occurring behavior or events. Observational research is often used to study phenomena that the researcher cannot or should not deliberately manipulate. For example, researchers studying interpersonal conflict among children would never force children to fight each other. One study resolved this issue by videotaping groups of adolescents discussing topics such as competition and peer pressure and then counting the number of times the adolescents teased or mocked each other (Connolly et al., 2015).

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Rather than observing or measuring behavior directly, **archival research** is the use of archives (records or documents of the activities of individuals, groups, or organizations) to examine research questions or hypotheses. One example of an archival research study studied the question of whether men are perceived to be more creative than women (Proudfoot et al., 2015). In this study, the 100 most viewed talks on TED.com were identified, and the researchers compared viewers' posted ratings of the talks given by men versus women.

Labeling variables in nonexperimental research.

One issue that often confuses students when learning about nonexperimental research is the practice of referring to the variables as the "independent variable" and the "dependent variable." For example, in the study described previously that involved biological sex and TED.com ratings, biological sex may be called the "independent variable." However, because biological sex isn't manipulated by the researcher, it would be inappropriate to say that biological sex "causes" differences in TED.com ratings. To avoid confusion, researchers sometimes refer to the independent variable and the dependent variable in a nonexperimental study as the **predictor variable** and **criterion variable**, respectively.

Analyzing the Data

Once data collection has been completed, the next step in the research process involves data analysis. This part of the research process addresses the primary topic of this book: statistics. Because this is the first chapter of this book, we won't describe analyzing data in detail. Instead, we introduce the notion that there are two main purposes of analyzing data: (1) to organize, summarize, and describe data that has been collected and (2) to test and draw conclusions about ideas and hypotheses. These two purposes are met by calculating two main types of statistics: descriptive statistics and inferential statistics.

Calculating Descriptive Statistics

Descriptive statistics are statistics used to summarize and describe a set of data for a variable. For example, you've probably heard of statistics used to describe or summarize aspects of our society, such as crime statistics and unemployment statistics. Using a research-related example, Caitlin Abar, a researcher at Brown University, conducted a study examining increases in students' alcohol-related behaviors after entering college (Abar, 2012). To measure students' level of alcohol use, she created a variable called "typical weekend drinking," which "was measured as the sum of drinks consumed on a typical Friday and Saturday within the past 30 days" (p. 22). One way to summarize students' responses to the typical weekend drinking variable (0, 1, 2, etc.) would be to calculate the mean, which is the mathematical average of a set of scores. The mean, described in Chapter 3, is an example of a descriptive statistic. The first part of this book describes a variety of descriptive statistics used by researchers to summarize data they've collected.

Calculating Inferential Statistics

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Besides summarizing and describing data, a second purpose of statistics is to test hypotheses. **Inferential statistics** are statistical procedures used to test hypotheses and draw conclusions from data collected during research studies. Using inferential statistics, researchers make inferences about what's believed to exist in populations based on data gathered from samples of the population.

As an example of an inferential statistic, let's return to the example of students' alcohol-related behaviors (Abar, 2012). This study was interested in seeing whether these behaviors were related to aspects of these students' relationships with their parents. One aspect was "alcohol communications," defined as "how often they discussed alcohol related topics with their parents" (p. 22). To test a hypothesis regarding the relationship between students' alcohol use and their alcohol communications, we could use an inferential statistic known as the Pearson correlation coefficient, which will be discussed in Chapter 13. The chapters in the last half of this book discuss a broad variety of inferential statistics, along with detailed instructions for analyzing and drawing conclusions from statistical data and presenting research findings.

Drawing a Conclusion Regarding the Research Hypothesis

Once statistical analyses have been completed, the next step is to interpret the results of the analyses as they relate to the study's research hypothesis. More specifically, do the findings of the analyses support or not support the research hypothesis? The word *support* in the previous sentence is very important. Students and researchers are sometimes tempted to say their findings "prove" their hypotheses are either "true" or "false." However, as we'll discuss in Chapter 5, because researchers typically don't collect data from the entire population, they can't know with 100% certainty whether their hypothesis has been proven to be true or false. A *huge* part of this book is learning how to draw appropriate conclusions from statistical analyses.

Communicating the Findings of the Study

Conducting research requires a variety of different skills: conceptual skills to develop research hypotheses, methodological skills to collect data, and mathematical skills to analyze these data. Another integral part of the research process is the communication skills needed to inform others about a study. Researchers must be not only skilled scientists but also effective writers.

Learning Check 3

Reviewing What You've Learned So Far



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1. Review questions
    a. What are the main differences between experimental and nonexperimental research methods?
    b. What are confounding variables? What do researchers do to minimize the effects of confounding variables?
    c. What are the main types of nonexperimental research methods?
    d. What are the relative strengths and weaknesses of experimental and nonexperimental research?
    e. What are the main purposes of descriptive and inferential statistics?

For many years, the American Psychological Association (APA), the professional association for psychologists, has provided researchers guidance on how to communicate the results of their research. In 1929, the APA published a seven-page article entitled, “Instructions in Regard to Preparation of Manuscript.” In contrast, the seventh edition of the *Publication Manual of the American Psychological Association* , published in 2020, consists of 428 pages! This increase in length highlights the challenge of communicating research effectively and appropriately.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1.4 Plan of the Book

The primary purpose of this chapter was to introduce statistics and place it within the larger research process. The remainder of this book discusses different statistics used by researchers to examine various questions of interest. Although it's critical for you to learn to correctly calculate statistics, it is equally important to understand the role of statistics within the research process and appreciate issues related to the use (and sometimes misuse) of statistics.

The first half of the book (Chapters 2 through 6) introduces conceptual and mathematical issues that are the foundation of statistical analyses. Chapters 2, 3, and 4 discuss how data may be examined, described, summarized, and presented in numeric, visual, and graphic form. Descriptive statistics are introduced and discussed in these chapters. Chapters 5 and 6 introduce characteristics of the inferential statistics used to test research hypotheses. The chapters in the second half of the book cover a number of different inferential statistics used with different types of variables.

In most cases, the chapters in this book share a uniform format and structure and are centered on the research process, using examples of actual published research studies that address questions with which you may be familiar. For each example, we'll clearly state the research hypothesis, briefly describe how the data in the study were collected, illustrate the calculation and interpretation of descriptive and inferential statistics, discuss the extent to which the statistical analyses support the research hypothesis, and show how to communicate one's research to a broader audience.

1.5 Looking Ahead

We began this chapter by defining statistics and providing reasons why learning about them may be of value to you. Next, we placed statistics within the stages of the larger research process, a process that will be the foundation of this textbook. As we've explained, the research process is centered on a research hypothesis, which is a predicted relationship between variables. Once a hypothesis has been stated, the next step is to collect data to be analyzed. However, before conducting statistical analyses on a set of data, there is preliminary work that must be completed. The next chapter will focus on methods used by researchers to examine data.

1.6 Summary

Statistics may be defined as a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data. As such, statistics not only is concerned about how data are analyzed but also recognizes the importance of understanding how data are collected and how the results of analyses are interpreted and communicated. There are a variety of reasons why students should learn statistics: Statistics are encountered in a wide variety of daily activities, students may be asked or required to read and interpret the results of statistical analyses in their courses, students may use statistics in conducting their own research, and statistics may help one's career.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Researchers conduct research using the **scientific method** of inquiry, a method of investigation that uses the objective and systematic collection and analysis of empirical data to test theories and hypotheses. The research process used within the scientific method of inquiry consists of five main steps: developing a research hypothesis to be tested, collecting data, analyzing the data, drawing a conclusion regarding the research hypothesis, and communicating the findings to the study.

A **research hypothesis** is a statement regarding an expected or predicted relationship between variables. Developing a research hypothesis involves identifying a question or issue to be examined, reviewing and evaluating relevant theories and research, and stating the research hypothesis to be tested in the study. A **theory** is a set of propositions used to describe or explain a phenomenon. A research hypothesis contains **variables** , which are properties or characteristics of some object, event, or person that can take on different values. More specifically, a research hypothesis states the nature and direction of a proposed relationship between an **independent variable** (a variable manipulated or controlled by the researcher) and a **dependent variable** (a variable measured by the researcher that is expected to change or vary as a function of the independent variable).

Collecting data involves drawing a sample from a population, determining how the variables will be measured, and selecting a method to collect the data. A **population** is the total number of possible units or elements that could potentially be included in a study; a **sample** is a subset or portion of a population. Variables can be measured at one of four levels of measurement: **nominal** (values differing in category or type), **ordinal** (values placed in an order relative to the other values), **interval** (values equally spaced along a numeric continuum), or **ratio** (values equally spaced along a numeric continuum with a true zero point).

There are two main types of methods used to collect data: **experimental research methods** , designed to test whether changes in independent variables produce or cause changes in dependent variables, and **nonexperimental (correlational) research methods** , designed to examine the relationship between variables without the ability to infer cause-effect relationships.

In experimental research, researchers are concerned about possible **confounding variables** , which are variables related to independent variables that provide an alternative explanation for the relationship between independent and dependent variables. To minimize the impact of confounding variables, a research study may exercise experimental control over various aspects of the situation, including the use of a **control group** , which is a group of participants not exposed to the independent variable the research is designed to test, or **random assignment** , which involves assigning participants to each category of an independent variable in such a way that each participant has an equal chance of being assigned to each category.

Examples of nonexperimental research methods are **survey research** , in which information is directly obtained from a group of people regarding their opinions, beliefs, or behavior; **observational research** , which involves the systematic and objective observation of naturally occurring events; and **archival research** , which uses archives (records or documents of the activities of individuals, groups, or organizations) to examine research questions or hypotheses. In nonexperimental research, the independent variable and

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

dependent variable may be referred to as the **predictor variable** and **criterion variable**, respectively.

Once data have been collected, the next step is to analyze them using two main types of statistics: **descriptive statistics**, which summarize and describe a set of data, and **inferential statistics**, which are statistical techniques used to test hypotheses and draw conclusions from data collected during research studies.

Once the statistical analyses have been completed, the results of the analyses are interpreted regarding whether they support or not support the study's research hypothesis. The final step in the research process is to communicate the study to a broader audience.

1.7 Important Terms

archival research

confounding variable

control group

criterion variable

dependent variable

descriptive statistics

experimental research methods

independent variable

inferential statistics

interval level of measurement

level of measurement (nominal, ordinal, interval, ratio)

measurement

nonexperimental research methods

observational research

population

predictor variable

quasi-experimental research

random assignment

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

research hypothesis

sample

scientific method

statistics

survey research

theory

variable

1.8 Exercises

1. Identify the independent and dependent variable for each of the following research questions and hypotheses from published articles.
    a. “We hypothesize that finance majors will embody narcissistic traits more than other business majors” (Wood et al., 2021, p. 924).
    b. “Men are expected to employ physical characteristics (particularly those that are directly related to sex) more often than women in selecting dating candidates” (Hetsroni, 2000, p. 91).
    c. “It is hypothesized that perceptions of emotional pain and support needs will be higher for the community of the Black victim than those of the officers’” (Kahn et al., 2024).
    d. “Gen Z students are more likely to believe that ‘Group projects bring out the worst in people’ than Millennials” (Schlee et al., 2020, p. 145).
    e. “We hypothesized to find . . . higher stress and anxiety post COVID in the Portuguese group in contrast with the British ” (Penalba-Sanchez et al., 2022, p. 322).
    f. “We hypothesized that consumers would think that a ‘sale price’ presentation would generate a greater monetary savings than an ‘everyday low price’ presentation” (Tom & Ruiz, 1997, p. 403).
2. The following are a number of research questions and hypotheses from actual published articles. For each hypothesis, identify the independent and dependent variable.
    a. “The use of color in a Yellow Pages advertisement will increase the perception of quality of the products for a particular business when compared with noncolor advertisements” (Lohse & Rosen, 2001, p. 75).
    b. “We hypothesized that parents who use more frequent corporal and verbal punishmen. . . . will report more problem behaviors in their children” (Brenner & Fox, 1998, p. 252).
    c. “It was hypothesized that adolescents with anorexia nervosa woul. . . . be more respectful when compared with peers with bulimia nervosa” (Pryor & Wiederman, 1998, p. 292).
    d. “The purpose of the present research was to assess brand name recognition as a function of humor in advertisements. . . . It was predicted that participants would recognize product brand names that had been presented with humorous

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

advertisements more often than brand names presented with nonhumorous advertisements" (Berg & Lippman, 2001, p. 197).

e. "The purpose of this study was to explore the relation between time spent in daycare and the quality of exploratory behaviors in 9-month-old infant. . . . it was hypothesized tha. . . . infants who spent greater amounts of time in center-based care would demonstrate more advanced exploratory behaviors than infants who did not spend as much time in center-based care" (Schuetze et al., 1999, p. 269).

f. "It was hypothesized that students would score higher on test items in which the narrative contains topics and elements that resonate with their daily experiences versus test items comprised of material that is unfamiliar" (Erdodi, 2012, p. 172).

3. Identify the independent and dependent variable for each of the following research questions and hypotheses from published articles.
    a. "It was hypothesized that the number of selfies posted . . . would be positively correlated with dimensions of narcissism" (Barry et al., 2017, p. 51).
    b. "We hypothesized that preschoolers' self-reported preferences for aggressive media . . . would be positively related to parent reports of children's aggressive behaviors" (Jamnik & DiLalla, 2018, p. 159).
    c. "We hypothesized that helicopter parenting would lead to higher levels of depression" (Darlow et al., 2017, p. 3).
    d. "We predicted that weight-related shame would be associated with greater binge eating" (Craven & Fekete, 2019, p. 45).
    e. "We hypothesize that men with strong gender identities will be more sensitive to manipulated threats to their masculinity compared with men with gender identities that are less central to their identity" (Harrison & Michelson, 2019, p. 65).
    f. "After exposure to fake news story with user comments identifying the news as fake, people are less likely to share the fake news than after exposure to the fake news story with user comments supportive of the content" (Colliander, 2019, p. 204).
4. Name the scale of measurement (nominal, ordinal, interval, ratio) for each of the following variables:
    a. The amount of time needed to react to a sound
    b. Gender
    c. Score on the Scholastic Aptitude Test (SAT)
    d. Political orientation (not at all conservative, conservative, very conservative)
    e. Political affiliation (Democrat, Republican, Independent)
5. Name the scale of measurement (nominal, ordinal, interval, ratio) for each of the following variables:
    a. One's age (in years)
    b. Size of soft drink (small, medium, large, extra large)
    c. Voting behavior (in favor vs. against)
    d. IQ score
    e. Parent (mother vs. father)
6. A faculty member wishes to assess the relationship between students' scores on the SAT and their performance in college.
    a. What is a possible research hypothesis in this situation?
    b. What are the independent and dependent variables?
    c. How could you measure the variable "performance in college" at each of the four levels of measurement?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

7. A researcher hypothesizes that drivers who use cellular phones will get into a greater number of traffic accidents than drivers who do not use these phones. For this example,
    a. What are the independent and dependent variables?
    b. How could you measure the dependent variable?
    c. How could you conduct this study using an experimental research method? How could you conduct this study using a nonexperimental method?

Answers to Learning Checks

Learning Check 1

1.

    a. IV: Instructor's dress; DV: Expertise

    b. IV: Amount of television viewing; DV: Posttraumatic stress symptoms

    c. IV: Student group; DV: Financial distress

    d. IV: Activity while driving; DV: Driving performance

Learning Check 2

2.

    a. Nominal

    b. Ratio

    c. Ordinal

    d. Ratio

Answers to Odd-Numbered Exercises

1.

    a. IV: Type of business major; DV: Narcissistic traits

    b. IV: Biological sex; DV: Employment of physical characteristics

    c. IV: Community; DV: Perceptions of emotional pain and support needs

    d. IV: Generation; DV: Beliefs about group projects

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

e. IV: Country; DV: Stress and anxiety

f. IV: Price presentation; DV: Perception of savings

3.

a. IV: Number of selfies posted; DV: Narcissism

b. IV: Preference for aggressive media; DV: Parent reports of children's aggressive behaviors

c. IV: Helicopter parenting; DV: Depression

d. IV: Weight-related shame; DV: Binge eating

e. IV: Strength of gender identity; DV: Sensitivity to threat to masculinity

f. IV: Type of user comments to news story; DV: Sharing of news story

5.

a. Ratio

b. Ordinal

c. Nominal

d. Interval

e. Nominal

7.

a. IV: Cellular phone use; DV: Number of traffic accidents

b. Example: Determine the number of traffic accidents each person has experienced in the past year.

c. Example of experimental: Use a driving simulation with the "driver" talking on a phone and measure the number of potential driving mistakes or accidents.

d. Example of nonexperimental: Take a survey of the number of accidents people have been in and whether or not they were using a phone during the accident.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CHAPTER 2 EXAMINING DATA: TABLES AND FIGURES

Learning Objectives

2.1 Describe the main purpose and the research hypothesis of the “lottery” study.

2.2 Describe different reasons for examining data.

2.3 Describe the steps used to create a frequency distribution table.

2.4 Describe the steps used to create a grouped frequency distribution table.

2.5 Explain the differences between bar charts, pie charts, histograms, and frequency polygons.

2.6 Describe the three aspects and types of distributions of data.

Chapter 1 described a process by which research may be conducted, which begins by stating a research hypothesis. The next step involves collecting data that will ultimately be used in statistical analyses designed to test the research hypothesis. However, researchers don’t conduct analyses until they’ve first examined the data to make sure appropriate conclusions can be drawn from the analyses. This chapter describes *why* researchers examine data and *how* data may be examined using tables, charts, and graphs. As with many of the chapters in this book, this discussion will be guided by the findings from a published research study.

2.1 An Example From the Research: Winning the Lottery

A friend of yours flips a coin and asks you to guess whether it’s “heads” or “tails.” You guess “tails,” but the answer is “heads.” Your friend flips the coin a second time, and again you predict “tails.” But again the answer is “heads.” On the third coin flip, you guess “tails” once more—but, once again, you’re wrong. As your friend prepares to flip the coin a fourth time, you tell yourself there’s little chance the coin could be “heads” again. But the reality is that the fourth coin flip is equally likely to be “heads” or “tails” because the outcome of the first three coin flips has absolutely no impact on the fourth flip. In other words, the outcome of a coin flip is a random event that can’t be controlled or predetermined. But do you really believe this?

The impact of random events can take on greater significance than the outcome of a coin flip. For example, government-run lotteries provide the opportunity to win huge amounts of money through the random selection of numbers. In 2001, Dr. Karen Hardoon, a researcher at McGill University in Canada, studied the thought processes people use when purchasing lottery tickets (Hardoon et al., 2001). As previous research had found gamblers believe they can control random events such as the rolling of dice or the spinning of a roulette wheel, Dr. Hardoon and her colleagues were interested in

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

determining “whether individuals with gambling problems perceive the purchase of lottery tickets in a similar manner as non-problem gamblers” (p. 752). On the basis of earlier findings, Dr. Hardoon hypothesized that problem gamblers are less likely to believe lottery outcomes are random than are nonproblem gamblers.

To test their research hypothesis, Dr. Hardoon and her associates located a group of gamblers and asked whether they had ever experienced problems as a result of their gambling. Each participant was classified into one of two groups: problem gamblers and nonproblem gamblers.

The participants in the study, which we’ll refer to as the lottery study, were shown four lottery tickets and asked, “If you were to buy a ticket to play in the lottery, which one would you select?” Each ticket contained six numbers selected from among the numbers 1 through 49; the four tickets were given a particular label. The *sequence* ticket had numbers in consecutive order (30-31-32-33-34-35); numbers on the *pattern* ticket increased by 5s (5-10-15-20-25-30); the *nonequilibrated* numbers (35-37-40-43-44-49) were clustered at the upper end of the 49 numbers, and the *random* ticket (7-8-23-34-36-42) shared none of the characteristics of the other three tickets. We’ll call the variable in this study “lottery ticket,” a variable with four categories: sequence, pattern, nonequilibrated, and random.

The researchers recorded which of the four tickets each participant chose. Table 2.1 lists the data for the lottery ticket variable for the nonproblem gamblers (the data for the problem gamblers are provided in the exercises at the end of this chapter). Looking at this table, we see that the nonproblem gamblers differed in their lottery ticket choices; however, it’s difficult to describe these differences given the unorganized nature of the collected data. Consequently, before conducting statistical analyses on their data, researchers typically organize and examine them. In the following sections, we will discuss why and how data are organized and examined.

2.2 Why Examine Data?

There are a variety of reasons why researchers examine data they’ve collected before analyzing the data. These reasons include the following:

To gain an initial sense of the data

To identify outliers

To evaluate research methodology

To determine whether data meet statistical criteria and assumptions

Each of these reasons is discussed in the following sections.

Table 2.1 The Lottery Ticket Choices of 22 Nonproblem Gamblers

Participant	Lottery Ticket
1	**Random**
2	**Random**
3	**Nonequilibrated**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Participant	Lottery Ticket
4	**Pattern**
5	**Random**
6	**Sequence**
7	**Random**
8	**Random**
9	**Pattern**
10	**Pattern**
11	**Random**
12	**Pattern**
13	**Random**
14	**Nonequilibrated**
15	**Random**
16	**Pattern**
17	**Nonequilibrated**
18	**Random**
19	**Random**
20	**Pattern**
21	**Random**
22	**Random**

Gaining an Initial Sense of the Data

Because researchers spend a great amount of time and energy designing their research studies, they're eager to analyze the data they've collected. Students taking courses in statistics, on the other hand, face a different challenge: They're given sets of data with which they have no prior experience. Although both researchers and students may be motivated to analyze their data as quickly as possible, it's important to first examine the data in order to gain an initial sense of them.

As an example, imagine you conduct a study regarding the fuel efficiency of automobiles and collect data on the miles per gallon (MPG) of different types of cars. You might expect the scores in your data to range from 10 to 50 MPG, with the majority between 25 and 35 MPG. Examining data helps researchers gain an initial sense of whether the data are in line with their expectations. Examining data also provides a way of confirming the results of later statistical analyses. We've found that students' errors in their calculations could have been avoided had they first looked at the set of data. For example, examining data will help you avoid making statements such as, "The average MPG in my sample was 127 MPG."

Identifying Outliers

Researchers also examine their data to identify **outliers** , defined as rare, extreme scores that lie outside of the range of the majority of scores in a set of data. Imagine, for example, that a group of 15 students is asked to take a 20-item test of mathematical calculations; each student's score is the number of test items answered correctly. In scoring the tests, a researcher finds that 14 students

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

correctly answered between 8 and 17 questions. However, one student provided the wrong answer to every question for a score of 0. In this case, the score of 0 would be considered an outlier.

If not identified, outliers may distort conclusions researchers draw about their sample and their data, particularly when the sample is relatively small. In the previous example, the single score of 0 might lead an observer to conclude the class as a whole performed more poorly on the test than they actually did.

Evaluating Research Methodology

Examining data is also useful in assessing the effectiveness of the methods used to collect the data. For the 20-item math test, for example, it might be reasonable to expect the majority of scores to fall in the middle of the 0 to 20 range. But what if every student answered at least 17 of the 20 questions correctly? You might conclude the test was too easy and that more difficult questions are needed to accurately assess students' math skills. By comparing the collected data for a variable with the range of possible values, researchers are able to evaluate and modify their measurement tools.

Determining Whether Data Meet Statistical Criteria and Assumptions

Researchers also examine their data to see whether they meet certain statistical criteria and assumptions. Many of the statistics in this book are based on assumptions regarding the shape and nature of a set of data that has been collected. For example, these procedures assume that in most data sets, the majority of scores will fall in the middle of the range of possible values, with a small amount of scores at the highest and lowest possible values (you may have heard this referred to as "the bell curve"). In the example of the 20-item math test, it might be assumed that most scores will be between 7 and 13, with smaller number of scores less than 7 or greater than 13. When assumptions such as this are not met, it's possible to draw inappropriate conclusions from statistical analyses that are conducted.

2.3 Examining Data Using Tables

One part of examining data that have been collected for a variable involves answering a series of questions:

> What are the values of the variable?
>
> How many participants in the sample have each value of the variable?
>
> What percentage of the sample has each value of the variable?

As it may be difficult to answer questions such as these just by looking at the data in Table 2.1 , researchers typically organize their data by creating a table known as a **frequency distribution table** , a table that summarizes the number and percentage of participants for the different values of a variable. This section discusses the steps involved in creating frequency distribution tables using these questions.

What Are the Values of the Variable?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The first step in organizing the data for a variable is to identify all of the possible values for the variable. Table 2.2 begins the construction of a frequency distribution table by identifying the four values for the lottery ticket variable: sequence, pattern, nonequilibrated, and random.

How Many Participants in the Sample Have Each Value of the Variable?

Once the values of the variable have been identified, the next step in organizing a set of data is to determine the **frequency** of each value, which is the number of participants in a sample corresponding to a value of a variable. These frequencies help indicate the nature and shape of the distribution of scores for a variable in a sample.

Table 2.2 Identifying the Values of the Lottery Ticket Variable

Number	Lottery Ticket
1	Sequence
2	Pattern
3	Nonequilibrated
4	Random

Table 2.3 Using the Tick Mark Method to Determine the Frequencies of the Lottery Ticket Variable

Lottery Ticket	Tick Marks	*f*
Sequence	|	1
Pattern	||||— |	6
Nonequilibrated	|||	3
Random	||||— ||||— ||	12
Total	||||— ||||— ||||—||||—||	22

One way to determine the frequencies for a variable is to sort the data into the different values and then count the number of participants having each value using a “tick mark method,” where a tick mark is placed next to a value of the variable each time that value appears in the sample. As shown in Table 2.3 , the tick marks are then counted to determine the frequency for each value. For example, the number of tick marks and therefore the frequency (f) for the sequence ticket is 1 because it was picked by one participant (Participant 6 in Table 2.1). Similarly, the tick marks for each of the other three types of tickets can be recorded and counted until all of the data in the sample are accounted for.

What Percentage of the Sample Has Each Value of the Variable?

Although it’s important to determine the frequency for each value of a variable, these frequencies by themselves are not always particularly informative. For example, how you interpret “There are 35 Democrats in my sample” is different when the sample consists of 40 voters rather than 400 voters.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Because the interpretation of frequencies depends on the size of the sample, it's useful to calculate the *percentage* of the sample having each value of a variable. These percentages can be calculated using the following formula:

$$\% = \frac{f}{total\,number\,of\,scores} * 100$$

where *f* is the frequency for a value of a variable.

Looking at Table 2.3 , 1 of the 22 nonproblem gamblers chose the sequence ticket. Therefore, the percentage of the sample choosing this ticket is

$$\begin{aligned} \% &= \frac{f}{total\,number\,of\,scores} * 100 \\ &= \frac{1}{22} * 100 = .05 * 100 \\ &= 5\% \end{aligned}$$

After calculating the percentage for each of the four tickets, Table 2.4 provides the frequency distribution table for the lottery ticket variable.

Once a frequency distribution table has been constructed for a variable, researchers can begin to draw conclusions about their sample. For the lottery ticket variable, the random ticket had the highest frequency and percentage (*f* = 12, 54%) compared to the pattern (*f* = 6, 27%), nonequilibrated (*f* = 3, 14%), and sequence (*f* = 1, 5%) tickets. As a result, the researchers in the lottery study made the following observation:

"The results of the present study indicated that, for the entire sample . . . the greatest percentage of tickets selected by the entire sampl. . . . were random tickets." (Hardoon et al., 2001, p. 760)

Table 2.4 Frequency Distribution Table for the Lottery Ticket Variable

Lottery Ticket	***f***	**%**
Sequence	1	**5%**
Pattern	6	**27%**
Nonequilibrated	3	**14%**
Random	12	**54%**
Total	22	**100%**

Learning Check 1

Reviewing What You've Learned So Far

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



1. Review questions
    a. What are some reasons for examining data before conducting statistical analyses?
    b. Why are outliers of concern to researchers?
    c. What questions do you address about a set of data for a variable in constructing frequency distribution tables?
    d. Why is it useful to calculate percentages in addition to frequencies for each value of a variable?
2. For each of the following situations, create a frequency distribution table.
    a. A college counselor asks a group of 116 seniors what their plans are after graduating from college; 71 say they are going to “work,” 34 are planning on going to “graduate school,” and 11 say they are “not sure” what their plans are.
    b. A high school instructor teaching a course finds that of the students she teaches, 69 are Freshmen, 18 are Sophomores, 12 are Juniors, and 9 are Seniors.
    c. A pollster stops 12 people and asks if they are either “in favor” or “against” a local proposition in Table 2.5 :

Table 2.5 Example of People’s Position

Person	Position
1	**In favor**
2	**In favor**
3	**Against**
4	**In favor**
5	**Against**
6	**Against**
7	**In favor**
8	**Against**
9	**In favor**
10	**Against**
11	**In favor**
12	**In favor**

    d. The company that makes M&Ms candy (www.mms.com) conducted a survey asking people which new color they would like to have added: purple, aqua, or pink. Table 2.6 are the votes of a hypothetical sample:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Table 2.6
Hypothetical Sample of New M&M Color

Person	Color
1	**Pink**
2	**Purple**
3	**Pink**
4	**Aqua**
5	**Purple**
6	**Purple**
7	**Purple**
8	**Aqua**
9	**Purple**
10	**Pink**
11	**Aqua**
12	**Pink**
13	**Pink**
14	**Purple**
15	**Aqua**
16	**Pink**
17	**Purple**
18	**Purple**
19	**Aqua**
20	**Purple**
21	**Purple**
22	**Pink**
23	**Pink**
24	**Purple**

2.4 Grouped Frequency Distribution Tables

The frequency distribution table in Table 2.4 was sufficient for presenting the frequencies for the lottery ticket variable. However, other situations may involve variables with a large number of possible values. Consider, for example, the variable grade point average (GPA). The values for GPA typically range from 0.00 to 4.00, with hundreds of possible values in between. Creating a frequency distribution table for GPA could result in an extremely large table, with many values having frequencies of zero.

To examine a variable with a large number of values, it is useful to construct a **grouped frequency distribution table** , a table that groups the values of a variable measured at the interval or ratio level

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

of measurement into a small number of intervals and then provides the frequency and percentage for each interval. To illustrate how grouped frequency distribution tables are constructed, Table 2.7 (a) provides a sample of 42 hypothetical GPAs. In Table 2.7 (b), each of these 42 GPAs has been placed into one of eight intervals that comprise a grouped frequency distribution table.

To illustrate how to read and interpret a grouped frequency distribution table, the lowest interval in Table 2.7 (b) consists of students with GPAs less than (<) 2.00. The frequency in this interval (*f* = 1) consists of one student: Student 29 (GPA = 1.96). The next interval consists of students with GPAs ranging from 2.00 to 2.29 and contains three students (4, 22, and 38). This grouping continues until all of the GPAs in the sample have been accounted for.

To create a grouped frequency distribution table, the real limits of each interval should be identified. The **real limits** of an interval are the values of the variable that are halfway between the top of one interval and the bottom of the next interval. For example, because of how values of GPA are rounded, the lowest value of GPA that would lead someone to be placed in the interval 2.30–2.59 is 2.295, which is halfway between the top of the 2.00–2.29 interval and the bottom of the 2.30–2.59 interval. The **real lower limit** is the smallest value of a variable that would be grouped into a particular interval. In this case, 2.295 represents the real lower limit for the 2.30–2.59 interval. On the other hand, the **real upper limit** is the largest value of a variable that would be grouped into a particular interval. For the 2.30–2.59 interval, the real upper limit is 2.594.

Once the grouped frequency table for the GPA example has been created (Table 2.7(b)), what preliminary conclusions might be drawn about these students? First, because the 2.90–3.19 interval has the greatest frequency (*f* = 10), you might conclude that the typical student in this sample has a GPA close to 3.00. Second, moving in both directions away from the 2.90–3.19 interval, the frequencies of the different intervals grow progressively smaller, with fewer and fewer students having either very high or very low GPAs. Table 2.8 provides a list of guidelines for creating grouped frequency distribution tables.

Cumulative Percentages

In addition to determining the percentage of the sample that has each individual value or interval for a variable, it is sometimes useful to combine these percentages. For the GPA example, you may be interested in knowing, “What percentage of the sample had GPAs less than 2.60?” To answer this question, the percentages in the “%” column of the frequency distribution table must be combined. This type of percentage is referred to as a **cumulative percentage** , defined as the percentage of a sample at or below a particular value of a variable.

Table 2.7a Grade Point Average (GPA), 42 Students: Individual GPAs

Student	GPA
1	**3.05**
2	**2.83**
3	**3.26**
4	**2.19**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Student	GPA
5	**3.52**
6	**3.34**
7	**3.02**
8	**2.97**
9	**3.71**
10	**2.50**
11	**2.77**
12	**3.10**
13	**3.06**
14	**3.83**
15	**2.39**
16	**3.16**
17	**3.92**
18	**2.37**
19	**3.65**
20	**3.70**
21	**3.00**
22	**2.26**
23	**2.65**
24	**3.40**
25	**2.87**
26	**2.92**
27	**2.89**
28	**3.41**
29	**1.96**
30	**3.77**
31	**3.20**
32	**2.86**
33	**2.94**
34	**2.71**
35	**3.43**
36	**3.95**
37	**3.32**
38	**2.28**
39	**3.08**
40	**3.25**
41	**2.40**
42	**3.22**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Table 2.7b Grade Point Average (GPA), 42 Students: Grouped Frequency Distribution Table

GPA	*f*	%
3.80+	3	7%
3.50−3.79	5	12%
3.20−3.49	9	21%
2.90−3.19	10	24%
2.60−2.89	7	17%
2.30−2.59	4	10%
2.00−2.29	3	7%
<2.00	1	2%
Total	42	100%

Table 2.8 Guidelines for Creating Grouped Frequency Distribution Tables

Guideline	Guideline Description
1. *Variables are grouped into approximately 10 intervals.*	The exact number of intervals depends on the data in each sample. For example, data that fall within a small range of values may be accurately represented with a relatively small number of intervals, whereas data extending across a wide range may require a larger number of intervals.
2. *The number of intervals should accurately represent the data.*	The intervals should represent the nature of the data as accurately as possible. For example, there should not be so large a number of intervals that many intervals have frequencies of zero (e.g., 3.00−3.05) or so few intervals that a large majority of the sample falls into only one or two intervals (e.g., 3.00−3.99).
3. *Intervals should be of equal size.*	To make the intervals comparable, they should be of equal size or width. For example, you would not want one interval of 2.01−2.50 (a width of .50) and another of 2.51−2.70 (a width of .20). One exception to this guideline is the lowest or highest interval (e.g., less than 2.00 or greater than 3.70), which is typically wider than the other intervals because it contains a relatively small frequency.
4. *Intervals should not overlap.*	The intervals should be distinct from each other, such that each score may be placed in only one interval. For example, intervals of 2.30− *2.60* and *2.60*− 2.90 are a problem because the GPA 2.60 is in both intervals leading to confusion regarding within which interval to place someone with a GPA of 2.60.

Table 2.9 includes the cumulative percentages for the GPA example. As an example of a cumulative percentage, the cumulative percentage for the 2.30–2.59 interval (19%) has been calculated by combining the percentages for three intervals: < 2.00, 2.00–2.29, and 2.30–2.59: 2% + 7% + 10% = 19%. As a result, we could conclude that 19% of this sample had GPAs less than 2.60. The accumulating of percentages continues until all (100%) of the scores in the sample are combined. As

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

shown in Table 2.9 , cumulative percentages are typically located in the final column of a frequency distribution table and labeled “Cum %.”

Table 2.9 Frequency Distribution Table With Cumulative Percentages for the GPA Variable

GPA	*f*	%	Cum %
3.80+	3	7%	**100%**
3.50−3.79	5	12%	**93%**
3.20−3.49	9	21%	**81%**
2.90−3.19	10	24%	**60%**
2.60−2.89	7	17%	**36%**
2.30−2.59	4	10%	**19%**
2.00−2.29	3	7%	**9%**
< 2.00	1	2%	**2%**
Total	42	100%	

2.5 Examining Data Using Figures

In addition to organizing data into frequency distribution tables, researchers also examine data visually using figures such as charts or graphs. There are four types of figures often used to visually portray data for variables: bar charts, pie charts, histograms, and frequency polygons. Which type of figure a researcher may use depends on the variable’s level of measurement (nominal, ordinal, interval, ratio). Data for variables measured at the nominal or ordinal level of measurement are typically displayed using bar charts or pie charts. The values of interval and ratio variables are graphed using histograms and frequency polygons. Each of these types of visual illustration is discussed in the following sections.

Displaying Nominal and Ordinal Variables: Bar Charts and Pie Charts

The values of variables measured at the *nominal* level of measurement differ in category or type; one example of a nominal variable is the lottery ticket variable discussed earlier, which consisted of four categories: sequence, pattern, nonequilibrated, and random. The values of variables measured at the *ordinal* level of measurement can be placed in an order relative to the other values; age (young, middle age, elderly) is one example of an ordinal variable.

One way to display nominal and ordinal variables is to use a bar chart. A **bar chart** is a figure that uses bars to represent the frequency or percentage of a sample corresponding to each value of a variable, with the different bars not touching each other. Figure 2.1 shows a bar chart for the data from the lottery ticket variable. In this figure, each bar corresponds to one of the four types of lottery tickets; the height of each bar corresponds to the frequency (*f*) of nonproblem gamblers who selected that type of ticket.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

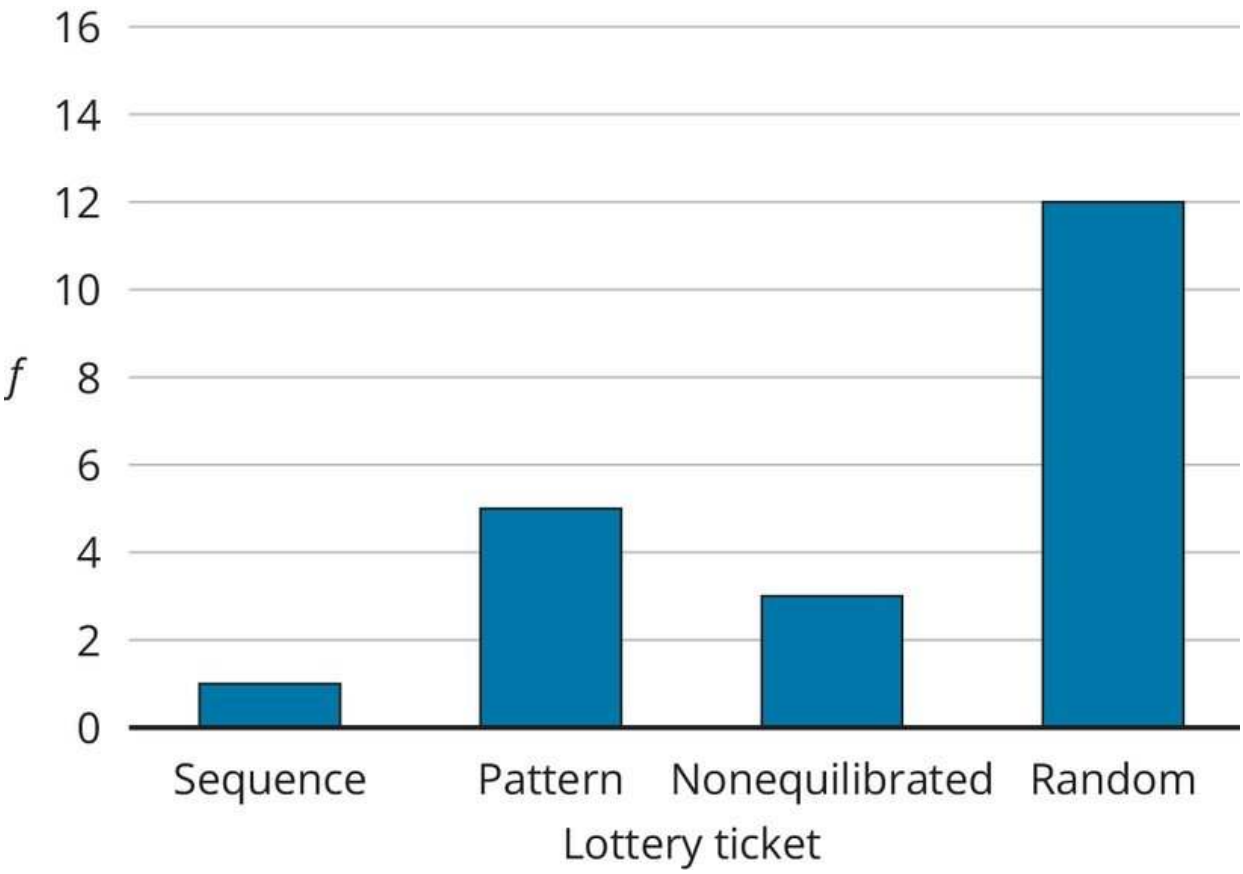


Figure 2.1 Bar Chart Showing the Frequencies of the Lottery Ticket Variable for the Nonproblem Gamblers

When creating a bar chart, the horizontal (X) axis contains the name of the variable and the values of the variable; the vertical (Y) axis contains the frequency (*f*) or percentage (%) for each value. Notice that the bars in the bar chart in Figure 2.1 do not touch each other; the gap between the bars indicates that the values of the variable represent distinct groups or categories that can't be placed along a numeric continuum.

A second type of figure used to display nominal or ordinal variables is a pie chart. A **pie chart** is a figure that uses a circle divided into proportions (slices) that represent the percentage of the sample for each value of a variable. Pie charts are created by dividing the 360 degrees of a circle into proportions corresponding to the percentage of each value.

Figure 2.2 shows a pie chart for the data for the lottery ticket variable. To illustrate how pie charts are constructed, the frequency distribution table in Table 2.4 shows that 5% of nonproblem gamblers chose the sequence ticket. Given that a pie chart consists of 360 degrees, 5% of the 360 degrees are needed to represent the sequence ticket. As 5% * 360 = 18, 18 degrees of the pie is assigned to the sequence ticket. Similarly, 97 degrees of the pie chart (27% of 360) are used to represent the pattern ticket. This process continues until all 360 degrees of the pie chart are accounted for.

The decision whether to use a bar chart or pie chart to illustrate a variable may depend on the number of values for the variable. For example, a pie chart for a variable that has 12 values may

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

result in many of the 12 “slices” being very thin and small, making them hard to see, meaning that a bar chart may be a better choice to illustrate this variable.

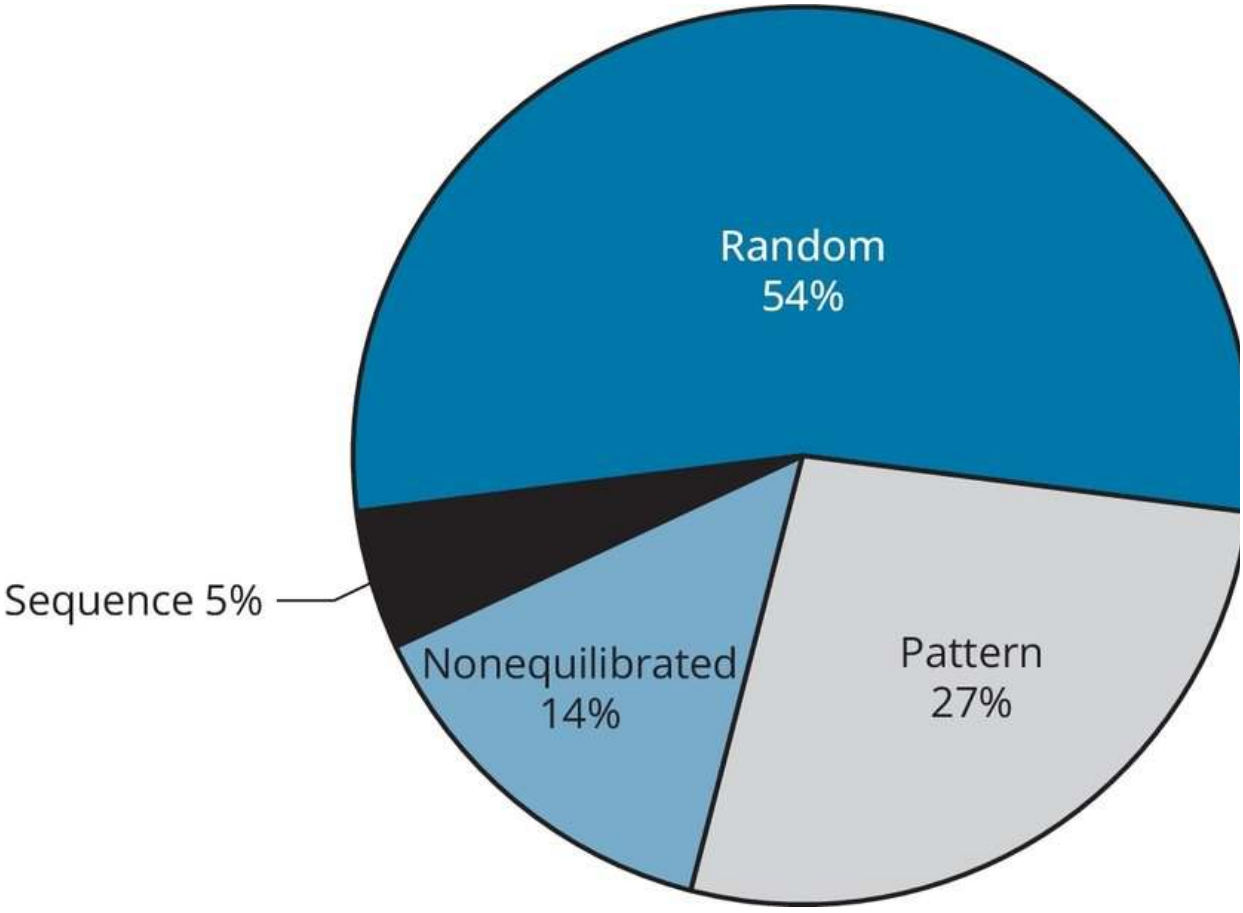


Figure 2.2 Pie Chart Showing the Percentages of the Lottery Ticket Variable for the Non−Problem Gamblers

Displaying Interval and Ratio Variables: Histograms and Frequency Polygons

Interval and ratio variables have values equally spaced along a numeric continuum and are identical to each other with one exception: Ratio scales possess a true zero point, for which the value of zero (0) represents the complete absence of the variable. The values of interval and ratio variables may be illustrated using histograms and frequency polygons, the choice of which depends on the number of values for the variable.

A **histogram** is a figure in which bars are used to represent the frequency of each value of a variable, with the different bars touching each other. The name and values of the variable are located on the X-axis, and the frequency or percentage for each value is on the Y-axis. Histograms are similar to bar charts, with one important difference: The bars in a histogram touch each other, indicating that the values of the variable are connected to each other along a numeric continuum.

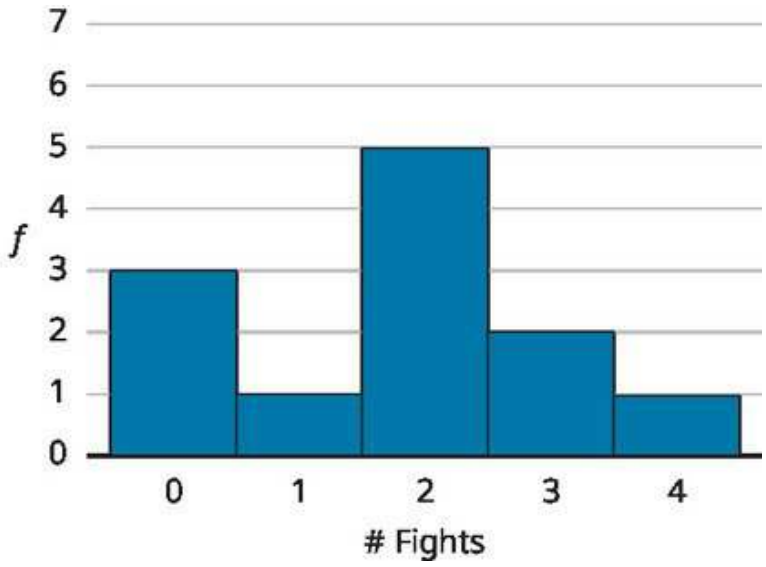
join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

As an example, imagine a teacher is interested in studying aggressive behavior in children and measures this variable by counting the number of fights each of her students is involved in over a 3-day period. “Number of fights” is a ratio variable rather than an interval variable because the value of 0 represents the complete absence of fights. Figure 2.3 provides a frequency distribution table and histogram for this variable.

A histogram is used when an interval or ratio variable consists of a relatively small number of values. But what if, for example, the number of fights could be anywhere from 0 to 20 rather than 0 to 5? In order to have the histogram fit on one page, the 21 bars would need to be very thin and crowded together, making the histogram confusing and difficult to read. In situations such as this, the values of the variable may be more clearly illustrated using a figure known as a **frequency polygon**, which is a line graph that uses data points to represent the frequency of each value of a variable, with straight lines connecting the data points. In a frequency polygon, the X-axis represents the name and values of the variable, while the Y-axis represents the frequency for each value.

# Fights	*f*
4	1
3	2
2	5
1	1
0	3
Total	12



▶ Long Description: Figure 2.3 Frequency Distribution Table and Histogram for the Number of Fights Variable

Figure 2.3 Frequency Distribution Table and Histogram for the Number of Fights Variable

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

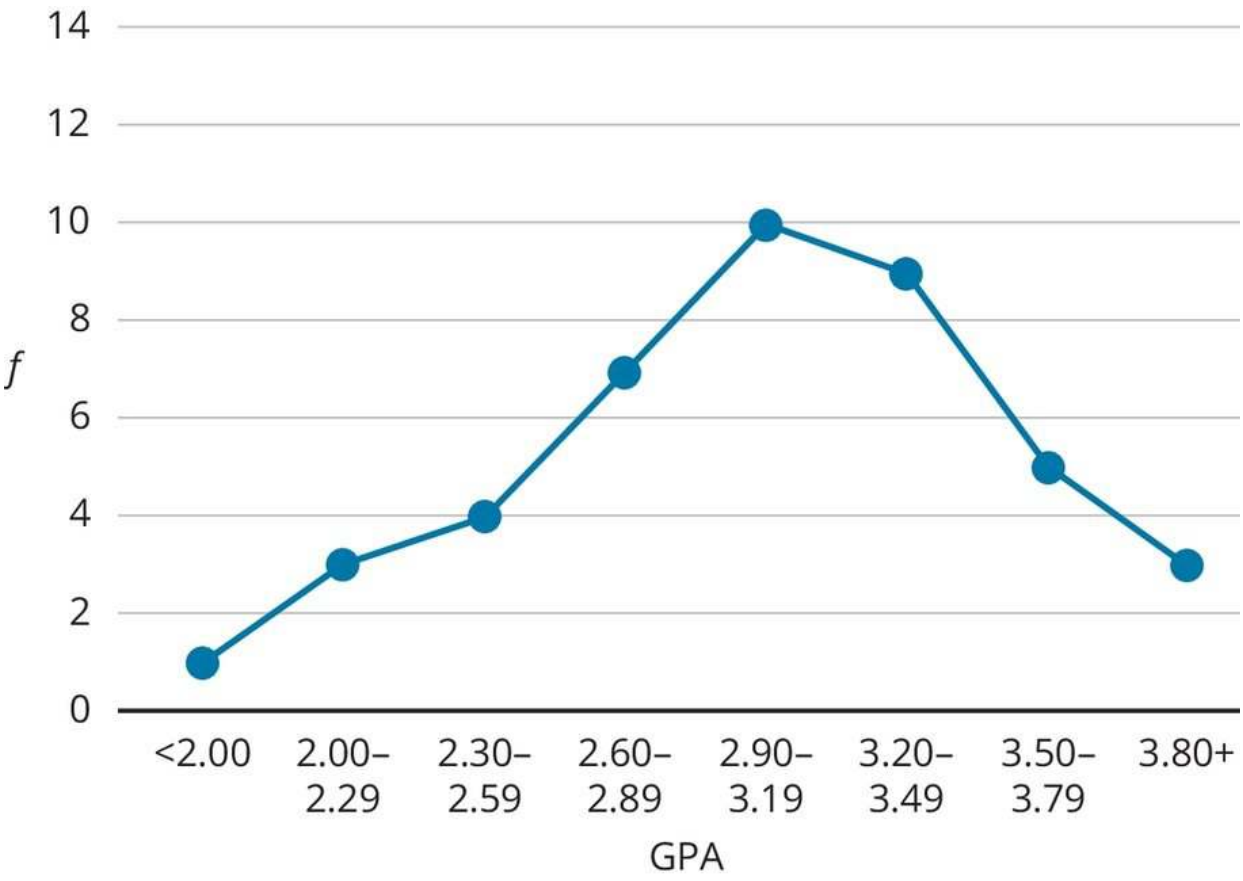


Figure 2.4 Frequency Polygon Showing the Frequency of Students in Each Interval of the Grade Point Average (GPA) Variable

Figure 2.4 shows a frequency polygon based on the grouped frequency distribution table of the earlier example of students' GPA (Table 2.7(b)). In this frequency polygon, the data points representing the frequencies of the eight GPA intervals are connected using straight, sharp lines. These lines indicate that the values of the variable can be connected together along a numeric continuum.

Although a frequency polygon like the one in Figure 2.4 may look jagged as it moves from one data point to the next, it provides a general sense of the shape of the distribution of scores for a variable. For example, looking at Figure 2.4 , you might conclude that the majority of GPAs in this sample are in the middle of the 1.00 to 4.00 range, with relatively low frequencies of students with either low or high GPAs.

Learning Check 2

Reviewing What You've Learned So Far

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



1. Review questions
    a. What is the difference between a frequency distribution table and a grouped frequency distribution table?
    b. For what levels of measurement would you create a pie chart, bar chart, histogram, or a frequency polygon?
    c. What is the difference between a bar chart and a histogram?
    d. Why would it be inappropriate to create a histogram or frequency polygon for a nominal or ordinal variable?
    e. Under what circumstances might you use a frequency polygon rather than a histogram to graph an interval or ratio variable?
2. For each of the following variables, identify whether you could use a bar chart, pie chart, histogram, and/or a frequency polygon to visually display the data.
    a. Miles per gallon (MPG)
    b. Living situation (on-campus, off-campus)
    c. Number of brothers and sisters
    d. Marital status (single, married, divorced)

Drawing Inappropriate Conclusions From Figures

A well-constructed figure can greatly facilitate the understanding of complex information. However, a poorly designed figure may not only be confusing, it can lead to inappropriate or misleading conclusions. Box 2.1 provides guidelines for constructing figures based on recommendations from the American Psychological Association (APA).

A simple example may demonstrate how altering the features of a figure can lead to different interpretations of the same set of data. Imagine a pollster assesses how voters with different political affiliations plan to vote on a ballot measure. She asks a sample of Democrats, Republicans, and Independents the question, “Do you plan on voting in favor of Measure A?” In examining her data, the pollster finds 56% of Democrats, 47% of Republicans, and 52% of Independents intend to vote “yes” on the measure.

Figure 2.5 graphs these three percentages using two bar charts; bar charts are used to display the data because the variable of interest (political affiliation) is a nominal variable (Democrat, Republican, or Independent). The bars don’t touch each other because the three political affiliations are distinct from each other and can’t be placed along a numeric continuum.

Although the two bar charts are similar, there is one important difference in how they are constructed. Following the guidelines in Box 2.1, the bar chart in Figure 2.5 (a) starts the Y-axis with the value 0% and ends with 100%. However, the bar chart in Figure 2.5 (b) starts with the value 45% and ends with 65%. By limiting the range of values of the percentages, the bar chart in Figure 2.5 (b) exaggerates differences in the heights of the three bars. As a result, someone looking at Figure 2.5 (b) may interpret the differences between the three political affiliations as being larger than someone looking at Figure 2.5 (a).

The saying “a picture is worth a thousand words” is particularly relevant for the presentation of statistical data. Because figures can provide a vivid representation of data, it’s the responsibility of researchers to portray their data in an accurate manner and to avoid creating images that may mislead or confuse.

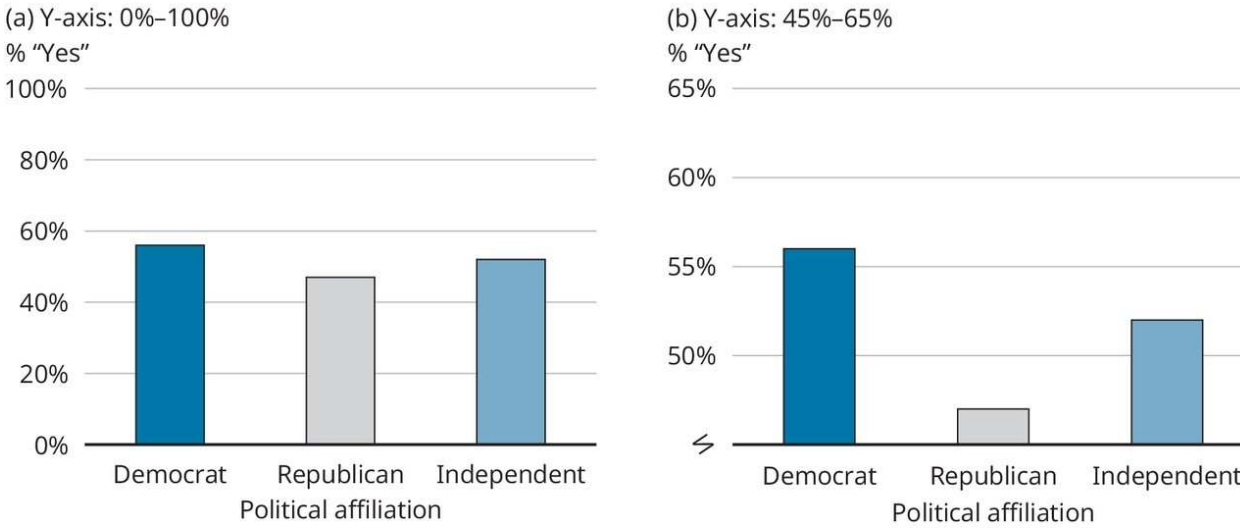
join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Box 2.1 Some APA Guidelines for Creating Figures

1. The figure should be formatted in a portrait rather than landscape page orientation (the reader should not turn the page sideways to read the table).
2. The figure is numbered using Arabic numbers (Figure 1, not Figure *I* or Figure *A*).
3. The figure includes a title that describes the contents of the figure. For example, it is not enough to name a figure simply " Figure 2.1 " or " Figure 2.1 . Bar chart."
4. Figures have a rectangular shape, with the height of the vertical (Y) axis two thirds to three fourths the length of the horizontal (X) axis.
5. The horizontal (X) axis contains the values of the variable; the vertical (Y) axis contains the frequency (*f*) or percentage (%) of the sample having each value of the variable.
6. The *X* - and *Y* -axes include descriptive labels written in both upper- and lowercase letters (as opposed to all capital letters).
7. The labels for the *X* - and *Y* -axes are short enough to fit on one line along the axes.
8. The label of the Y-axis is often placed parallel to the Y-axis (the letters face sideways rather than vertically) but can sometimes be seen on top of the y-axis for accessibility.
9. The Y-axis starts with the value zero (0) and is divided into equally spaced values or intervals.
10. The Y-axis is divided into 6 to 10 values (just enough values to accommodate bar length). The bars should neither be flattened along the bottom of the Y-axis nor bumped up against the top of the Y-axis.

Political affiliation	% "Yes"
Democrat	56%
Republican	47%
Independent	52%



▶ Long Description: Figure 2.5 Different Displays of the Same Data

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Figure 2.5 Different Displays of the Same Data

2.6 Examining Data: Describing Distributions

We've discussed how researchers construct tables and figures to organize their data. Looking at a table or figure enables researchers to describe the distribution of values or scores for a variable. Describing a distribution involves addressing questions such as, "Which values of the variable occur most often?," "How do the frequencies of the values of the variable change in relation to the most common values?," and "What is the amount of differences among the different values?"

Questions such as these pertain to three aspects of distributions: modality, symmetry, and variability. These three aspects of distributions are introduced in the following sections; they'll also be discussed in greater detail in the next two chapters of this book.

Modality

The first feature of a distribution of data for a variable is its modality. The **modality** of a distribution refers to the value or values of the variable that have the highest frequency or occur most often in a set of data. Using the lottery study as an example, identifying the modality of the lottery ticket variable involves answering the question, "Which of the four lottery tickets was chosen the most often?"

The modality of a distribution can be identified by looking at a frequency distribution table; the modality is the value(s) of the variable with the highest frequency (*f*) or percentage (%). The modality may also be determined by examining a figure and locating the tallest bar in a bar chart or histogram, the highest point in a frequency polygon, or the biggest "slice" in a pie chart. For the lottery ticket variable, the modality is the random ticket.

In identifying the modality of a distribution, it may be that one value or a limited number of values adjacent to one another clearly have the highest frequency. In such cases, the distribution is referred to as a unimodal distribution. A **unimodal distribution** is a distribution where one value of a variable occurs with the greatest frequency. Looking at the frequency distribution table (Table 2.4) or figure (Figure 2.1) for the lottery ticket variable, this distribution would be considered unimodal as one value, the random ticket, clearly occurred more often than the other three ticket choices.

In contrast to a unimodal distribution, it's possible that two distinct values may both have relatively high frequencies; if so, the distribution is referred to as a **bimodal distribution** , which is a distribution where two values clearly distinct from each other have the greatest frequency. A bimodal distribution sometimes resembles the two humps on a camel's back. As an example of a bimodal distribution, in April 2024 the Pew Research Center (www.pewresearch.org) asked a sample of 7,000 voters whether they felt Donald Trump had the mental fitness needed to be president, with five responses ranging from "extremely" to "not at all." The majority of voters chose one of the two extreme choices, either "extremely" (23%) or "not at all" (35%), with only 14% choosing the middle response of "somewhat."

And, in addition to unimodal and bimodal distributions, it's possible for a distribution to have more than two values with relatively equally high frequencies; this type of distribution is known as a **multimodal distribution** —a distribution where more than two values have the greatest frequency. In examining people's attitudes toward gun control, in May 2023, the news network CNN asked 1,227 people the question, "Do you think allowing gun owners to carry their guns in public places makes those places safer, less safe, or makes no difference in how safe a public place is?" They found people differed greatly in their responses, with 32% saying "makes public places safer," 36% saying

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

"makes public places less safe," and 32% saying "makes no difference," Figure 2.6 presents an example of a unimodal, bimodal, and multimodal distribution.

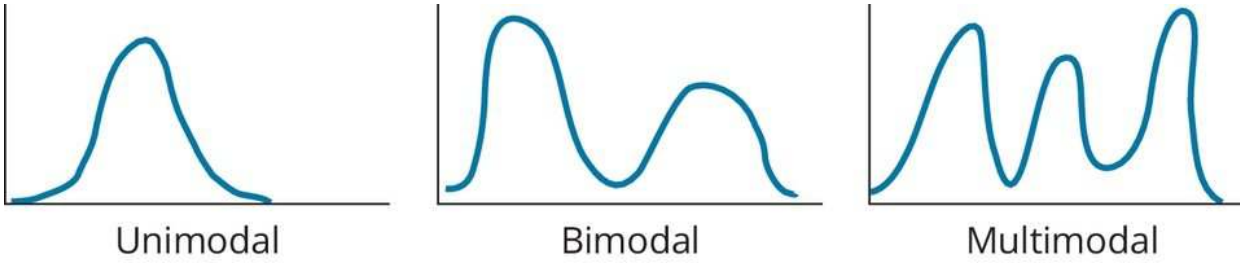


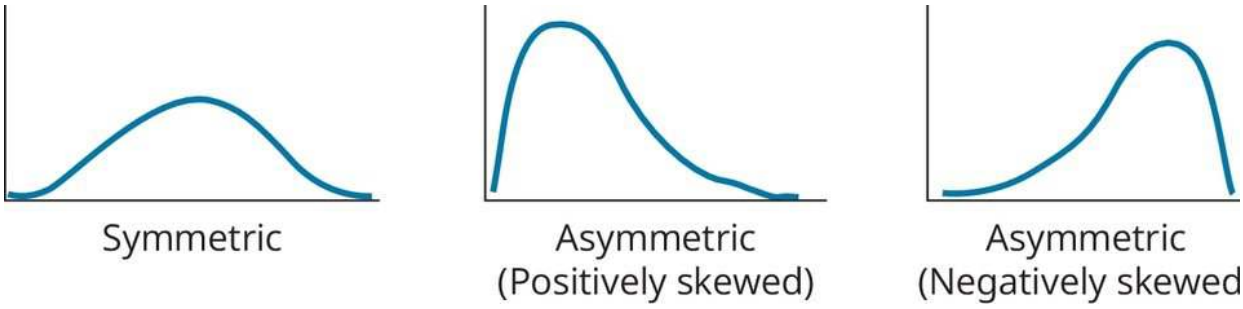
Figure 2.6 Distributions With Different Modalities (Unimodal, Bimodal, and Multimodal)

Symmetry

The second feature of a distribution of data is its symmetry. The **symmetry** of a distribution refers to how the frequencies of values of a variable change in relation to the most common or frequently occurring values. Describing the symmetry of a distribution means answering the question, "What is the shape of the distribution?"

The symmetry of a distribution can be determined by looking at the frequencies in a frequency distribution table or the bars or points in a figure. By doing so, a distribution may be described as being either symmetric or asymmetric. A **symmetric distribution** is one in which the frequencies change in a similar manner moving away in both directions from the most frequently occurring values. An example of a symmetric distribution is illustrated in Figure 2.7 . One way to determine whether a distribution is symmetric is to draw the distribution on a piece of paper and then fold the distribution vertically down the middle so that the two halves are on top of each other. A distribution is symmetric when the two halves are similar in shape such that they resemble mirror images of each other.

However, a distribution may also be asymmetric. An **asymmetric distribution** (also known as a *skewed* distribution) is one in which the frequencies change in a different manner moving away in both directions from the most frequently occurring values. In asymmetric distributions, the highest frequencies are located at one end of the distribution rather than in the middle, such that if you fold an asymmetric distribution down the middle, the shapes of the two halves do not look the same.



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Figure 2.7 Distributions With Different Symmetries (Symmetric, Positively Skewed, and Negatively Skewed)

In an asymmetric distribution, the frequencies take on the shape of a long tail as the values move toward the end of the distribution; scores in the tail of the distribution are often considered to be outliers. For example, in the 2022 Current Population Survey, the U.S. Census Bureau (www.census.gov) estimated that most Americans (63%) had annual incomes less than $50,000. However, a very small percentage (3%) were greater than $200,000; these higher incomes would be considered outliers.

Depending on the location of its tail, asymmetric distributions are referred to either as positively skewed or negatively skewed. A **positively skewed distribution** is one in which the higher frequencies are at the lower end of the distribution, with the tail on the upper (right) end of the distribution; a **negatively skewed distribution** is one in which the higher frequencies are at the upper end of the distribution, with the tail on the lower (left) end of the distribution. Figure 2.7 provides an illustration of a symmetric distribution and two asymmetric distributions (one positively skewed and one negatively skewed). Annual income, illustrated by the U.S. Census Bureau example mentioned earlier, is an example of a positively skewed distribution.

Variability

The third way in which distributions may be described is in terms of the amount or nature of differences among the different values for a variable. The **variability** of a distribution refers to the amount of differences in a distribution of data for a variable. The variability in a distribution is described by answering the question, “To what extent are the scores in the distribution different from each other?”

The amount of variability in a distribution can be estimated by looking at a frequency distribution table or figure. Imagine, for example, an instructor gives an 11-item quiz to her class and finds almost all of the students answer six, seven, or eight of the items correctly. In this situation, the distribution would resemble the shape of a tall mountain: thin and tall, with many of the scores residing in a few values of the variable. A distribution with a small amount of variability such as this one is referred to as a **peaked distribution** , defined as a distribution where much of the data are in a small number of values of a variable.

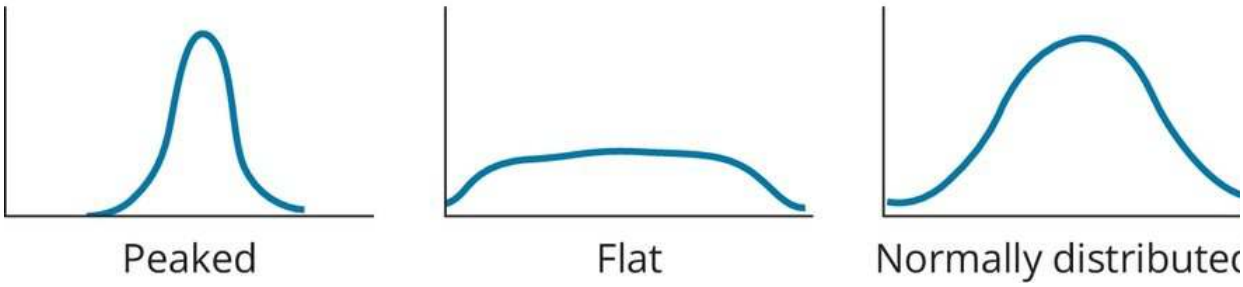


Figure 2.8 Distributions With Different Variability (Peaked, Flat, and Normally Distributed)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

On the other hand, what would the distribution look like if she had found her students' scores were distributed relatively evenly across the entire range of possible scores? In this situation, the distribution would not be tall and thin like a mountain but rather would be low and wide like a plateau. This type of distribution is referred to as a **flat distribution** , which is a distribution where the data are spread evenly across the values of a variable. The CNN gun control example described earlier is an example of a flat distribution.

In discussing modality, symmetry, and variability, distributions may have more than one mode, may be asymmetric (skewed), and may be peaked or flat. But what if a distribution has none of these qualities? A distribution that is unimodal, symmetric, and neither peaked nor flat may be described as "bell shaped." If you look at a bell, you'll notice that it is the highest in the middle, and as you move away from the top of a bell in both the left and right directions, the shape of the bell slopes downward in the same manner. A bell-shaped distribution for a variable is given a special name: a normally distributed variable.

A **normally distributed variable** is a distribution for a variable that is considered unimodal, symmetric, and neither peaked nor flat. Note that the word *normal* is a statistical concept and not a value judgment. In other words, asymmetric, peaked, or flat distributions are not considered "abnormal." The concept of normal distributions will be discussed in greater detail in Chapter 5. Figure 2.8 provides an illustration of three distributions with different amounts of variability: peaked, flat, and normally distributed.

Learning Check 3

Reviewing What You've Learned So Far



1. Review questions
    a. What are the three aspects used to describe distributions of variables?
    b. What is the difference between unimodal, bimodal, and multimodal distributions?
    c. What is the difference between symmetric and asymmetric (skewed) distributions?
    d. What is the difference between positively and negatively skewed distributions?
    e. What is the difference between peaked and flat distributions?
    f. What are the characteristics of a normally distributed variable?
2. For each of the following sets of data, determine whether the distribution of scores is unimodal, bimodal, or multimodal.
    a. 4, 2, 11, 6, 10, 2, 6, 13, 2, 10, 6, 10, 1
    b. 15, 17, 13, 18, 17, 14, 20, 18, 17, 14, 17
    c. 20, 10, 50, 10, 80, 30, 70, 10, 60, 10, 80, 0, 30, 80, 10, 100, 80
3. For each of the following sets of data, determine whether the distribution of scores is symmetric or asymmetric (skewed).
    a. 4, 5, 1, 4, 2, 4, 3, 5, 3, 4, 4, 3
    b. 220, 230, 250, 230, 200, 240, 210, 230, 270, 220, 240, 230, 240, 230
    c. 3, 1, 9, 2, 3, 6, 2, 7, 3, 5, 2, 6, 3, 2, 2
4. For each of the following sets of data, determine whether the distribution of scores is peaked or flat.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

a. 16, 10, 19, 12, 13, 15, 14, 17, 15, 20, 11, 13, 19, 16, 15
b. 12, 15, 12, 11, 13, 11, 12, 12, 14, 10, 12, 11, 16, 12, 13, 12, 11
c. 14, 12, 16, 14, 13, 15, 12, 14, 14, 15, 14, 14, 13, 14

2.7 Looking Ahead

This chapter has discussed why and how researchers examine data. Examining data by creating tables and figures requires a number of different skills, including organizational skills (to prepare a set of data for statistical analysis), communication skills (to present data in tables or figures), and conceptual skills (to describe a distribution). The next two chapters focus on mathematical skills related to the statistical analysis of data, describing how researchers describe distributions numerically rather than visually. Given that this is a statistics textbook, we will focus on calculating statistics to describe and analyze data. However, and we cannot say this strongly enough, it is *extremely* important to examine a set of data before conducting statistical analyses as this will prevent errors both in your calculations of statistics as well as your interpretation of the results of these calculations.

2.8 Summary

Before conducting statistical analyses, researchers typically examine their data to gain an initial sense of the data, detect data coding or data entry errors, identify ***outliers*** (rare, extreme scores that lie outside of the range of the majority of scores in a set of data), evaluate research methodology, and determine whether the data meet statistical criteria and assumptions.

One common way of examining variables involves creating tables known as ***frequency distribution tables*** (tables that summarize the number and percentage of participants that express the different values for a variable) and ***grouped frequency distribution tables*** (tables that group the values of a variable measured at the interval or ratio level of measurement into a small number of intervals and then provide the frequency and percentage within each interval). To create a grouped frequency table, the ***real limits*** of each interval (the values of the variable that fall halfway between the top of one interval and the bottom of the next interval) need to be identified.

Data for a variable may also be examined by creating figures such as bar charts, pie charts, histograms, and frequency polygons. ***Bar charts*** and ***pie charts*** are used for variables measured at the nominal or ordinal level of measurement; a bar chart uses bars to represent the frequency or percentage of a sample corresponding to each value of a variable, with the different bars not touching each other; a pie chart is a circle divided into proportions that represent the percentage of the sample corresponding to each value of a variable. ***Histograms*** and ***frequency polygons*** are used for variables measured at the interval or ratio level of measurement; a histogram is a figure in which bars are used to represent the frequency of each value of a variable, with the different bars touching each other; a frequency polygon is a line graph that uses data points to represent the frequency of each value of a variable, with lines connecting the data points.

From examining their data, researchers can describe the distribution of values or scores for a variable. There are three main aspects of distributions: modality, symmetry, and variability.

The ***modality*** of a distribution is the value(s) of the variable that have the highest frequencies or occur most often in a set of data. ***Unimodal distributions*** are distributions where one value occurs with the greatest frequency, ***bimodal distributions*** have two values with the greatest frequency, and ***multimodal distributions*** have more than two values with the greatest frequencies.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The ***symmetry*** of a distribution refers to how the frequencies of values of a variable change in relation to the most common or frequently occurring values of the variable. In ***symmetric distributions*** , the frequencies change in a similar manner, moving away in both directions from the most common values of the variable. In ***asymmetric (skewed) distributions*** , the frequencies change in a different manner, moving away in both directions from the most common values of the variable.

The ***variability*** of a distribution refers to the amount of differences in a distribution of data. In ***peaked distributions*** , much of the data are in a few values of a variable; in ***flat distributions*** , the data are spread evenly across the values of a variable. A ***normally distributed variable*** has a symmetric distribution that is neither peaked nor flat.

2.9 Important Terms

asymmetric (skewed) distribution

bar chart

bimodal distribution

cumulative percentage

flat distribution

frequency (*f*)

frequency distribution table

frequency polygon

grouped frequency distribution table

histogram

modality

multimodal distribution

negatively skewed distribution

normally distributed variable

outlier

peaked distribution

pie chart

positively skewed distribution

real limits

real lower limit

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

real upper limit

symmetric distribution

symmetry

unimodal distribution

variability

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

2.10 Formula Introduced in This Chapter

Percentage (%)

$$\% = \frac{f}{total\ number\ of\ scores} * 100$$

2.11 Using IBM SPSS Software *

Creating Data Files: Defining Variables and Entering Data

* IBM SPSS software. SPSS is a registered trademark of International Business Machines Corporation.

1. Open a new data file (See Figure 2.9).

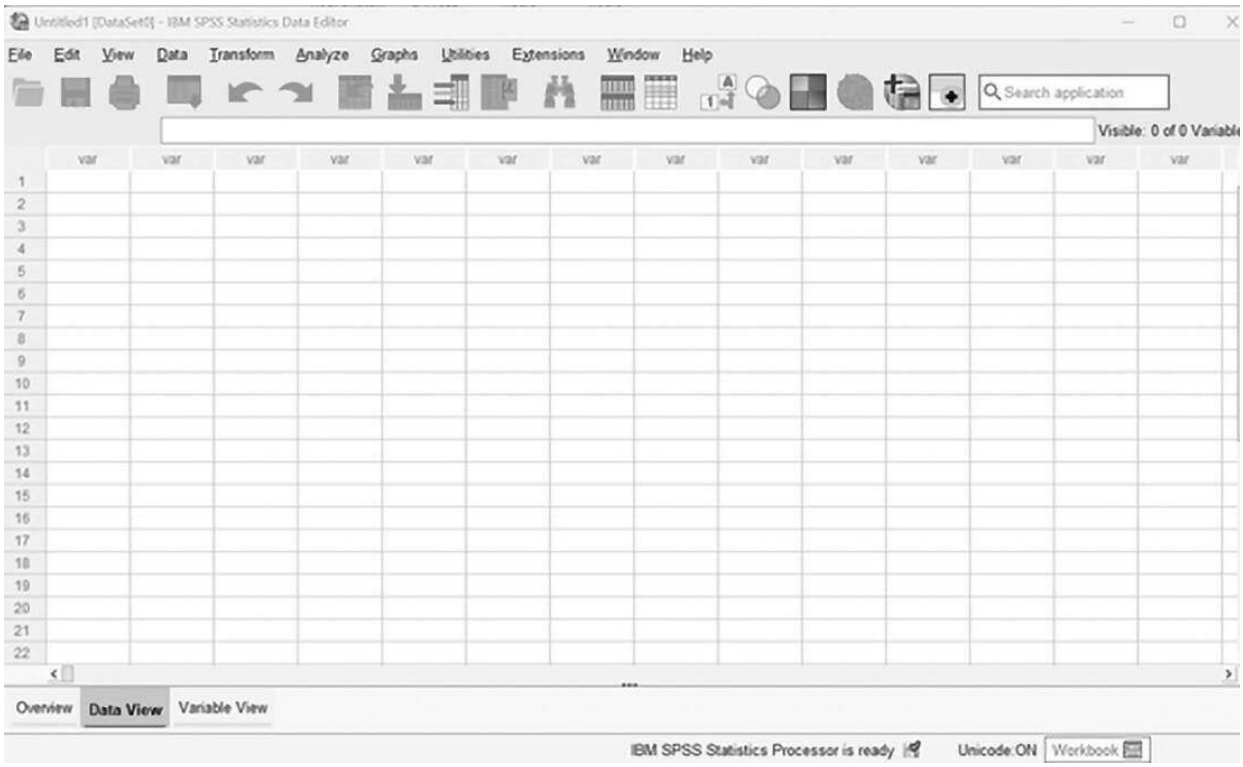


Figure 2.9 Data View

2. Define and name variable(s) (See Figure 2.10).

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

How? (1) Click **Variable View** tab, (2) click below **Name** , and (3) type name of variable in the cell.

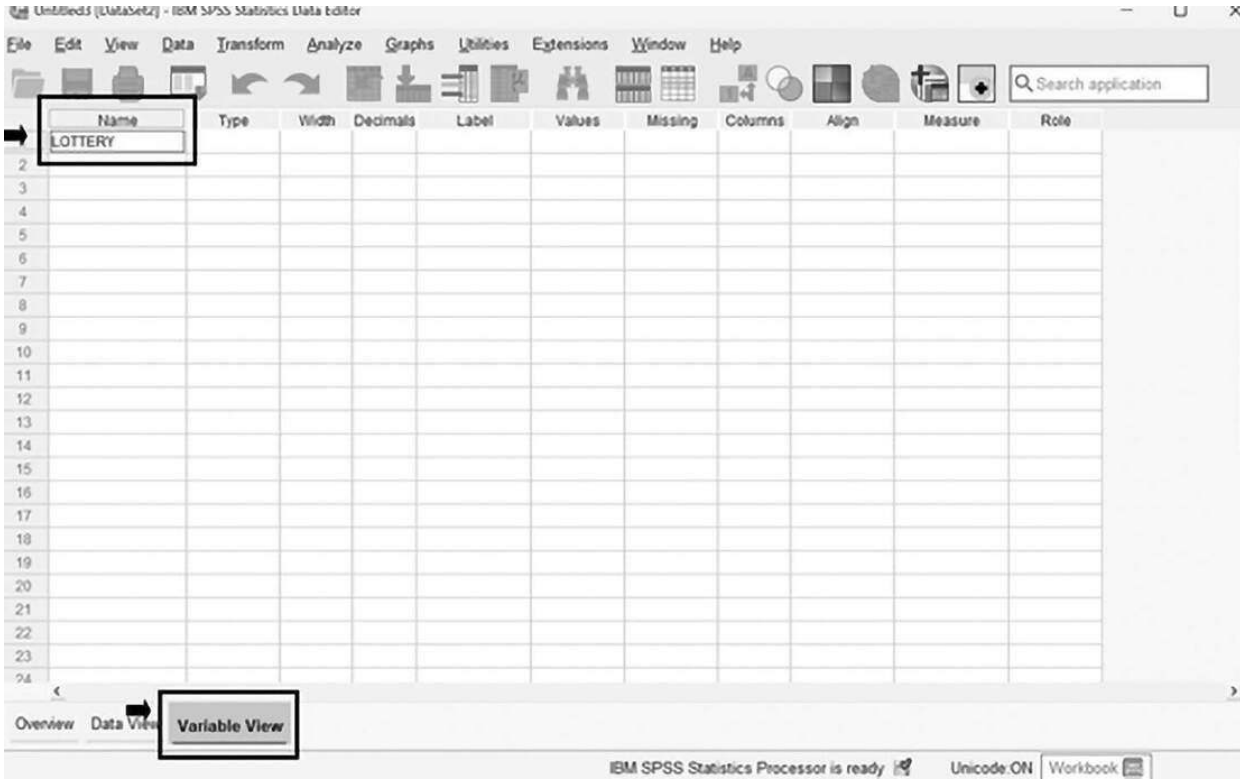


Figure 2.10 Variable View

3. Determine number of decimals for your variable(s) (See Figure 2.11).

How? Click desired number of decimal places in the **Decimals** box.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

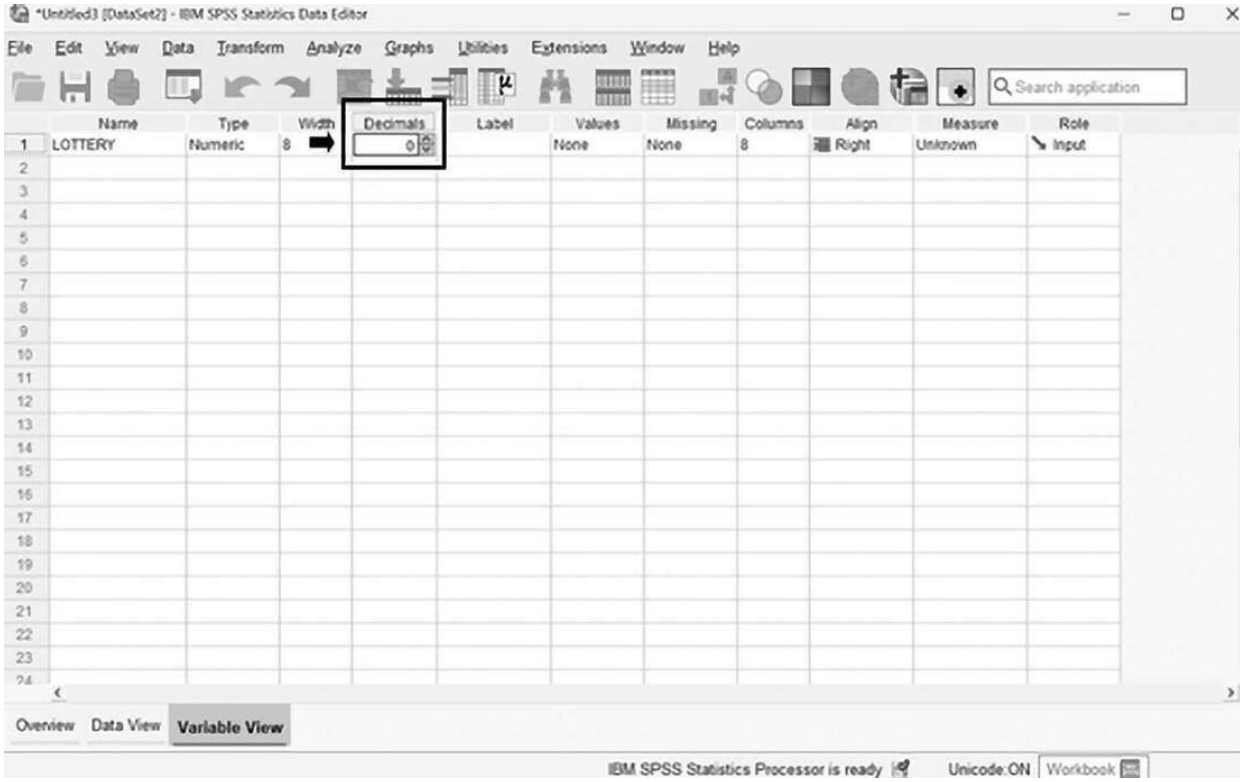


Figure 2.11 Decimals Box

4. Provide labels for your variable(s) (See Figure 2.12).

How? (1) Click below **Label** , and (2) type the label for the variable in the cell.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

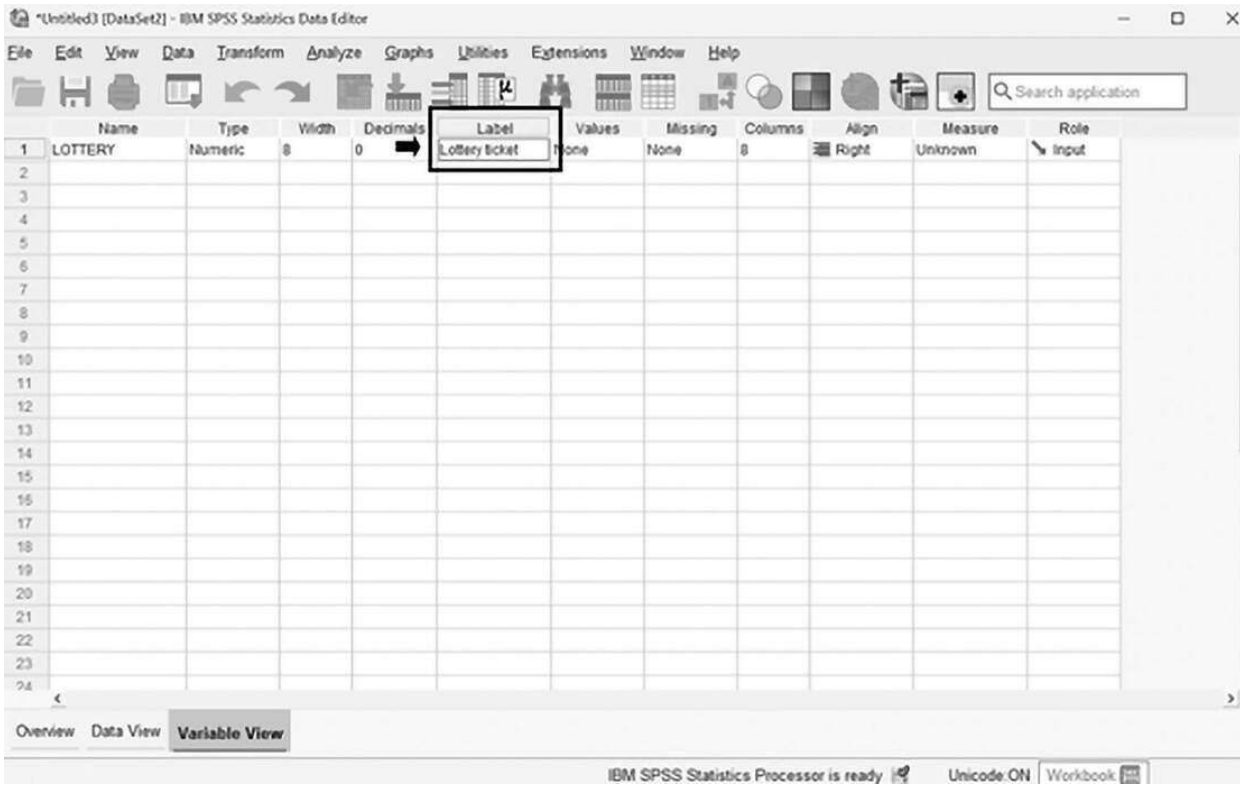


Figure 2.12 Label Box

5. Provide labels for values of nominal (categorical) variable(s) (See Figure 2.13).

How? (1) Click below **Values** , and (2) type labels for coded values of the variable.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

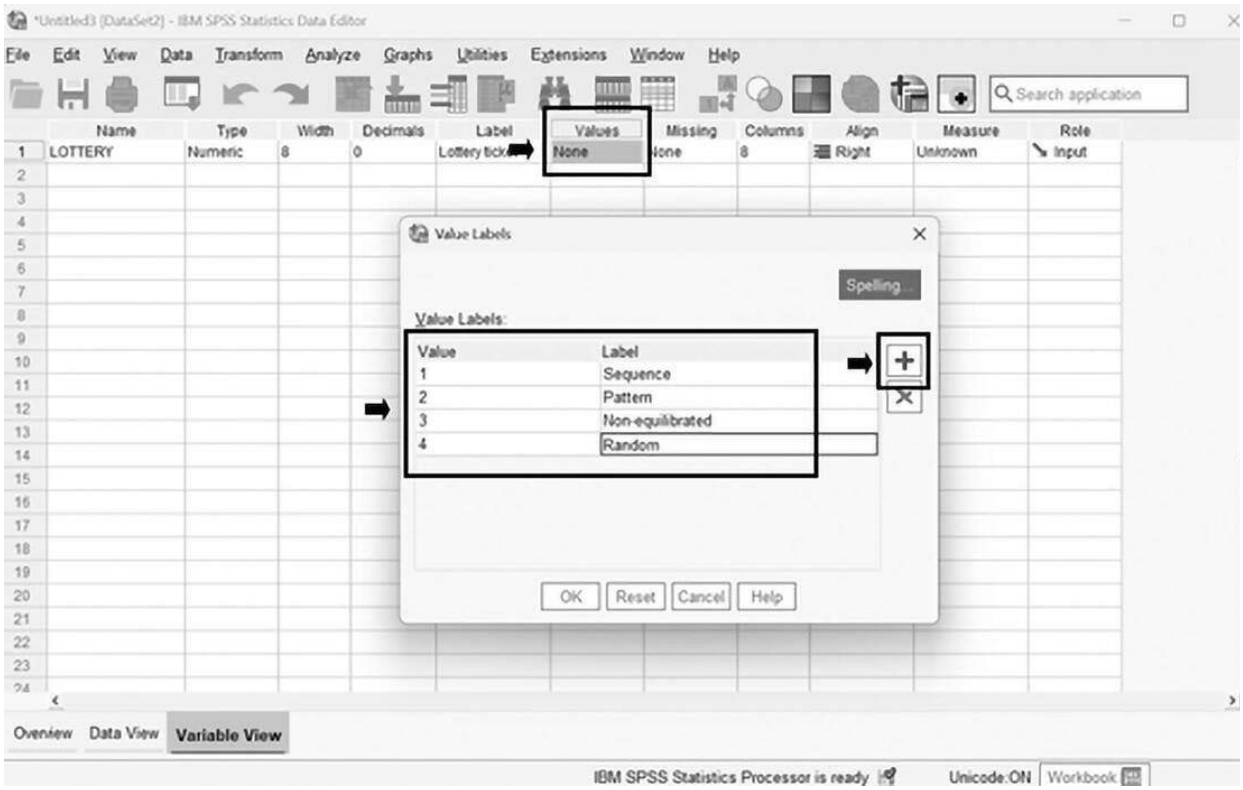


Figure 2.13 Nominal Variables

6. Enter data (See Figure 2.14).

How? (1) Click **Data View** tab and (2) enter data for the variable in the cells.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

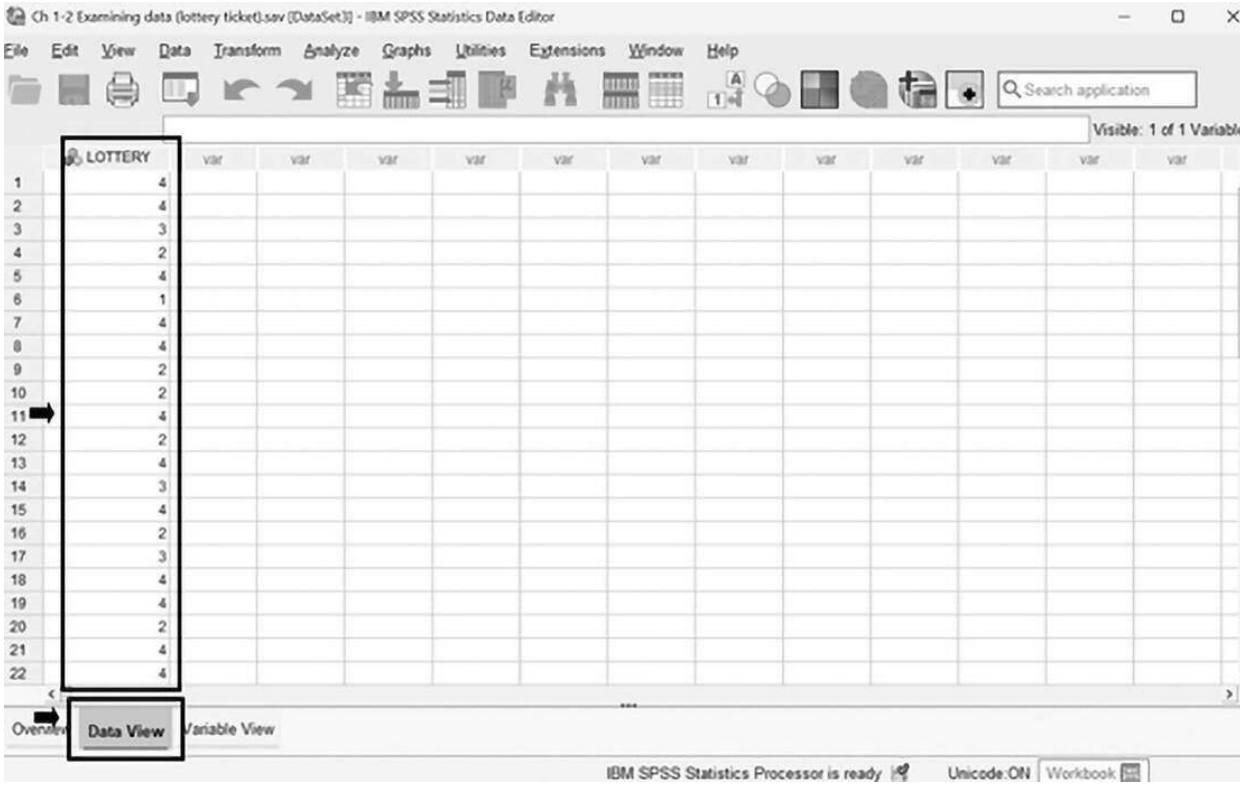


Figure 2.14 Enter Data View

Examining Data: Frequency Distribution Tables and Figures: The Lottery Study (2.1)

1. Select the frequency distribution procedure within SPSS (See Figure 2.15).

   How? (1) Click **Analyze** menu, (2) click **Descriptive Statistics** , and (3) click **Frequencies** .

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

To Jennifer , my kids Meagan and Eika , and my parents, Katsumi and Grayce Tokunaga.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

As a global academic publisher, Sage is driven by the belief that research and education are critical in shaping society. Our mission is building bridges to knowledge—supporting the development of ideas into scholarship that is certified, taught, and applied in the real world.

Sage's founder, Sara Miller McCune, transferred control of the company to an independent trust, which guarantees our independence indefinitely. This enables us to support an equitable academic future over the long term by building lasting relationships, championing diverse perspectives, and co-creating social and behavioral science resources that transform teaching and learning.

▶ Long Description:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

FUNDAMENTAL STATISTICS FOR THE SOCIAL AND BEHAVIORAL SCIENCES

Third Edition

Howard T. Tokunaga

San Jose State University



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

eISBN: 978-1-0718-6160-8

Copyright © 2027 by Sage.

All rights reserved. Except as permitted by U.S. copyright law, no part of this work may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without permission in writing from the publisher.

All third party trademarks referenced or depicted herein are included solely for the purpose of illustration and are the property of their respective owners. Reference to these trademarks in no way indicates any relationship with, or endorsement by, the trademark owner.



FOR INFORMATION:

2455 Teller Road, Thousand Oaks, California 91320, E-mail: orders@sagepub.com

3rd floor, HYLO, 103-105 Bunhill Row, London, EC1Y 8LZ, United Kingdom

10th Floor, Emaar Capital Tower 2, MG Road Sikanderpur, Sector 26, Gurugram, Haryana – 122002, India

18 Cross Street #10-10/11/12, China Square Central, Singapore 048423

Library of Congress Control Number: 2026943277

ISBN: 978-1-0718-6144-8

Printed in the United States of America

This book is printed on acid-free paper.

26 27 28 29 30 10 9 8 7 6 5 4 3 2 1

Acquisitions Editor: Anna Villarruel

Content Development Editor: Kenzie Offley

Editorial Assistant: Daria Terry

Production Editor: Syeda Aina Rahat Ali

Copy Editor: Michelle Ponce

Typesetter: Lumina Datamatics

Cover Designer: Candice Harman

Marketing Manager: Victoria Velasquez

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CONTENTS

Preface
Acknowledgments
About the Author
Chapter 1 Introduction to Statistics
Chapter 2 Examining Data: Tables and Figures
Chapter 3 Measures of Central Tendency
Chapter 4 Measures of Variability
Chapter 5 Normal Distributions
Chapter 6 Probability and Introduction to Hypothesis Testing
Chapter 7 Testing One Sample Mean
Chapter 8 Estimating the Mean of a Population
Chapter 9 Testing the Difference Between Two Means
Chapter 10 Errors in Hypothesis Testing, Statistical Power, and Effect Size
Chapter 11 One-Way Analysis of Variance (ANOVA)
Chapter 12 Two-Way Analysis of Variance (ANOVA)
Chapter 13 Correlation
Chapter 14 Linear Regression and Multiple Correlation
Chapter 15 Chi-Square
Appendix A: Tables
Appendix B: Review of Basic Mathematics
Glossary
References
Index

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

DETAILED CONTENTS

Preface

Acknowledgments

About the Author

Chapter 1 Introduction to Statistics

- 1.1 What Is Statistics?
- 1.2 Why Learn Statistics?
- 1.3 Introduction to the Stages of the Research Process
  - Developing a Research Hypothesis to Be Tested
    - Identifying a Question or Issue to Be Examined
    - Reviewing and Evaluating Relevant Theories and Research
    - Stating a Research Hypothesis: Independent and Dependent Variables
  - Collecting Data
    - Drawing a Sample From a Population
    - Determining How Variables Will Be Measured: Levels of Measurement
    - Selecting a Method to Collect the Data: Experimental and Nonexperimental Research Methods
  - Analyzing the Data
    - Calculating Descriptive Statistics
    - Calculating Inferential Statistics
  - Drawing a Conclusion Regarding the Research Hypothesis
  - Communicating the Findings of the Study
- 1.4 Plan of the Book
- 1.5 Looking Ahead
- 1.6 Summary
- 1.7 Important Terms
- 1.8 Exercises

Answers to Learning Checks

Learning Check 1

Learning Check 2

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Odd-Numbered Exercises

Chapter 2 Examining Data: Tables and Figures

2.1 An Example From the Research: Winning the Lottery

2.2 Why Examine Data?

Gaining an Initial Sense of the Data

Identifying Outliers

Evaluating Research Methodology

Determining Whether Data Meet Statistical Criteria and Assumptions

2.3 Examining Data Using Tables

What Are the Values of the Variable?

How Many Participants in the Sample Have Each Value of the Variable?

What Percentage of the Sample Has Each Value of the Variable?

2.4 Grouped Frequency Distribution Tables

Cumulative Percentages

2.5 Examining Data Using Figures

Displaying Nominal and Ordinal Variables: Bar Charts and Pie Charts

Displaying Interval and Ratio Variables: Histograms and Frequency Polygons

Drawing Inappropriate Conclusions From Figures

2.6 Examining Data: Describing Distributions

Modality

Symmetry

Variability

2.7 Looking Ahead

2.8 Summary

2.9 Important Terms

2.10 Formula Introduced in This Chapter

Percentage (%)

2.11 Using IBM SPSS Software *

Creating Data Files: Defining Variables and Entering Data

Examining Data: Frequency Distribution Tables and Figures: The Lottery Study (2.1)

2.12 Exercises

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 3 Measures of Central Tendency
3.1 An Example From the Research: The 10% Myth
3.2 Understanding Central Tendency
3.3 The Mode
Determining the Mode
3.4 The Median
Calculating the Median: Odd Number of Scores
Calculating the Median: Even Number of Scores
Calculating the Median When Multiple Scores Have the Median Value
3.5 The Mean
Calculating the Mean
Reporting Calculated Numbers
Rounding Calculated Numbers
Calculating the Mean From Frequency Distribution Tables
The Mean as a Balancing Point in a Distribution
The Sample Mean vs. the Population Mean: Statistics vs. Parameters
3.6 Comparison of the Mode, Median, and Mean
Strengths and Weaknesses of the Mode
Strengths and Weaknesses of the Median
Strengths and Weaknesses of the Mean
Choosing a Particular Measure of Central Tendency
3.7 Measures of Central Tendency: Drawing Conclusions
3.8 Looking Ahead
3.9 Summary
3.10 Important Terms
3.11 Formulas Introduced in This Chapter
Median (Odd Number of Scores)
Median (Even Number of Scores)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Mean
Mean (Frequency Distribution Table)
Mean (Balancing Point in a Distribution)
3.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 4 Measures of Variability
4.1 An Example From the Research: How Many Sometimes" in an Always"?
4.2 Understanding Variability
4.3 The Range
Strengths and Weaknesses of the Range
4.4 The Interquartile Range
Strengths and Weaknesses of the Interquartile Range
4.5 The Variance (s^2)
Definitional Formula for the Variance
Computational Formula for the Variance
Why Not Use the Absolute Value of the Deviation in Calculating the Variance?
Why Divide by N − 1 Rather Than N in Calculating the Variance?
4.6 The Standard Deviation (s)
Definitional Formula for the Standard Deviation
Computational Formula for the Standard Deviation
Eyeball Estimating" the Mean and Standard Deviation
4.7 Measures of Variability for Populations
The Population Variance (σ^2)
The Population Standard Deviation (σ)
Measures of Variability for Samples vs. Populations
4.8 Measures of Variability for Nominal Variables
The Index of Qualitative Variation
4.9 Measures of Variability: Drawing Conclusions

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

4.10 Looking Ahead
4.11 Summary
4.12 Important Terms
4.13 Formulas Introduced in This Chapter
Range
Interquartile Range
Variance (s^2) (Definitional Formula)
Variance (s^2) (Computational Formula)
Standard Deviation (s) (Definitional Formula)
Standard Deviation (s) (Computational Formula)
Population Variance (σ^2) (Definitional Formula)
Population Standard Deviation (σ) (Definitional Formula)
Index of Qualitative Variation (IQV)
4.14 Using SPSS
Calculating Measures of Central Tendency and Variability: The Frequency Rating Study (4.1)
4.15 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 5 Normal Distributions
5.1 Normal Distributions
What Are Normal Distributions?
Why Are Normal Distributions Important?
Is There Only One Normal Distribution?
5.2 The Standard Normal Distribution
What Is the Standard Normal Distribution?
Interpreting z-Scores: Distance From the Mean
Interpreting z-Scores: Relative Position Within the Distribution
Finding the Area Between the Mean and a z-Score

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Finding the Area Less or Greater Than a z-Score
Finding the Area Between Two z-Scores
5.3 Applying z-Scores to Normal Distributions
Transforming Normal Distributions Into the Standard Normal Distribution
Understanding the z-Score Formula
Example: Transforming SAT Scores Into z-Scores
Interpreting z-Scores: Distance From the Mean
Interpreting z-Scores: Relative Position Within Distribution
Finding the Area Between the Mean and a Score
Finding the Area Less or Greater Than a Score
Finding the Area Between Two Scores
Transforming Scores Into z-Scores: Drawing Conclusions
5.4 Standardizing Frequency Distributions
Example: Measuring Politicians' Behavior
Standardizing a Frequency Distribution
Interpreting Standardized Scores
Using Standardized Scores to Compare or Combine Variables
Characteristics of Standardized Distributions
When Can You Apply the Normal Curve Table to a Standardized Distribution?
5.5 Looking Ahead
5.6 Summary
5.7 Important Terms
5.8 Formulas Introduced in This Chapter
z-Score
Standardized Score
5.9 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 6 Probability and Introduction to Hypothesis Testing

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

6.1 A Brief Introduction to Probability
- What Is Probability?
- Why Is Probability Important to Researchers?
- Applying Probability to Normal Distributions
- Applying Probability to Binomial Distributions

6.2 Example: Making Heads or Tails of the Super Bowl

6.3 Introduction to Hypothesis Testing
- State the Null and Alternative Hypotheses (H_0 and H_1)
  - The Null Hypothesis (H_0)
  - The Alternative Hypothesis (H_1)
- Make a Decision About the Null Hypothesis
  - Defining a Low" Probability of a Statistic: Setting Alpha (α)
  - Identifying the Values of the Statistic With a Low Probability
  - Stating a Decision Rule
  - Calculating a Value of a Statistic
  - Making the Decision Whether to Reject the Null Hypothesis
- Draw a Conclusion From the Analysis
- Relate the Result of the Analysis to the Research Hypothesis
- Summary of Steps in Hypothesis Testing

6.4 Issues Related to Hypothesis Testing: An Introduction
- The Issue of Proof" in Hypothesis Testing
- Errors in Decision Making
- Factors Influencing the Decision About the Null Hypothesis
  - Sample Size
  - Alpha (α)
  - Directionality of the Alternative Hypothesis

6.5 Looking Ahead

6.6 Summary

6.7 Important Terms

6.8 Formula Introduced in This Chapter

Probability

6.9 Exercises

Answers to Learning Checks

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 7 Testing One Sample Mean
7.1 An Example From the Research: Do You Read Me?
7.2 The Sampling Distribution of the Mean
Characteristics of the Sampling Distribution of the Mean
7.3 Inferential Statistics: Testing One Sample Mean (σ Known)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: z-Test for One Mean
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the z-Test for One Mean
Summary
7.4 A Second Example From the Research: Unique Invulnerability
7.5 Introduction to the t-Distribution
Characteristics of the Student t-Distribution
7.6 Inferential Statistics: Testing One Sample Mean (σ Not Known)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: t-Test for One Mean
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the t-Test for One Mean

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Summary
7.7 Factors Affecting the Decision About the Null Hypothesis
Sample Size
Alpha (α)
Directionality of the Alternative Hypothesis
7.8 Looking Ahead
7.9 Summary
7.10 Important Terms
7.11 Formulas Introduced in This Chapter
z-Test for One Mean
Population Standard Error of the Mean $\left(\mathbf{i}t\sigma_{\mathbf{i}t\overline{X}}\right)$
Degrees of Freedom (df) for One Sample
t-Test for One Mean
Standard Error of the Mean $\left(\mathbf{i}ts_{\mathbf{i}t\overline{X}}\right)$
7.12 Using SPSS
Testing One Sample Mean (σ Not Known): The Unique Invulnerability Study (7.6)
7.13 Exercises
Answers to Learning Checks
Learning Check 2
Learning Check 3
Learning Check 5
Learning Check 6
Answers to Odd-Numbered Exercises
Chapter 8 Estimating the Mean of a Population
8.1 An Example From the Research: Salary Survey
8.2 Introduction to the Confidence Interval for the Mean
Using the Sampling Distribution of the Mean to Create the Confidence Interval for the Mean
8.3 The Confidence Interval for the Mean (σ Not Known)
State the Desired Level of Confidence
Calculate the Confidence Interval and Confidence Limits
Calculate the Standard Error of the Mean $\left(s_{\overline{X}}\right)$

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Calculate the Degrees of Freedom (df) and Identify the Critical Value of t

Calculate the Confidence Interval and Confidence Limits

Draw a Conclusion About the Confidence Interval

Probability of an Interval or the Probability of a Population Mean?

Summary

8.4 The Confidence Interval for the Mean (σ Known)

State the Desired Level of Confidence

Calculate the Confidence Interval and Confidence Limits

Calculate the Population Standard Error of the Mean $\left(\sigma_{\overline{X}}\right)$

Identify the Critical Value of z

Calculate the Confidence Interval and Confidence Limits

Draw a Conclusion About the Confidence Interval

Summary

8.5 Factors Affecting the Width of the Confidence Interval for the Mean

Sample Size

Estimating the Sample Size Needed for a Desired Confidence Interval

Level of Confidence

Choosing Between Different Levels of Confidence: Probability Versus Precision

8.6 Interval Estimation and Hypothesis Testing

Concerns About Hypothesis Testing

The Interpretation of Statistical Significance and Nonsignificance

The Influence of Sample Size on the Decision About the Null Hypothesis

Interval Estimation as an Alternative to Hypothesis Testing

8.7 Looking Ahead

8.8 Summary

8.9 Important Terms

8.10 Formulas Introduced in This Chapter

Confidence Interval for the Mean (σ Not Known)

Confidence Interval for the Mean (σ Known)

Estimated Sample Size for the Confidence Interval for the Mean

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

8.11 Using SPSS
Calculating the confidence interval for the mean (The salary survey study (8.1))
8.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 9 Testing the Difference Between Two Means
9.1 An Example From the Research: You Can Just Wait
9.2 The Sampling Distribution of the Difference
Characteristics of the Sampling Distribution of the Difference
9.3 Inferential Statistics: Testing the Difference Between Two Sample Means
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: t-Test for Independent Means
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the t-Test for Independent Means
Summary
9.4 Inferential Statistics: Testing the Difference Between Two Sample Means (Unequal Sample Sizes)
An Example From the Research: Kids' Motor Skills and Fitness
Inferential Statistics: Testing the Difference Between Two Sample Means (Unequal Sample Sizes)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Relate the Result of the Analysis to the Research Hypothesis

Assumptions of the t-Test for Independent Means (Unequal Sample Sizes)

Summary

9.5 Inferential Statistics: Testing the Difference Between Paired Means

Research Situations Appropriate for Within-Subjects Research Designs

An Example From the Research: Web-Based Family Interventions

Inferential Statistics: Testing the Difference Between Paired Means

Calculate the Difference Between the Paired Scores

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Draw a Conclusion From the Analysis

Relate the Result of the Analysis to the Research Hypothesis

Assumptions of the t-Test for Dependent Means

Summary

9.6 Looking Ahead

9.7 Summary

9.8 Important Terms

9.9 Formulas Introduced in This Chapter

Degrees of Freedom (df), Difference Between Two Sample Means

t-Test for Independent Means

Standard Error of the Difference $\left(s_{\overline{X}_1-\overline{X}_2}\right)$

Standard Error of the Difference (Unequal Sample Sizes)

Degrees of Freedom (df), Difference Between Paired Means

t-Test for Dependent Means

Standard Error of Difference Scores (s D)

9.10 Using SPSS

Testing the Difference Between Two Sample Means: The Parking Lot Study (9.1)

Testing the Difference Between Paired Means: The Web-Based Intervention Study (9.5)

9.11 Exercises

Answers to Learning Checks

Learning Check 1

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Answers to Odd-Numbered Exercises
Chapter 10 Errors in Hypothesis Testing, Statistical Power, and Effect Size
10.1 Hypothesis Testing vs. Criminal Trials
10.2 An Example From the Research: Truth or Consequences
Correct Decisions and Mistakes in Criminal Trials
10.3 Two Errors in Hypothesis Testing: Type I and Type II Error
Type I Error: The Risk in Rejecting the Null Hypothesis
What Is the Probability of Making a Type I Error?
What Is the Probability of Not Making a Type I Error?
Why Is Type I Error a Concern?
Type II Error: The Risk in Not Rejecting the Null Hypothesis
What Is the Probability of Making a Type II Error?
What Is the Probability of Not Making a Type II Error?
Why Is Type II Error a Concern?
Why Does Type II Error Occur?
Type I and Type II Error: Summary
10.4 Controlling Type I and Type II Error
Controlling Type I Error
Concerns About Controlling Type I Error
Controlling Type II Error
Increasing Sample Size
Raising Alpha (α)
Using a Directional Alternative Hypothesis (H1)
Increasing Between-Group Variability
Decreasing Within-Group Variability
Controlling Type I and Type II Error: Summary
10.5 Measures of Effect Size
Interpreting Inferential Statistics: The Influence of Sample Size
The Inappropriateness of “Highly Significant”

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Effect Size as Variance Accounted for

Measure of Effect Size for the Difference Between Two Sample Means: r^2

Interpreting Measures of Effect Size

Presenting Measures of Effect Size

Reasons for Calculating Measures of Effect Size

A Second Measure of Effect Size for the Difference Between Two Sample Means: Cohen's d

10.6 Looking Ahead

10.7 Summary

10.8 Important Terms

10.9 Formulas Introduced in This Chapter

Probability of Type I Error

Probability of Type II error

r Squared (r^2)

Cohen's d

10.10 Exercises

Answers to Learning Checks

Learning Check 1

Learning Check 2

Learning Check 3

Learning Check 4

Answers to Odd-Numbered Exercises

Chapter 11 One-Way Analysis of Variance (ANOVA)

11.1 An Example From the Research: It's Your Move

11.2 Introduction to Analysis of Variance (ANOVA)

Review: Testing the Difference Between the Means of Two Samples

Understanding Between-Group and Within-Group Variability

Relating Between-Group and Within-Group Variability

Introduction to Analysis of Variance (ANOVA) and the F-Ratio

Characteristics of the F-Ratio Distribution

11.3 Inferential Statistics: One-Way Analysis of Variance (ANOVA)

State the Null and Alternative Hypotheses (H_0 and H_1)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Value, and State a Decision Rule
Calculate a Statistic: F-Ratio for the One-Way ANOVA
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (R^2)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Summary
11.4 A Second Example: The Parking Lot Study Revisited
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Value, and State a Decision Rule
Calculate a Statistic: F-Ratio for the One-Way ANOVA
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (R^2)
Draw a Conclusion From the Analysis
The Relationship Between the t-Test and the F-Ratio
11.5 Analytical Comparisons Within the One-Way ANOVA
Planned Versus Unplanned Comparisons
Between-Group and Within-Group Variance in Analytical Comparisons
Conducting Planned Comparisons
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Conducting Unplanned Comparisons
Concerns Regarding Unplanned Comparisons
Methods of Controlling Familywise Error

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Controlling Familywise Error: The Tukey Test

11.6 Looking Ahead

11.7 Summary

11.8 Important Terms

11.9 Formulas Introduced in This Chapter

Degrees of Freedom for Between-Group Variance, One-Way ANOVA (df_{BG})

Degrees of Freedom for Within-Group Variance, One-Way ANOVA (df_{WG})

Between-Group Variance, One-Way ANOVA (MS_{BG})

Total Sample Mean, One-Way ANOVA $\left(\overline{X}_T\right)$

Within-Group Variance, One-Way ANOVA (MS_{WG})

F-Ratio, One-Way ANOVA (F)

Measure of Effect Size, One-Way ANOVA (R^2)

Relationship Between t-Test and F-Ratio for the One-Way ANOVA

F-Ratio, Analytical Comparison (F_{comp})

Between-Group Variance, Analytical Comparison $\left(MS_{\mathrm{BG}_{\mathrm{comp}}}\right)$

Measure of Effect Size, Analytical Comparison $\left(R^2_{\mathrm{comp}}\right)$

Critical Value, Tukey Test for Unplanned Comparisons (F_T)

11.10 Using SPSS

One-Way Analysis of Variance (ANOVA): The Chess Study (11.1)

Analytical Comparisons: Planned Comparisons—Explanation vs. Prediction (11.5)

11.11 Exercises

Answers to Learning Checks

Learning Check 2

Learning Check 3

Learning Check 4

Answers to Odd-Numbered Exercises

Chapter 12 Two-Way Analysis of Variance (ANOVA)

12.1 An Example From the Research: Vote—or Else!

12.2 Introduction to Factorial Research Designs

Testing Interaction Effects

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

An Example of Interaction Effect

An Example of No Interaction Effect

Testing Main Effects

The Relationship Between Main Effects and Interaction Effects

Factorial Designs vs. Single-Factor Designs: Advantages and Disadvantages

12.3 The Two-Factor (A × B) Research Design

Notational System for Components of the Two-Factor Research Design

Analyzing the Two-Factor Research Design: Descriptive Statistics

Notational System for Means in the Two-Factor Research Design

Examining the Descriptive Statistics

12.4 Introduction to Analysis of Variance (ANOVA) for the Two-Factor Research Design

F-Ratios in the Two-Factor Research Design

12.5 Inferential Statistics: Two-Way Analysis of Variance (ANOVA)

State the Null and Alternative Hypotheses (H_0 and H_1)

Main Effect of Factor A

Main Effect of Factor B

A × B Interaction Effect

Make Decisions About the Null Hypotheses

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Values, and State Decision Rules

Calculate Statistics: F-Ratios for the Two-Way ANOVA

Make Decisions Whether to Reject the Null Hypotheses

Determine the Levels of Significance

Calculate Measures of Effect Size (R^2)

Draw Conclusions From the Analyses

Relate the Results of the Analyses to the Research Hypothesis

12.6 Investigating a Significant A × B Interaction Effect: Analysis of Simple Effects

12.7 Looking Ahead

12.8 Summary

12.9 Important Terms

12.10 Formulas Introduced in This Chapter

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Degrees of Freedom for Main Effect of Factor A, Two-Way ANOVA (df_A)
Degrees of Freedom for Main Effect of Factor B, Two-Way ANOVA (df_B)
Degrees of Freedom for A × B Interaction Effect, Two-Way ANOVA ($df_{A\times B}$)
Degrees of Freedom for Within-Group Variance, Two-Way ANOVA (dfWG)
Between-Group Variance for Main Effect of Factor A, Two-Way ANOVA (MS_A)
Between-Group Variance for Main Effect of Factor B, Two-Way ANOVA (MS_B)
Sums of Squares for Between-Group Variance, Two-Way ANOVA (SS_{BG})
Between-Group Variance for A × B Interaction, Two-Way ANOVA ($MS_{A\times B}$)
Within-Group Variance, Two-Way ANOVA (MS_{WG})
F-Ratio for Main Effect of Factor A, Two-Way ANOVA (F_A)
F-Ratio for Main Effect of Factor B, Two-Way ANOVA (F_B)
F-Ratio for A × B Interaction, Two-Way ANOVA ($F_{A\times B}$)
Measure of Effect Size for Main Effect of Factor A, Two-Way ANOVA(R^2_A)
Measure of Effect Size for Main Effect of Factor B, Two-Way ANOVA(R^2_B)
Measure of Effect Size for A × B Interaction, Two-Way ANOVA(R^2_{AB})
12.11 Using SPSS
Two-Way Analysis of Variance (ANOVA): The Voting Message Study (12.1)
12.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 13 Correlation
13.1 An Example From the Research: It's Good for You!
13.2 Introduction to the Concept of Correlation
Describing the Relationship Between Variables
Nature of the Relationship
Direction of the Relationship
Strength of the Relationship

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Measuring the Relationship Between Variables: Correlational Statistics
Introduction to the Pearson Correlation Coefficient (r)
Calculating Correlational Statistics: The Role of Variance and Covariance
13.3 Inferential Statistics: Pearson Correlation Coefficient
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: Pearson Correlation Coefficient (r)
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (r^2)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the Pearson Correlation
A Computational Formula for the Pearson r
Represent the Covariance Between X and Y (SPXY)
Represent the Variance of Variable X (SSX)
Represent the Variance of Variable Y (SSY)
Calculate the Pearson Correlation Coefficient (r)
13.4 Correlating Two Sets of Ranks: The Spearman Rank-Order Correlation
Introduction to Ranked Variables
An Example From the Research: Do Students and Teachers Think Alike?
Inferential Statistics: The Spearman Rank-Order Correlation Coefficient
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
13.5 Correlational Statistics vs. Correlational Research
Does ANOVA = Causality?
“Correlation Is Not Causality”—What Does That Mean?
13.6 Looking Ahead

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

13.7 Summary
13.8 Important Terms
13.9 Formulas Introduced in This Chapter
Degrees of Freedom for Pearson Correlation (df)
Sum of Products (Definitional Formula) (SPXY)
Sum of Squares for the X Variable (Definitional Formula) (SSX)
Sum of Squares for the Y Variable (Definitional Formula) (SSY)
Pearson Correlation (Definitional Formula) (r)
Measure of Effect Size for the Pearson Correlation Coefficient (r^2)
Sum of Products (Computational Formula) (SPXY)
Sum of Squares for the X Variable (Computational Formula) (SSX)
Sum of Squares for the Y Variable (Computational Formula) (SSY)
Pearson Correlation (Computational Formula) (r)
Spearman Rank-Order Correlation (rs)
Measure of Effect Size for the Spearman Rank-Order Correlation $\left(r_s^2\right)$
13.10 Using SPSS
Pearson Correlation: The Picky Eater Study (13.1)
Spearman Rank-Order Correlation: The Business Student Study (13.4)
13.11 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Answers to Odd-Numbered Exercises
Chapter 14 Linear Regression and Multiple Correlation
14.1 Predicting One Variable From Another: Linear Regression
The Linear Regression Equation
Calculating the Linear Regression Equation
Calculate the Slope of the Equation (b)
Calculate the Y-Intercept of the Equation (a)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Report the Linear Regression Equation

Drawing the Linear Regression Equation

A Second Example: Put Down Your Pencils

Calculating the Linear Regression Equation

Drawing the Linear Regression Equation

14.2 Correlation With Two or More Predictors: Introduction to Multiple Correlation and Regression

The Limitations of Using Only One Predictor

Introduction to Multiple Correlation

Multiple Correlation With Two Predictors: An Example

Measuring the Relationship Between Two Predictors and a Criterion: The Multiple Correlation (R)

Testing the Multiple Correlation for Statistical Significance

Predicting a Criterion From Two Predictors: The Multiple Regression Equation

Conclusion

14.3 Looking Ahead

14.4 Summary

14.5 Important Terms

14.6 Formulas Introduced in This Chapter

Linear Regression Equation

Estimate of Slope (b)

Estimate of Y-Intercept (a)

Multiple Correlation Coefficient (R)

F-Ratio for the Multiple Correlation Coefficient (F)

Multiple Regression Equation

Regression Coefficient

Estimate of Y-Intercept (a) (Multiple Regression)

14.7 Using SPSS

Linear Regression: The Picky Eater Study (14.1)

Multiple Regression: The College Admission Example (14.2)

14.8 Exercises

Answers to Learning Checks

Learning Check 1

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 2

Answers to Odd-Numbered Exercises

Chapter 15 Chi-Square

15.1 An Example From the Research (One Categorical Variable): Are You My Type?

15.2 Introduction to the Chi-Square Statistic

Observed and Expected Frequencies: The Basis of the Chi-Square Statistic

Characteristics of the Distribution of Chi-Square Statistics

15.3 Inferential Statistic: Chi-Square Goodness-of-Fit Test

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Value, and State a Decision Rule

Calculate a Statistic: Chi-Square (χ^2)

Make a Decision Whether to Reject the Null Hypothesis

Determine the Level of Significance

Calculate a Measure of Effect Size (Cramér's φ)

Draw a Conclusion From the Analysis

Relate the Result of the Analysis to the Research Hypothesis

Assumptions Underlying the Chi-Square Statistic

Conducting the Chi-Square Goodness-of-Fit Test With Unequal Hypothesized Proportions

Calculate Expected Frequencies (f_e)

Calculate the Chi-Square Statistic (χ^2)

15.4 An Example From the Research (Two Categorical Variables): Seeing Red

15.5 Inferential Statistic: Chi-Square Test of Independence

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Value, and State a Decision Rule

Calculate a Statistic: Chi-square (χ^2)

Make a Decision Whether to Reject the Null Hypothesis

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Determine the Level of Significance
Calculate a Measure of Effect Size (Cramér's φ)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
15.6 Parametric and Nonparametric Statistical Tests
Parametric vs. Nonparametric Statistical Tests
Reasons for Using Nonparametric Statistical Tests
Examples of Nonparametric Statistical Tests
15.7 Looking Ahead
15.8 Summary
15.9 Important Terms
15.10 Formulas Introduced in This Chapter
Degrees of Freedom for the Chi-Square Goodness-of-Fit Test
Expected Frequency for the Chi-Square Goodness-of-Fit Test
Chi-Square Statistic (χ^2)
Cramér's φ for the Chi-Square Goodness-of-Fit Test
Degrees of Freedom for the Chi-Square Test of Independence
Expected Frequency for the Chi-Square Test of Independence
Cramér's φ for the Chi-Square Test of Independence
15.11 Using SPSS
Chi-Square Goodness-of-Fit Test—Equal Hypothesized Proportions: The Personality Study (15.1)
Chi-Square Goodness-of-Fit Test—Unequal Hypothesized Proportions: The Personality Study (15.3)
Chi-Square Test of Independence: The Color Preference Study (15.4)
Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Learning Check 6

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Odd-Numbered Exercises

Appendix A: Tables

Appendix B: Review of Basic Mathematics

I. Symbols

II. Terms

III. Addition, Subtraction, Multiplication, Division

A. All Positive Numbers

B. Adding All Negative Numbers

C. Adding a Combination of Positive and Negative Numbers

D. Subtracting Negative Numbers

E. Multiplying Two Negative Numbers

F. Multiplying a Combination of Positive and Negative Numbers

G. Dividing Two Negative Numbers

H. Dividing One Positive Number and One Negative Number

IV. Order of Operations

V. Fractions, Decimals, Percentages

A. Fraction

B. Decimal

C. Percentage

VI. Solving Simple Equations With One Unknown (X)

A. X Has a Value Subtracted From It (Solve for X Using Addition)

B. X Has a Value Added to It (Solve for X Using Subtraction)

C. X Is Multiplied by a Value (Solve for X Using Division)

D. X Is Divided by a Value (Solve for X Using Multiplication)

VII. Solving More Complex Equations With One Unknown (X)

A. X Is Multiplied by a Value and Then Has a Value Either Added to It or Subtracted From It (Solve for X Using Either Subtraction or Addition and Then Division)

B. X Is Divided by a Value and Then Has a Value Either Added to It or Subtracted From It (Solve for X Using Either Addition or Subtraction and Then Multiplication)

C. X Has a Value Added to It or Subtracted From It and Is Then Multiplied by a Value (Solve for X Using Division and Then Either Subtraction or Addition)

D. X Has a Value Either Added to It or Subtracted From It and Is Then Divided by a Value (Solve for X Using Multiplication and Then Either Subtraction or

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Addition)
Glossary
References
Index

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

PREFACE

A Few (?) Words About Me

When I entered college, I thought my path would take me to law school; however, taking an Intro to Psych class made me realize I enjoy the challenge of trying to understand the human mind. The college I attended, U.C. Santa Cruz, was a fairly unconventional university, but somehow in the midst of a sea of humanistic psychologists I became attracted to the empirical and methodological aspects of psychology by taking classes from instructors such as David Harrington and Dane Archer, who showed me that statistics could appeal to students if taught with a gentle, guiding approach that addresses questions relevant to students' lives. After graduating from college, I was able to get a job as a research assistant for a human resource consulting firm. Despite my lack of work experience, I was hired primarily as a result of having taken statistics and research methods courses, which taught me that learning statistics has benefits both inside and outside of the classroom.

Several years later, I started grad school at U.C. Berkeley, where two events critical to this book took place. First, serving as a teaching assistant, I found I really enjoyed helping students, particularly in statistics and research methods classes often approached with fear and suspicion. Second, I took graduate classes from Geoff Keppel, who had developed his own unique system for analyzing experimental research designs. His lectures and books showed me that statistics can be taught in a systematic way that highlights similarities rather than differences between different research situations. Geoff managed to transform something as daunting-sounding as a "3 x 2 x 4 research design" into the mathematical equivalent of playing with wooden toy alphabet blocks labeled "A," "B," and "C."

After leaving Cal, I was hired by San Jose State, where I teach an introductory statistics course aimed at students with a wide range of backgrounds, ability, and motivation. When I went to select a textbook for the course, for the first time I looked carefully at the wide range of offerings. What I found striking (and still find striking) was that the majority of books focused on providing formulas and very small sets of data designed to demonstrate how to correctly calculate the correct numbers from these formulas. Little emphasis, however, was given to what these numbers meant or implied. Given my own experiences learning and using statistics, I thought a book was needed that focusses on how statistics may be used to answer questions and test ideas within the larger research process.

Goal of this Book

This textbook, aimed primarily at students majoring in the behavioral and social sciences, includes many of the same topics and chapters as the majority of existing introductory statistics books. However, what I believe distinguishes this book is its thematically-based approach that places statistics within the research process. The primary purpose of this

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

book is to not just teach students how to calculate statistics but how to interpret the results of statistical analyses in light of a study's research hypothesis, and to communicate one's results and interpretations to a broader audience. Hopefully, this book will not only help students understand the purpose and use of statistics, but it will also give them a greater appreciation of how research studies are conceived, conducted, and communicated.

The book is intended for students with a limited knowledge of mathematics, statistics, and research design and methodology. Some common obstacles faced by students are their math anxiety, fear of conducting arithmetic calculations, and an inability to see the relevance of the material to their daily lives. To address these concerns, this books attempts to judiciously use mathematical terminology, write in a conversational manner, and utilize actual research situations rather than contrived sets of numbers.

Organization of the Book

The 15 chapters of this book may be placed into three general categories. Chapters 1 to 4 are designed to introduce students to the research process and how data that have been collected may be organized, presented, and summarized:

Chapter 1 (Introduction to Statistics) defines the concept of statistics, explains why statistics are important to learn, and describes the steps in the research process that are the framework for the rest of the book.

Chapter 2 (Examining Data: Tables and Figures) discusses the importance of examining data that have been collected and describes how tables and figures may be created to represent data.

Chapter 3 (Measures of Central Tendency) presents measures of central tendency (the mode, median, and the mean) and compares them in terms of their relative strengths and weaknesses.

Chapter 4 (Measures of Variability) presents measures of variability (range, interquartile range, variance, standard deviation, index of qualitative variation) and the concept of parameters (population variance and standard deviation).

Chapters 5 through 10 discuss the process of conducting statistical analyses to test research questions and hypotheses, as well as issues and controversies regarding this process:

Chapter 5 (Normal Distributions) introduces theoretical distributions, starting with normal distributions, before moving on to the standard normal distribution (normal curve).

Chapter 6 (Probability and Introduction to Hypothesis Testing) presents the structure for later chapters by introducing, defining, and illustrating steps in hypothesis testing.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Chapter 7 (Testing One Sample Mean) tests one sample mean using the *z* -test and *t* -test and introduces sampling distributions, the central limit theorem, and the Student *t* distribution.

Chapter 8 (Estimating the Mean of a Population) introduces interval estimation by developing the confidence interval for the mean and discusses interval estimation as an alternative to null hypothesis significance testing (NHST).

Chapter 9 (Testing the Difference Between Two Means) covers the *t* -test for independent means for both equal and unequal sample size and introduces within-subjects research designs using the *t* -test for dependent means.

Chapter 10 (Errors in Hypothesis Testing, Statistical Power, and Effect Size) discusses Type I and Type II error and introduces measures of effect size as a way of drawing appropriate conclusions about an analysis.

Chapters 11 to 15 discuss different statistical procedures used in research situations that vary in the number of independent variables as well as how the variables have been measured:

Chapter 11 (One-way Analysis of Variance (ANOVA)) discusses the one-way ANOVA, the relationship between the *t* -statistic and the F-ratio, and planned and post-hoc comparisons.

Chapter 12 (Two-way Analysis of Variance (ANOVA)) introduces factorial research designs, the calculation of F-ratios for main effects and interaction effects, and illustrates the use of simple effects analyses to test research hypotheses.

Chapter 13 (Correlation) defines and illustrates the concept of correlation, discusses the Pearson and Spearman correlation coefficients, and makes the distinction between correlational statistics and correlational research methods.

Chapter 14 (Linear Regression and Multiple Correlation) introduces the concept of regression, discusses linear regression, and works through an example of multiple regression with two predictor variables.

Chapter 15 (Chi-Square) describes the analysis of categorical variables using the chi-square goodness-of-fit test and the chi-square test of independence before ending the book with a brief introduction to nonparametric statistics.

In most cases, the chapters in this book share a uniform format and structure, centered on the research process, based on examples of actual published research studies addressing topics relevant to students' lives. For each example, the research hypothesis is clearly identified; after a brief description of how the data in the study were collected, the calculation and interpretation of both descriptive and inferential statistics are presented,

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ending with the extent to which the results of the statistical analyses support the research hypothesis and how these findings may be communicated to a broader audience.

Pedagogical Features

The chapters in this textbook contain several reoccurring features designed to enhance students' comprehension of the material. These include the following:

Periodic learning checks within chapters. Short learning checks, in the form of conceptual review questions and computational exercises, are provided throughout each chapter.

Explicit presentation of formulas and calculations. This book provides both definitional formulas to highlight their relation to the concept being measured and computational formulas to facilitate calculations. In applying these formulas, all of the steps in the calculation of statistics such as the *t* -test and F-ratio are included to help students in completing exercise and homework assignments.

Illustration of the presentation and communication of analyses. Every chapter includes a discussion and illustration of how statistical analyses may be reported as well how they may be communicated to a larger audience in both written and visual form.

Highlighted definition of terms. Every term and concept introduced in this book is boldfaced, immediately followed by an explicit definition.

End of chapter summaries and exercises. The end of each chapter contains a summary of critical points, a list of terms and formulas introduced in the chapter, and a large number of exercises, many of which refer to published research studies.

Calculation of statistics using SPSS. Many chapters include end-of-chapter content that provides an annotated presentation of how statistics may be calculated using the SPSS software. Screen shots of data files, menu selections, and output files are included and annotated to locate and identify relevant aspects of the analysis.

What's New to the 3rd Edition

I'd like to start by expressing my most heartfelt appreciation to the instructors and reviewers of the first two editions of this book as their evaluation, insights, criticisms, and suggestions were absolutely vital to the changes made in this third edition. I'd like to particularly thank the instructors who've adopted the book who were kind enough to spontaneously send me their thoughts and suggestions. It is my sincere hope that all of you see the incorporation of your input in this edition.

There were two main goals in this edition. The first goal was to make the writing as brief and concise as possible, manifested in the goal of reducing the length of each chapter by 5% to

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

10%. This goal was achieved by simplifying the language and shortening or eliminating what I, the book's adopters, and the book's reviewers felt was unnecessary detail or discussion. The second goal was to increase and update the exercises at the end of many of the chapters, including the use of recent published research as the basis of these exercises. Another goal was to update the demonstration of the use of SPSS to analyze data using the most recent version (Version 31 at the time of this book's publication).

More specifically, in terms of content, the following changes were made to these chapters:

Chapter 1 (Introduction to Statistics): The published research studies used to illustrate concepts in the text, learning checks, and exercises were added to and updated using topics relevant to students (e.g., fear of missing out (FOMO), cryptocurrency, legalization of cannabis, Gen Z, COVID, fake news).

Chapter 6 (Probability and Introduction to Hypothesis Testing): The Super Bowl example used to introduce and illustrate the process of hypothesis testing was updated.

Chapter 8 (Estimating the Mean of a Population): The main research example, which involves data from an actual national salary survey, has been updated throughout this chapter with more current salary data.

As with the first two editions, any and all questions, concerns, suggestions, and complaints you have about the book are most welcome and can be sent to me at howard.tokunaga@sjsu.edu .

A Few Tips for Students

To you, the college student about to read this book as part of taking a statistics course: "Welcome!" and "Great job!" We welcome you because you're about to embark on a semester-long journey that we hope enhances your skills and widens your perspective; we congratulate you because it's a journey not everyone is willing to take. We say to our students that by taking just one statistics course they are ahead of 95% of the population in their knowledge of statistics.

We also tell our students that our primary goal for them is not to become a statistics genius but just to become comfortable with the process of analyzing and interpreting data. In that regard, we agree with a 2012 *Forbes* magazine article entitled *The Six Classes That Will Make Any College Grad Employable* that noted "a good statistics class will help the student become comfortable pushing numbers around and avoiding false interpretations."

At the present moment, we know our encouragement may be of little comfort because you might be anxious about having to learn statistics. Some of you might be anxious about having to learn statistical concepts and formulas; others might be anxious because the research process seems pretty complicated. Talking with students who have taken our courses over the years has helped us assemble the following advice:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Master the Material Presented in the Early Chapters

Chapters 1 through 6 discuss how research is typically conducted, how data are summarized and described, and the process by which researchers conduct statistical analyses to test their ideas. It is absolutely critical for you to have a firm understanding of these chapters as they lay the groundwork for later chapters that discuss a variety of statistical procedures used by researchers to test hypotheses.

What does it mean to master this material? First, *read the chapters both before and after they're discussed in class* . Reading the chapters beforehand helps you identify anything that's unclear and have your questions addressed by your instructor; re-reading the chapters afterward helps confirm your understanding of the material. Next, *be able to define and explain key conceptspresented in these chapters* . These concepts are highlighted in bold-faced letters, and it's important to learn them when they're first introduced because they'll appear throughout the remainder of the book. Next, *do the learning checks within the chapters and the exercises at the end of the chapters* , because they include exercises and review questions you can ask yourself to assess your understanding of the material. Most importantly, *do not miss class during the early part of the semester* ! It's been our experience that students who miss critical lectures at the beginning of the term often have difficulty keeping up as the semester continues.

Review High School Algebra

If you happen to move beyond the introductory statistics course for which you're reading this book, you'll find that statistics are often calculated using computers and statistical software rather than hand calculations. Consequently, some statistics textbooks no longer include mathematical formulas but instead have students use statistical software or Microsoft Excel. However, this book takes a different approach for two main reasons. First, we believe that in order to understand the results of statistical analyses, one must understand their underlying foundation. Learning statistics via computer software can lead to a "brain-dead" approach in which students are at the mercy (rather than control) of their computers. We refer to this as the "I'm only as smart as my printout" method of learning statistics, also known as "Because this is what my computer told me."

Second, this book contains mathematical formulas and calculations for one simple reason: The statistics in this book are not hard to calculate. To assess whether you're prepared to read this book, take the following test:

1. Do you know how to add, subtract, multiply, and divide?
2. Do you know how to calculate the square or square root of a number?
3. Do you understand the order of operations and parentheses within mathematical formulas?

If your answers to these questions are "Yes," congratulations! You possess the ability to conduct *every* mathematical calculation in this book because *none* of the formulas involve geometry, trigonometry, or calculus. However, if you're unsure of your mathematical ability, we suggest you review your high school algebra. Also, there's an appendix at the end of this

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

book that includes the mathematical concepts and operations needed to conduct the statistical analyses in this book. We highly recommend you read this appendix to review and assess your mathematical skills.

Learn to Use a Statistical Calculator

Although learning the statistics in this book doesn't require the use of computers or software, you will need a hand calculator. Calculators with statistical capabilities are typically identified as "statistical" or "scientific" calculators. Although there's a wide selection of brands and models, know that the calculators most appropriate for this book cost only $10 to $15. We do *not* recommend you use or purchase a more expensive calculator! We prefer these simpler calculators because students are able to do their calculations more quickly and with fewer errors. We do suggest you purchase a calculator similar to the one your instructor uses as they will be more familiar with its features and idiosyncrasies.

In selecting a calculator, look closely at the keyboard; you'll need a calculator whose keys have labels such as $\bar{X}$, σ, and ΣX (Chapters 3 and 4 discuss what these symbols represent). When you go calculator shopping, you'll find calculators that have keys with labels such as ΣY and XY. Although calculators with these Y symbols are only slightly more expensive than calculators with just the X symbols, we do *not* recommend their purchase because these features are not needed to perform the calculations in this book. In fact, we've found students with Y calculators have a more difficult time performing simpler calculations.

Ask Questions

Conducting research is the process of asking and answering questions. Accordingly, throughout this book issues are often framed in the form of questions, which we hope encourages you to ask questions as well. The first person you should direct any questions to is yourself. As you work your way through the chapters, get into the habit of asking yourself, "What does this mean?" A good test of your level of comprehension is whether you can explain the material to yourself in a meaningful way.

You should also be sure to work with your instructor to confirm your level of understanding and clarify unanswered questions. Instructors often say to students, "The only bad question is the one you don't ask." Don't be hesitant about asking questions in class. Instructors will tell you they use one student's question to assess an entire class's level of understanding. This is what we call the "pencil test"—when one student asks a question, we look to see how many other pencils get raised, which reflects the number of students who had the same question. Asking questions enhances the learning experience for yourself, your classmates, and your instructor.

Form a Study Group

No matter how much we encourage students to ask questions in class, it seems they hesitate for fear of drawing attention to themselves. As a result, we turn to another well-

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

traveled saying: “Misery loves company.” Forming a study group with other students will help you keep up with reading assignments, confirm your comprehension of the material, clarify any confusion you have, prepare for exams, and provide social support. We do suggest that your group contain at least one person who will raise the group’s questions and concerns to the instructor during class or office hours.

Practice, Practice, Practice

The best way to learn the topics covered by this book is to obtain as much practice as possible. Throughout the chapters are examples in which the calculations have been worked out for you; in the middle and end of each chapter are a number of exercises. The answers to some of the exercises are provided at the back of this book, and your instructor has access to the answers to the other exercises. Do the end-of-chapter exercises even if your instructor does not assign them as homework in order to identify any reoccurring mistakes you make. We find that the vast majority of errors students make are simple computational errors, committed when students do the calculations too quickly.

Also note that the publisher of this book, Sage Publications, has developed a webpage for this book that contains a huge amount of supplemental tools such as videos and quizzes for each chapter. Find it by going to the Sage Publishing website **collegepublishing.sagepub.com** , searching for my name (Tokunaga), and looking for the **Resources** tab.

Enough cheerleading and preaching – let’s get started!

WELCOME!

GREAT JOB!

Additional Teaching Resources

Visit **collegepublishing.sagepub.com** , and navigate to the Resources tab on your book’s page to find the teaching materials designed to accompany this textbook. On this site you will find an array of materials that will save you time and to help you keep students engaged, including the following:

Learning management system cartridges that easily integrate with your course management system so student test results and graded assignments seamlessly flow into your gradebook

Test banks , aligned to Bloom’s Taxonomy, that provide a diverse range of test items, including multiple choice, true/false, and essay questions

Lecture notes that provide an outline and the key concepts in each chapter to aid in lecture preparation

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

PowerPoint ® slides that offer a flexible, accessible, and customizable solution for creating multimedia lectures

Tables and Figures from the book to support lecture preparation and class discussions

Sample course syllabi , including suggested models for structuring your course

Answers to even-numbered questions from the text help facilitate grading

SPSS datasets for use with exercises from the text.

Multimedia content , including original SAGE videos featuring tutorials with author Howard T. Tokunaga that bring concepts to life and appeal to diverse learners

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ACKNOWLEDGMENTS

This book has traveled a long journey, with many people playing both transient and persevering roles in its development. First and foremost, I'd like to thank my kids, Meagan and Will, for providing me perspective and inspiration amid all of the twists and turns we faced while this book was being written. I'm so very blessed to have you as my "burden." Away from home, I'm grateful to my graduate students at San Jose State—working on this book has given me much empathy for their thesis-related travails. However, they should be aware that the completion of this book ends the unspoken agreement that I not ask, "So how's your thesis?" if they don't ask, "So how's your book?" A number of friends have provided most welcome relief after long days staring at a computer screen, with particular thanks going to Ellie, Eileen, Susan, Bonnie, Ann, Lisa, and Terry. I would like to thank Charles Linsmeier and David Shirley for their many helpful comments and advice on earlier versions of this book, Julie West for her help in developing chapter exercises, and Youn-Jeong Choi and Naomi Ran for their careful examination of the chapters and exercises.

I would like to express my appreciation to everyone at SAGE who has played a role in shepherding this book to completion. First and foremost, I must express my gratitude to my editor, Vicki Knight. It has been a joy and an honor to work with someone with such a truly exceptional combination of experience, knowledge, perspective, and humor. In production I was fortunate enough to work with others on the SAGE team who made the process a pleasure: Yvonne McDuffee, editorial assistant; Katie Guarino, assistant editor; Nicole Elliott, executive marketing manager; Candice Harman, cover designer; and the copyeditor, Michelle Ponce. Furthermore, I am grateful to the reviewers who took the time to read the chapters, raise questions, and provide constructive comments and suggestions to assist me in clarifying and strengthening the manuscript. The following are the reviewers for third edition:

Erin M. Fekete, University of Indianapolis

Brooke Magnus, Boston College

Loretta Neal McGregor; Arkansas State University

Michael T. Moore, Adelphi University

Megan M. Ringel, St. Edward's University

Bettina Viereck, University of Hartford

Janice Weaver, Community College of Allegheny County

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ABOUT THE AUTHOR

Howard T. Tokunaga is professor of Psychology at San Jose State University, where he serves as coordinator of the MS Program in Industrial/Organizational (I/O) Psychology and teaches undergraduate and graduate courses in statistics, research methods, and I/O psychology. He received his bachelor's degree in psychology at U.C. Santa Cruz and his PhD in psychology at U.C. Berkeley. In addition to his teaching, he has consulted with a number of public sector and private sector organizations on a wide variety of management and human resource issues. He is author of *Moving from IBM SPSS to R and RStudio: A Statistics Companion* , and coauthor (with G. Keppel) of *Introduction to Design and Analysis* : *A Student's Handbook* . In his spare time, he enjoys the outdoors, museums, live theater, and being as physically, socially, and politically active as possible.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CHAPTER 1 INTRODUCTION TO STATISTICS

Learning Objectives

1.1 Define the concept of statistics.

1.2 Describe different reasons for learning statistics.

1.3 Describe the main steps in the research process and the tasks and concepts within each step.

In introducing this book to you, we assume you're a college student taking what is perhaps your first course in statistics to fulfill a requirement for your major or a general education requirement. If so, you may be asking yourself two questions:

What is statistics?

Why learn statistics?

The ultimate goal of this book is to help you begin to answer these two questions.

1.1 What Is Statistics?

You may not know it, but you encounter a variety of "statistics" in your day-to-day activities: the typical cost of going to college, the yearly income of the average college graduate, the average price of a car, and so on. So what exactly is "statistics"?

The Merriam-Webster dictionary defines **statistics** as *a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data* . When most people think about statistics, they focus on the "analysis" aspect of this definition—that is to say, they focus on numbers that result from analyzing data. However, statistics not only is concerned about how data are *analyzed* but also how data are *collected* and how the results of statistical analyses are *interpreted* and *communicated* . The main purpose of this book is to introduce, describe, and illustrate the role of statistics within the larger research process.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1.2 Why Learn Statistics?

There are many reasons why you should learn statistics. First, not only do you currently encounter statistics in your daily activities, but *throughout your life you are affected by the results of research and statistical analyses* . Which college you attend is based in part on test scores developed by researchers. You may have to take a personality or intelligence test to get a job. The choices of drugs and medicines available to you are based on medical research. Learning about statistics will help you become a more informed and aware consumer of research and statistical analyses that affect many aspects of your life.

A second reason for learning statistics is that *you may be required to interpret the results of statistical analyses* . Many college courses require students to read academic research journal articles. Evaluating published research is complicated by the fact that different people studying the same topic may come up with diverse or even opposing conclusions. Understanding statistics and their role in the research process will help you decide whether conclusions drawn in research articles are appropriate and justified.

Another reason for learning statistics is that *you might be required to do your own research* . College courses sometimes have students design and conduct mini-research studies; undergraduate majors might encourage students to do senior honors theses; graduate programs often require masters' theses and doctoral dissertations. Learning to collect and analyze data will help you address your own questions in an objective, systematic manner.

A final reason for learning statistics is that it may *help you in your future career* . The website Careercast.com conducts an annual survey in which they evaluate jobs on four dimensions: environmental factors, income, employment outlook, and stress. Recently, their #1 rated job was "data scientist," and the #2 rated job was "statistician." They noted, "A statistician's skill set can be used to break down and analyze large quantities of data. The demand for these skills spans a variety of industries, including marketing, banking, government, sports, retail, and even healthcare."

It's generally a good idea for students to maintain a healthy level of curiosity or even skepticism in regard to their education. However, we find that when it comes to learning statistics, students' feelings may be characterized as fear or anxiety. We understand these feelings, and we hope the benefits of learning statistics will become clear to you and help overcome any concerns you have.

1.3 Introduction to the Stages of the Research Process

Much of scientific research involves asking questions. Throughout this book, we'll examine a broad range of questions regarding human attitudes and behavior that contemporary researchers have asked and attempted to answer. The following are research questions we'll address in this chapter to introduce the stages of the research process:

> Does appealing to one's fear of missing out (FOMO) affect the intention to invest in cryptocurrencies?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Do pro-cannabis messages on social media lead to cannabis use?

Do parents differ in their awareness of their children's online activities?

Do nonbinary college students have the same experiences as cis men or cis women?

How might you try to answer questions such as these? You could base your answers on your personal beliefs, or you could adopt the answers given to you by others. But rather than relying on subjective beliefs and feelings, researchers in many academic disciplines test their ideas using the scientific method. The **scientific method** is a method of investigation that uses the objective and systematic collection and analysis of empirical data to test theories and hypotheses.

At its simplest, this book will portray the scientific method as consisting of five main steps or phases:

developing a research hypothesis to be tested,

collecting data,

analyzing the data,

drawing a conclusion regarding the research hypothesis, and

communicating the findings of the study.

Accomplishing each of these five steps requires completing a number of tasks, as shown in Figure 1.1 . Because this sequence of steps serves as the model for the research process we'll discuss throughout this book, each step is briefly introduced in the following sections. It's important to understand that the research process depicted in Figure 1.1 represents an ideal way of doing research. The "real" way, as you may discover in your own efforts, is often anything but a smooth ride but rather is filled with false starts, dead ends, and wrong turns.

Developing a Research Hypothesis to Be Tested

The initial stage—and the first step—of the research process is to develop a research hypothesis to be tested. A **research hypothesis** may be defined as a statement regarding an expected or predicted relationship between variables. A **variable** is a property or characteristic of an object, event, or person that can take on different values. One example of a variable is "U.S. state," a variable with 50 possible values (Alabama, Arkansas, etc.).

Research hypotheses are usually developed through the completion of several tasks:

identifying a question or issue to be examined,

reviewing and evaluating relevant theories and research, and

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

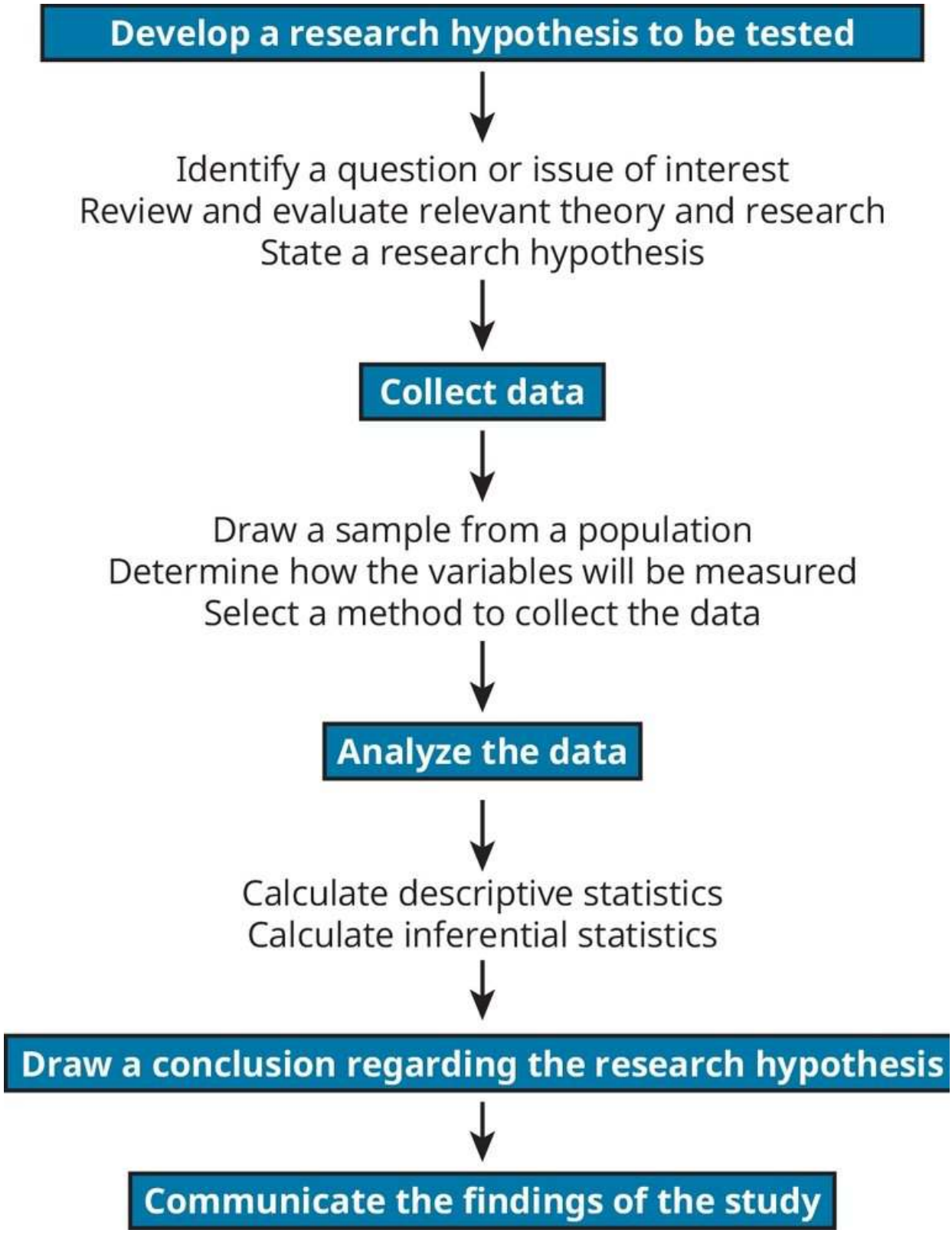
stating a research hypothesis.

Identifying a Question or Issue to Be Examined

Most research starts with a question posed by the researcher that often comes from the researcher's own ideas and daily observations. Although this may not seem terribly scientific, there's an advantage in using one's own experience as a starting point: People are generally much more motivated to explore issues that concern them personally. In teaching statistics, we advise students developing their own research projects to study something of interest to them. Conducting research can be tedious, difficult, and frustrating. At various points, you may ask yourself, "Why am I doing this?" Being able to provide a satisfactory answer to this question will help you overcome obstacles you encounter along the way.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

▶ Long Description: Figure 1.1 Steps in the Research Process Within the Scientific Method

Figure 1.1 Steps in the Research Process Within the Scientific Method

Reviewing and Evaluating Relevant Theories and Research

Beyond your own curiosity, research questions often arise from studying the theories, ideas, and research of others. A **theory** is a set of propositions used to describe or explain a phenomenon. The purpose of a theory is to summarize and explain specific facts using a logically consistent framework. Placing a question within a theoretical framework provides guidance and structure to research.

Reviewing and evaluating existing theories and research helps you decide whether it's worth the time and energy required to conduct a study. By seeing what others have done, you may conclude that the current way of thinking on a topic is incomplete or mistaken. By doing this review, you're able to ensure that the research you conduct adds to and improves upon an existing body of knowledge.

Stating a Research Hypothesis: Independent and Dependent Variables

Understanding and evaluating an existing literature not only helps develop a question of interest, but it may also lead to a predicted answer to that question. Within the scientific method, this answer is stated as a research hypothesis, defined earlier as a statement regarding an expected or predicted relationship between variables. Table 1.1 lists the research questions and research hypotheses for the studies listed at the beginning of this section. For example, the first research hypothesis states that "consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos."

One characteristic of research hypotheses such as those listed in Table 1.1 is that they identify the variables that are the focus of their research. As mentioned earlier, a variable is a property or characteristic with different values. Variables can be classified in several ways. In specifying a research hypothesis, researchers often speak in terms of "independent" and "dependent" variables. An **independent variable** may be defined as a variable manipulated or controlled by the researcher. A **dependent variable** , on the other hand, is a variable measured by the researcher that is expected to change or vary as a function of the independent variable. The independent variable is seen as the "cause" or "treatment," whereas the dependent variable is the "effect" or "outcome."

Researchers are interested in examining the effect of the independent variable on the dependent variable. Consider the first research hypothesis provided in Table 1.1 : "Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos." Here, the independent variable is the type of appeal consumers are exposed to: FOMO or non-FOMO. The dependent variable is the intention to invest in cryptos. In this study, the effect of the independent variable on the dependent variable is that

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

differences in the intention to invest in cryptos (the dependent variable) may “depend” upon which type of appeal (FOMO or non-FOMO) a consumer receives. Table 1.2 lists the independent and dependent variables for each of the research hypotheses in Table 1.1 .

In addition to identifying variables, a second characteristic of research hypotheses is that they specify the nature and direction of the relationship between variables. For example, the first research hypothesis in Table 1.1 states that “Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *higher* intention to invest in cryptos.” The word *higher* indicates the nature and direction of the relationship between the independent variable (type of appeal) and the dependent variable (intention to invest in cryptos). The direction of the relationship would not have been stated if the hypothesis had included the less specific phrase “will have a *different* intention,” which simply indicates that the two groups of consumers are not expected to be the same. Table 1.3 provides directional and nondirectional research hypotheses for the research studies from Table 1.1 .

Table 1.1 Examples of Research Questions and Hypotheses

Research Question	Research Hypothesis
Does appealing to one’s fear of missing out (FOMO) affect the intention to invest in cryptocurrencies? (Friederich et at., 2024)	Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.
Do pro-cannabis messages on social media lead to cannabis use? (Willoughby et al., 2024)	Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.
Do parents differ in their awareness of their children’s online activities? (Symons et al., 2017)	Mothers have more accurate knowledge of their children’s online activities than do fathers.
Do nonbinary college students have the same experiences as cis men or cis women? (Marx et al., 2024)	Nonbinary college students experience higher levels of victimization than do cis men or cis women.

The ability to form a directional research hypothesis is dependent on the state of the literature on the question of interest. If little or perhaps conflicting research has been conducted, researchers may not be able to form a directional hypothesis. For example, one study examining the relationship between exercise deprivation (not allowing people to exercise) and depression. reported, “We did not have a directional hypothesis, and when participants asked what we expected to find in this study, we replied: ‘We really are not sure since the results of earlier work on exercise deprivation are mixed’” (Mondin et al., 1996, p. 1200).

Table 1.2 Research Hypotheses and Their Independent and Dependent Variables

Research Hypothesis	Independent Variable	Dependent Variable
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher	Type of appeal	Intention to invest in cryptos

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Independent Variable	Dependent Variable
intention to invest in cryptos.		
Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.	Exposure to pro-cannabis messages	Intention to use cannabis
Mothers have more accurate knowledge of their children's online activities than do fathers.	Parent	Knowledge of children's online activities
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	Gender identity	Amount of victimization

Collecting Data

Once a research hypothesis has been formulated, researchers are ready to proceed to the second stage in the research process: collecting data. This step is seen as being composed of three tasks:

drawing a sample from a population,

determining how the variables will be measured, and

selecting a method by which to collect the data.

Table 1.3 Directional and Nondirectional Research Hypotheses

Directional Research Hypothesis	Nondirectional Research Hypothesis
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *higher* intention to invest in cryptos.	Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *different* intention to invest in cryptos.
Exposure to pro-cannabis messages on social media is related to *increased* intentions to use cannabis.	Exposure to pro-cannabis messages on social media is *related to* intentions to use cannabis.
Mothers have *more* accurate knowledge of their children's online activities than do fathers.	Mothers have *different* knowledge of their children's online activities than do fathers.
Nonbinary college students experience *higher* levels of victimization than do cis men or cis women.	Nonbinary college students experience *different* levels of victimization than do cis men or cis women.

Drawing a Sample From a Population

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The first step in collecting data is to identify the group of participants to which the research hypothesis applies. The group to which the results of a study may be applied or generalized is called a population. A **population** is the total number of possible units or elements that could potentially be included in a study. Although people often think the word "population" refers to everyone in the world, researchers typically define their populations more narrowly. For example, the population for the first research hypothesis in Table 1.1 could be "consumers," "U.S. consumers," "Georgia consumers," or "University of Georgia consumers." Researchers typically try to define their populations as broadly as possibly (e.g., "consumers" rather than "University of Georgia consumers") to maximize the applications or implications of their research.

Learning Check 1

Reviewing What You've Learned So Far



1. Review questions
    a. What are the five steps involved in the research process?
    b. Why is it useful to review and evaluate theories and research before conducting a study?
    c. What are the two main characteristics of research hypotheses?
    d. What is the difference between an independent variable and a dependent variable?
2. The following are several research hypotheses from published studies. For each research hypothesis, identify the independent and dependent variable.
    a. "College students will rate instructors who dress formally (i.e., business suit and tie) as having more expertise than instructors who dress casually (i.e., slacks and shirt)" (Sebastian & Bristow, 2008, p. 197).
    b. "It was expected tha. . . . greater amounts of television viewin. . . . would predict greate. . . . posttraumatic stress symptoms" (McLeish & Del Ben, 2008, p. 417).
    c. "Do borrowers who did not complete college have greater financial distress compared with . . . non-borrowers who did not complete college [or] borrowers who earned a degree?" (Jabbari et al., 2023, p. 6).
    d. "We expected that drivers would believe that texting while driving is more detrimental to their driving performance than browsing social media" (Hashash et al., 2019, p. 70).

It's extremely difficult to collect data from all members of a population. Imagine, for example, the time and money needed to collect information from every consumer in the United States! For this reason, researchers typically draw conclusions about populations based on information collected from a sample of the population. A **sample** is a subset or portion of a

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

population. Table 1.4 describes the samples used in the four studies introduced in Table 1.1 . As you can see from Table 1.4 , samples vary greatly in terms of their targeted population and size (the number of participants). The extent to which the results of a study can be generalized to the target population depends on the extent to which the sample is representative of the population.

Determining How Variables Will Be Measured: Levels of Measurement

The research hypotheses described in Table 1.1 involve variables such as type of appeal, intentions to use cannabis, and gender identity. To conduct a research study, the researcher must decide how to measure the variables stated in the research hypothesis. **Measurement** is the assignment of numbers or categories to objects or events according to rules.

For example, to measure the variable “height,” a researcher might use “number of inches from the ground in bare feet” as a form of measurement. To measure “success in college,” a student’s grade point average (GPA) may be obtained from school transcripts. “Self-esteem” might be measured by having people complete a questionnaire and be categorized as having either “low” or “high” self-esteem. As these examples demonstrate, different types of variables have different forms of measurement. Consequently, researchers have identified four distinct levels for measuring variables: nominal, ordinal, interval, and ratio.

The values of variables measured at the **nominal level of measurement** differ in category or type. The word *nominal* implies having to do with “names,” such that we use first and last names to distinguish between people. Gender identity is an example of a nominal variable, in that it consists of categories or types (such as cis male and cis female) rather than numeric values. In the first research hypothesis in Table 1.1 , the independent variable of “type of appeal” is measured at the nominal level with two categories: FOMO and non-FOMO.

Variables measured at the **ordinal level of measurement** have values that can be placed in an order relative to the other values. Rankings (such as finishing first, second, or third in a race), size (small, medium, large, or extra-large), and ratings (below average, average, or above average) are examples of ordinal variables. The values of ordinal variables indicate that one value represents relatively more or less of the variable than other values, but they don’t specify the precise difference between values. For example, although you can say that a “medium” drink is smaller than a “large” drink which is smaller than an “extra-large” drink, you can’t specify the exact difference in size between the three drinks.

Table 1.4 Research Hypotheses and Study Samples

Research Hypothesis	Study Sample
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.	“We recruited 150 participants from Amazon MTurk” (Friederich et at., 2024, p. 105).
Exposure to pro-cannabis messages on social media is	“Teens (N = 350) . . . needed to be between 13 and 17 years old and from Washington state” (Willoughby et

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Study Sample
related to increased intentions to use cannabis.	al., 2024, p. 185).
Mothers have more accurate knowledge of their children's online activities than do fathers.	"triadic data including reports from children aged 13 to 18, mothers and fathers . . . recruited in Flanders, the northern, Dutch speaking region of Belgium . . . a total of 357 valid triads was achieved" (Symons et al., 2017, p. 405).
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	"undergraduate college students aged 18–30 who were enrolled at 4-year institutions. This yielded a sample of 46,510 students" (Marx et al., 2024, p. 334).

The values of variables measured at the **interval level of measurement** are numbers equally spaced along a continuum. One example of an interval variable is the Fahrenheit scale of temperature. Here, a difference of 5 degrees has the same meaning anywhere along the scale; for example, the difference between 45°F and 50°F is the same as the difference between 65°F and 70°F. Many variables studied in the behavioral sciences (e.g., personality characteristics or attitudes) are considered to be measured at the interval level of measurement.

Variables at the **ratio level of measurement** are identical to interval variables, with one exception: Ratio scales possess what is known as a true zero point, for which the value of zero (0) represents the complete absence of the variable. Variables that describe a physical dimension (such as height, weight, distance, and time duration) typically have a true zero point. In the last research hypothesis in Table 1.1 , the variable "number of times victimized in last 12 months" is a ratio variable because it has a true zero point, where zero would indicate the complete absence of victimizations.

One advantage of ratio measurement versus interval is that ratio variables allow for a greater number of comparisons among values. Consider, for example, the value 6 for the ratio variable "inches." Not only is the difference between 4 and 6 inches the same as the difference between 6 and 8 inches (thereby involving addition and subtraction, which also can be used with interval variables), 6 inches is twice as long as 3 inches and half as long as 12 inches (involving multiplication and division). Multiplication and division comparisons cannot be made with interval variables. You cannot, for example, say that a temperature of 90°F is three times as much temperature as 30°F.

The boldfaced and italicized text in Table 1.5 illustrates how the four studies described in the earlier tables measured their independent and dependent variables. As you can see, the variables in these studies have been measured in a variety of ways at different levels of measurement.

Table 1.5 Research Hypotheses and Measurement of Variables

Research Hypothesis	Independent Variable	Dependent Variable

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Independent Variable	Dependent Variable
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.	Type of appeal **FOMO, non-FOMO (nominal)**	Intention to invest in cryptos **Likelihood of investing** **1-7 scale (1 = Highly unlikely to 7 = Highly likely) (ordinal)**
Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.	Exposure to pro-cannabis messages on social media **How often saw pro-cannabis messages (ratio)**	Intentions to use cannabis **Likelihood of using cannabis in the future (ratio)**
Mothers have more accurate knowledge of their children's online activities than do fathers.	Type of parent **Mother, Father (nominal)**	Accuracy of online activities **Accurate, somewhat accurate, inaccurate (nominal)**
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	Gender identity **Nonbinary, cis man, cis woman** **(nominal)**	Number of times victimized in last 12 months **(ratio)**

What difference does the level of measurement make?

A variable's level of measurement has important implications for researchers in that it influences how research hypotheses are stated and how data are analyzed. For example, imagine you study the variable "success in college." You could measure this variable at the *nominal* level of measurement by classifying students into one of two groups: successful or unsuccessful. Researchers typically use nominal variables when they are interested in studying differences between groups, such as, "Do successful and unsuccessful students differ in their study habits?"

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

You could instead measure success in college as an *ordinal* variable by ranking students from top to bottom, which enables a researcher to study relative differences with less concern for the precise size of these differences. For example, using ranked data, you could ask, “How similar are younger and older faculty members’ rankings of their students?”

To study success in college at the *interval* level, you could rate each student on a scale from 1 (low) to 100 (high). You could then use these ratings to address a research question such as, “Is there a relationship between faculty ratings of their students and students’ ratings of themselves?”

Finally, to measure success in college using the *ratio* level of measurement, you could use students’ salary after graduation (in dollars) as an indicator of success. Salary is a ratio variable because it contains a true zero point. In practice, ratio variables can be used to ask many of the same questions as those involving interval variables. For example, you could ask, “Is there a relationship between faculty members’ ratings of students and students’ salaries after graduation?”

As you see, how researchers measure their variables influences how they state their research questions and hypotheses. The level of measurement has another important implication: It helps determine the statistics researchers use to analyze their data. Certain statistics can be applied to variables measured at the interval or ratio levels of measurement, whereas others are appropriate for nominal or ordinal variables. As we move from chapter to chapter of this book, the level of measurement of the variables being analyzed will be noted.

Selecting a Method to Collect the Data: Experimental and Nonexperimental Research Methods

In addition to drawing a sample from a population and determining how the variables will be measured, the third step in collecting data is to determine the type of research method to use to collect the data on these variables. Research methods may be classified into two main types:

experimental research methods, and

nonexperimental research methods.

Experimental research methods.

Experimental research methods are methods designed to test causal relationships between variables—more specifically, whether changes in independent variables produce or cause changes in dependent variables. To make inferences about cause-effect relationships, researchers conducting an experiment must first eliminate other possible causes or explanations for changes in the dependent variable besides the independent variable. If it can be claimed that a variable other than the independent variable created the changes in the dependent variable, the research results are said to be confounded. A **confounding variable** is a variable related to an independent variable that provides an alternative explanation for the relationship between the independent and dependent variables.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

To illustrate the concept of confounding variables, consider the following question: Is there a relationship between a mother's ethnicity and the birth weight of her baby? Although research has shown differences in the birth weights of babies of different ethnicities, one study noted that some of this research did not eliminate the possibility that these differences may be partly due to differences between ethnic groups on factors associated with low birth weights, such as the mother's average age at the time of pregnancy and behaviors such as smoking and drinking (Ma, 2008). These other factors are considered confounding variables in that they provide alternative explanations for the relationship between ethnicity and babies' birth weights. Confounding variables make it difficult for researchers to draw firm conclusions regarding the relationship between the independent and dependent variable.

Learning Check 2

Reviewing What You've Learned So Far



1. Review questions
    a. What is the difference between a population and a sample?
    b. Within the research process, what is the relationship between populations and samples?
    c. What are the four levels of measurement? How do they differ?
2. For each of the following variables, name the scale of measurement (nominal, ordinal, interval, or ratio).
    a. Type of school attended (public, private)
    b. Probability of graduating college in 4 years (0% to 100%)
    c. Rating of television (good, better, best)
    d. Number of computers in home

Researchers minimize the influence of confounding variables in two main ways. First, researchers exercise experimental control by making the research setting (i.e., characteristics of the research participants, the instruments or measures administered, the methods used to collect data) the same for all participants. In the birth weight study, for example, the research design might control for the effect of parental smoking on birth weight by only including nonsmokers in the study sample, thereby excluding those who smoke.

Another way researchers exert control over the research setting is by including a condition known as a **control group** , a group of participants in an experiment not exposed to the independent variable. For example, if you conducted an experiment examining the effects of caffeine on one's health, one group (the experimental group) might be instructed to drink coffee while a second group (the control group) drinks decaffeinated coffee. By comparing the results of the two groups, researchers can then assess the impact of caffeinated coffee on the health of those who drink it.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Because researchers cannot possibly identify and control for all potential confounding variables, a second strategy used to minimize the influence of confounding variables is **random assignment** , which is assigning participants to each category of an independent variable in such a way that each participant has an equal chance of being assigned to each category. For example, a researcher could randomly assign a participant to the experimental or control condition by flipping a coin: heads for experimental, tails for control. The purpose of random assignment is to equalize or neutralize the effects of confounding variables by distributing them equally over the levels of the independent variable.

Nonexperimental (correlational) research methods.

Experimental research designs are one of the best tools researchers have for making causal inferences. However, the ability to make causal inferences requires a great deal of control that may not always be possible or desirable. For this reason, researchers often employ **nonexperimental research methods** , which are research methods designed to measure relationships between variables without having the ability to infer cause-effect relationships. These research methods are sometimes referred to as *correlational* research methods. It is important to understand that correlational research methods are different from correlational statistics we'll introduce in Chapter 13. Correlational research methods pertain to how data are *collected* ; correlational statistics pertain to how data are *analyzed* . Some of the most common types of nonexperimental research designs include quasi-experiments, survey research, observational research, and archival research.

Quasi-experimental research compares naturally formed or preexisting groups rather than randomly assigning participants to conditions. One example of this type of research was a study examining the effects of the COVID-19 pandemic on three groups of students: those living on campus, those living off campus but within commuting distance to school, and those living far away in another city, state, or country (Bisconer & McGill, 2023). This study is a quasi-experiment because the students were not randomly assigned to one of the three living conditions before collecting data.

Survey research methods obtain information directly from a group of people regarding their opinions, beliefs, or behavior. The goal of survey research, which can include using questionnaires and interviews, is to obtain information from a sample to represent or estimate the views of a larger population. One example of survey research examined college students' mental health by having them fill out a survey in which they reported their spirituality (the desire for truth, meaning, and significance of one's life) and their self-acceptance (knowledge and acceptance of one's self and limitations) (Nonis et al., 2024). Survey research is nonexperimental in that the variables being measured are not manipulated by the researcher.

Observational research is the systematic and objective observation of naturally occurring behavior or events. Observational research is often used to study phenomena that the researcher cannot or should not deliberately manipulate. For example, researchers studying interpersonal conflict among children would never force children to fight each other. One study resolved this issue by videotaping groups of adolescents discussing topics such as competition and peer pressure and then counting the number of times the adolescents teased or mocked each other (Connolly et al., 2015).

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Rather than observing or measuring behavior directly, **archival research** is the use of archives (records or documents of the activities of individuals, groups, or organizations) to examine research questions or hypotheses. One example of an archival research study studied the question of whether men are perceived to be more creative than women (Proudfoot et al., 2015). In this study, the 100 most viewed talks on TED.com were identified, and the researchers compared viewers' posted ratings of the talks given by men versus women.

Labeling variables in nonexperimental research.

One issue that often confuses students when learning about nonexperimental research is the practice of referring to the variables as the "independent variable" and the "dependent variable." For example, in the study described previously that involved biological sex and TED.com ratings, biological sex may be called the "independent variable." However, because biological sex isn't manipulated by the researcher, it would be inappropriate to say that biological sex "causes" differences in TED.com ratings. To avoid confusion, researchers sometimes refer to the independent variable and the dependent variable in a nonexperimental study as the **predictor variable** and **criterion variable**, respectively.

Analyzing the Data

Once data collection has been completed, the next step in the research process involves data analysis. This part of the research process addresses the primary topic of this book: statistics. Because this is the first chapter of this book, we won't describe analyzing data in detail. Instead, we introduce the notion that there are two main purposes of analyzing data: (1) to organize, summarize, and describe data that has been collected and (2) to test and draw conclusions about ideas and hypotheses. These two purposes are met by calculating two main types of statistics: descriptive statistics and inferential statistics.

Calculating Descriptive Statistics

Descriptive statistics are statistics used to summarize and describe a set of data for a variable. For example, you've probably heard of statistics used to describe or summarize aspects of our society, such as crime statistics and unemployment statistics. Using a research-related example, Caitlin Abar, a researcher at Brown University, conducted a study examining increases in students' alcohol-related behaviors after entering college (Abar, 2012). To measure students' level of alcohol use, she created a variable called "typical weekend drinking," which "was measured as the sum of drinks consumed on a typical Friday and Saturday within the past 30 days" (p. 22). One way to summarize students' responses to the typical weekend drinking variable (0, 1, 2, etc.) would be to calculate the mean, which is the mathematical average of a set of scores. The mean, described in Chapter 3, is an example of a descriptive statistic. The first part of this book describes a variety of descriptive statistics used by researchers to summarize data they've collected.

Calculating Inferential Statistics

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Besides summarizing and describing data, a second purpose of statistics is to test hypotheses. **Inferential statistics** are statistical procedures used to test hypotheses and draw conclusions from data collected during research studies. Using inferential statistics, researchers make inferences about what's believed to exist in populations based on data gathered from samples of the population.

As an example of an inferential statistic, let's return to the example of students' alcohol-related behaviors (Abar, 2012). This study was interested in seeing whether these behaviors were related to aspects of these students' relationships with their parents. One aspect was "alcohol communications," defined as "how often they discussed alcohol related topics with their parents" (p. 22). To test a hypothesis regarding the relationship between students' alcohol use and their alcohol communications, we could use an inferential statistic known as the Pearson correlation coefficient, which will be discussed in Chapter 13. The chapters in the last half of this book discuss a broad variety of inferential statistics, along with detailed instructions for analyzing and drawing conclusions from statistical data and presenting research findings.

Drawing a Conclusion Regarding the Research Hypothesis

Once statistical analyses have been completed, the next step is to interpret the results of the analyses as they relate to the study's research hypothesis. More specifically, do the findings of the analyses support or not support the research hypothesis? The word *support* in the previous sentence is very important. Students and researchers are sometimes tempted to say their findings "prove" their hypotheses are either "true" or "false." However, as we'll discuss in Chapter 5, because researchers typically don't collect data from the entire population, they can't know with 100% certainty whether their hypothesis has been proven to be true or false. A *huge* part of this book is learning how to draw appropriate conclusions from statistical analyses.

Communicating the Findings of the Study

Conducting research requires a variety of different skills: conceptual skills to develop research hypotheses, methodological skills to collect data, and mathematical skills to analyze these data. Another integral part of the research process is the communication skills needed to inform others about a study. Researchers must be not only skilled scientists but also effective writers.

Learning Check 3

Reviewing What You've Learned So Far



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1. Review questions
    a. What are the main differences between experimental and nonexperimental research methods?
    b. What are confounding variables? What do researchers do to minimize the effects of confounding variables?
    c. What are the main types of nonexperimental research methods?
    d. What are the relative strengths and weaknesses of experimental and nonexperimental research?
    e. What are the main purposes of descriptive and inferential statistics?

For many years, the American Psychological Association (APA), the professional association for psychologists, has provided researchers guidance on how to communicate the results of their research. In 1929, the APA published a seven-page article entitled, “Instructions in Regard to Preparation of Manuscript.” In contrast, the seventh edition of the *Publication Manual of the American Psychological Association* , published in 2020, consists of 428 pages! This increase in length highlights the challenge of communicating research effectively and appropriately.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1.4 Plan of the Book

The primary purpose of this chapter was to introduce statistics and place it within the larger research process. The remainder of this book discusses different statistics used by researchers to examine various questions of interest. Although it's critical for you to learn to correctly calculate statistics, it is equally important to understand the role of statistics within the research process and appreciate issues related to the use (and sometimes misuse) of statistics.

The first half of the book (Chapters 2 through 6) introduces conceptual and mathematical issues that are the foundation of statistical analyses. Chapters 2, 3, and 4 discuss how data may be examined, described, summarized, and presented in numeric, visual, and graphic form. Descriptive statistics are introduced and discussed in these chapters. Chapters 5 and 6 introduce characteristics of the inferential statistics used to test research hypotheses. The chapters in the second half of the book cover a number of different inferential statistics used with different types of variables.

In most cases, the chapters in this book share a uniform format and structure and are centered on the research process, using examples of actual published research studies that address questions with which you may be familiar. For each example, we'll clearly state the research hypothesis, briefly describe how the data in the study were collected, illustrate the calculation and interpretation of descriptive and inferential statistics, discuss the extent to which the statistical analyses support the research hypothesis, and show how to communicate one's research to a broader audience.

1.5 Looking Ahead

We began this chapter by defining statistics and providing reasons why learning about them may be of value to you. Next, we placed statistics within the stages of the larger research process, a process that will be the foundation of this textbook. As we've explained, the research process is centered on a research hypothesis, which is a predicted relationship between variables. Once a hypothesis has been stated, the next step is to collect data to be analyzed. However, before conducting statistical analyses on a set of data, there is preliminary work that must be completed. The next chapter will focus on methods used by researchers to examine data.

1.6 Summary

Statistics may be defined as a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data. As such, statistics not only is concerned about how data are analyzed but also recognizes the importance of understanding how data are collected and how the results of analyses are interpreted and communicated. There are a variety of reasons why students should learn statistics: Statistics are encountered in a wide variety of daily activities, students may be asked or required to read and interpret the results of statistical analyses in their courses, students may use statistics in conducting their own research, and statistics may help one's career.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Researchers conduct research using the **scientific method** of inquiry, a method of investigation that uses the objective and systematic collection and analysis of empirical data to test theories and hypotheses. The research process used within the scientific method of inquiry consists of five main steps: developing a research hypothesis to be tested, collecting data, analyzing the data, drawing a conclusion regarding the research hypothesis, and communicating the findings to the study.

A **research hypothesis** is a statement regarding an expected or predicted relationship between variables. Developing a research hypothesis involves identifying a question or issue to be examined, reviewing and evaluating relevant theories and research, and stating the research hypothesis to be tested in the study. A **theory** is a set of propositions used to describe or explain a phenomenon. A research hypothesis contains **variables** , which are properties or characteristics of some object, event, or person that can take on different values. More specifically, a research hypothesis states the nature and direction of a proposed relationship between an **independent variable** (a variable manipulated or controlled by the researcher) and a **dependent variable** (a variable measured by the researcher that is expected to change or vary as a function of the independent variable).

Collecting data involves drawing a sample from a population, determining how the variables will be measured, and selecting a method to collect the data. A **population** is the total number of possible units or elements that could potentially be included in a study; a **sample** is a subset or portion of a population. Variables can be measured at one of four levels of measurement: **nominal** (values differing in category or type), **ordinal** (values placed in an order relative to the other values), **interval** (values equally spaced along a numeric continuum), or **ratio** (values equally spaced along a numeric continuum with a true zero point).

There are two main types of methods used to collect data: **experimental research methods** , designed to test whether changes in independent variables produce or cause changes in dependent variables, and **nonexperimental (correlational) research methods** , designed to examine the relationship between variables without the ability to infer cause-effect relationships.

In experimental research, researchers are concerned about possible **confounding variables** , which are variables related to independent variables that provide an alternative explanation for the relationship between independent and dependent variables. To minimize the impact of confounding variables, a research study may exercise experimental control over various aspects of the situation, including the use of a **control group** , which is a group of participants not exposed to the independent variable the research is designed to test, or **random assignment** , which involves assigning participants to each category of an independent variable in such a way that each participant has an equal chance of being assigned to each category.

Examples of nonexperimental research methods are **survey research** , in which information is directly obtained from a group of people regarding their opinions, beliefs, or behavior; **observational research** , which involves the systematic and objective observation of naturally occurring events; and **archival research** , which uses archives (records or documents of the activities of individuals, groups, or organizations) to examine research questions or hypotheses. In nonexperimental research, the independent variable and

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

dependent variable may be referred to as the **predictor variable** and **criterion variable**, respectively.

Once data have been collected, the next step is to analyze them using two main types of statistics: **descriptive statistics**, which summarize and describe a set of data, and **inferential statistics**, which are statistical techniques used to test hypotheses and draw conclusions from data collected during research studies.

Once the statistical analyses have been completed, the results of the analyses are interpreted regarding whether they support or not support the study's research hypothesis. The final step in the research process is to communicate the study to a broader audience.

1.7 Important Terms

archival research

confounding variable

control group

criterion variable

dependent variable

descriptive statistics

experimental research methods

independent variable

inferential statistics

interval level of measurement

level of measurement (nominal, ordinal, interval, ratio)

measurement

nonexperimental research methods

observational research

population

predictor variable

quasi-experimental research

random assignment

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

research hypothesis

sample

scientific method

statistics

survey research

theory

variable

1.8 Exercises

1. Identify the independent and dependent variable for each of the following research questions and hypotheses from published articles.
    a. “We hypothesize that finance majors will embody narcissistic traits more than other business majors” (Wood et al., 2021, p. 924).
    b. “Men are expected to employ physical characteristics (particularly those that are directly related to sex) more often than women in selecting dating candidates” (Hetsroni, 2000, p. 91).
    c. “It is hypothesized that perceptions of emotional pain and support needs will be higher for the community of the Black victim than those of the officers’” (Kahn et al., 2024).
    d. “Gen Z students are more likely to believe that ‘Group projects bring out the worst in people’ than Millennials” (Schlee et al., 2020, p. 145).
    e. “We hypothesized to find . . . higher stress and anxiety post COVID in the Portuguese group in contrast with the British ” (Penalba-Sanchez et al., 2022, p. 322).
    f. “We hypothesized that consumers would think that a ‘sale price’ presentation would generate a greater monetary savings than an ‘everyday low price’ presentation” (Tom & Ruiz, 1997, p. 403).
2. The following are a number of research questions and hypotheses from actual published articles. For each hypothesis, identify the independent and dependent variable.
    a. “The use of color in a Yellow Pages advertisement will increase the perception of quality of the products for a particular business when compared with noncolor advertisements” (Lohse & Rosen, 2001, p. 75).
    b. “We hypothesized that parents who use more frequent corporal and verbal punishmen. . . . will report more problem behaviors in their children” (Brenner & Fox, 1998, p. 252).
    c. “It was hypothesized that adolescents with anorexia nervosa woul. . . . be more respectful when compared with peers with bulimia nervosa” (Pryor & Wiederman, 1998, p. 292).
    d. “The purpose of the present research was to assess brand name recognition as a function of humor in advertisements. . . . It was predicted that participants would recognize product brand names that had been presented with humorous

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

advertisements more often than brand names presented with nonhumorous advertisements" (Berg & Lippman, 2001, p. 197).

e. "The purpose of this study was to explore the relation between time spent in daycare and the quality of exploratory behaviors in 9-month-old infant. . . . it was hypothesized tha. . . . infants who spent greater amounts of time in center-based care would demonstrate more advanced exploratory behaviors than infants who did not spend as much time in center-based care" (Schuetze et al., 1999, p. 269).

f. "It was hypothesized that students would score higher on test items in which the narrative contains topics and elements that resonate with their daily experiences versus test items comprised of material that is unfamiliar" (Erdodi, 2012, p. 172).

3. Identify the independent and dependent variable for each of the following research questions and hypotheses from published articles.
    a. "It was hypothesized that the number of selfies posted . . . would be positively correlated with dimensions of narcissism" (Barry et al., 2017, p. 51).
    b. "We hypothesized that preschoolers' self-reported preferences for aggressive media . . . would be positively related to parent reports of children's aggressive behaviors" (Jamnik & DiLalla, 2018, p. 159).
    c. "We hypothesized that helicopter parenting would lead to higher levels of depression" (Darlow et al., 2017, p. 3).
    d. "We predicted that weight-related shame would be associated with greater binge eating" (Craven & Fekete, 2019, p. 45).
    e. "We hypothesize that men with strong gender identities will be more sensitive to manipulated threats to their masculinity compared with men with gender identities that are less central to their identity" (Harrison & Michelson, 2019, p. 65).
    f. "After exposure to fake news story with user comments identifying the news as fake, people are less likely to share the fake news than after exposure to the fake news story with user comments supportive of the content" (Colliander, 2019, p. 204).
4. Name the scale of measurement (nominal, ordinal, interval, ratio) for each of the following variables:
    a. The amount of time needed to react to a sound
    b. Gender
    c. Score on the Scholastic Aptitude Test (SAT)
    d. Political orientation (not at all conservative, conservative, very conservative)
    e. Political affiliation (Democrat, Republican, Independent)
5. Name the scale of measurement (nominal, ordinal, interval, ratio) for each of the following variables:
    a. One's age (in years)
    b. Size of soft drink (small, medium, large, extra large)
    c. Voting behavior (in favor vs. against)
    d. IQ score
    e. Parent (mother vs. father)
6. A faculty member wishes to assess the relationship between students' scores on the SAT and their performance in college.
    a. What is a possible research hypothesis in this situation?
    b. What are the independent and dependent variables?
    c. How could you measure the variable "performance in college" at each of the four levels of measurement?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

7. A researcher hypothesizes that drivers who use cellular phones will get into a greater number of traffic accidents than drivers who do not use these phones. For this example,
    a. What are the independent and dependent variables?
    b. How could you measure the dependent variable?
    c. How could you conduct this study using an experimental research method? How could you conduct this study using a nonexperimental method?

Answers to Learning Checks

Learning Check 1

1.

    a. IV: Instructor's dress; DV: Expertise

    b. IV: Amount of television viewing; DV: Posttraumatic stress symptoms

    c. IV: Student group; DV: Financial distress

    d. IV: Activity while driving; DV: Driving performance

Learning Check 2

2.

    a. Nominal

    b. Ratio

    c. Ordinal

    d. Ratio

Answers to Odd-Numbered Exercises

1.

    a. IV: Type of business major; DV: Narcissistic traits

    b. IV: Biological sex; DV: Employment of physical characteristics

    c. IV: Community; DV: Perceptions of emotional pain and support needs

    d. IV: Generation; DV: Beliefs about group projects

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

e. IV: Country; DV: Stress and anxiety

f. IV: Price presentation; DV: Perception of savings

3.

a. IV: Number of selfies posted; DV: Narcissism

b. IV: Preference for aggressive media; DV: Parent reports of children's aggressive behaviors

c. IV: Helicopter parenting; DV: Depression

d. IV: Weight-related shame; DV: Binge eating

e. IV: Strength of gender identity; DV: Sensitivity to threat to masculinity

f. IV: Type of user comments to news story; DV: Sharing of news story

5.

a. Ratio

b. Ordinal

c. Nominal

d. Interval

e. Nominal

7.

a. IV: Cellular phone use; DV: Number of traffic accidents

b. Example: Determine the number of traffic accidents each person has experienced in the past year.

c. Example of experimental: Use a driving simulation with the "driver" talking on a phone and measure the number of potential driving mistakes or accidents.

d. Example of nonexperimental: Take a survey of the number of accidents people have been in and whether or not they were using a phone during the accident.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CHAPTER 2 EXAMINING DATA: TABLES AND FIGURES

Learning Objectives

2.1 Describe the main purpose and the research hypothesis of the “lottery” study.

2.2 Describe different reasons for examining data.

2.3 Describe the steps used to create a frequency distribution table.

2.4 Describe the steps used to create a grouped frequency distribution table.

2.5 Explain the differences between bar charts, pie charts, histograms, and frequency polygons.

2.6 Describe the three aspects and types of distributions of data.

Chapter 1 described a process by which research may be conducted, which begins by stating a research hypothesis. The next step involves collecting data that will ultimately be used in statistical analyses designed to test the research hypothesis. However, researchers don’t conduct analyses until they’ve first examined the data to make sure appropriate conclusions can be drawn from the analyses. This chapter describes *why* researchers examine data and *how* data may be examined using tables, charts, and graphs. As with many of the chapters in this book, this discussion will be guided by the findings from a published research study.

2.1 An Example From the Research: Winning the Lottery

A friend of yours flips a coin and asks you to guess whether it’s “heads” or “tails.” You guess “tails,” but the answer is “heads.” Your friend flips the coin a second time, and again you predict “tails.” But again the answer is “heads.” On the third coin flip, you guess “tails” once more—but, once again, you’re wrong. As your friend prepares to flip the coin a fourth time, you tell yourself there’s little chance the coin could be “heads” again. But the reality is that the fourth coin flip is equally likely to be “heads” or “tails” because the outcome of the first three coin flips has absolutely no impact on the fourth flip. In other words, the outcome of a coin flip is a random event that can’t be controlled or predetermined. But do you really believe this?

The impact of random events can take on greater significance than the outcome of a coin flip. For example, government-run lotteries provide the opportunity to win huge amounts of money through the random selection of numbers. In 2001, Dr. Karen Hardoon, a researcher at McGill University in Canada, studied the thought processes people use when purchasing lottery tickets (Hardoon et al., 2001). As previous research had found gamblers believe they can control random events such as the rolling of dice or the spinning of a roulette wheel, Dr. Hardoon and her colleagues were interested in

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

determining “whether individuals with gambling problems perceive the purchase of lottery tickets in a similar manner as non-problem gamblers” (p. 752). On the basis of earlier findings, Dr. Hardoon hypothesized that problem gamblers are less likely to believe lottery outcomes are random than are nonproblem gamblers.

To test their research hypothesis, Dr. Hardoon and her associates located a group of gamblers and asked whether they had ever experienced problems as a result of their gambling. Each participant was classified into one of two groups: problem gamblers and nonproblem gamblers.

The participants in the study, which we’ll refer to as the lottery study, were shown four lottery tickets and asked, “If you were to buy a ticket to play in the lottery, which one would you select?” Each ticket contained six numbers selected from among the numbers 1 through 49; the four tickets were given a particular label. The *sequence* ticket had numbers in consecutive order (30-31-32-33-34-35); numbers on the *pattern* ticket increased by 5s (5-10-15-20-25-30); the *nonequilibrated* numbers (35-37-40-43-44-49) were clustered at the upper end of the 49 numbers, and the *random* ticket (7-8-23-34-36-42) shared none of the characteristics of the other three tickets. We’ll call the variable in this study “lottery ticket,” a variable with four categories: sequence, pattern, nonequilibrated, and random.

The researchers recorded which of the four tickets each participant chose. Table 2.1 lists the data for the lottery ticket variable for the nonproblem gamblers (the data for the problem gamblers are provided in the exercises at the end of this chapter). Looking at this table, we see that the nonproblem gamblers differed in their lottery ticket choices; however, it’s difficult to describe these differences given the unorganized nature of the collected data. Consequently, before conducting statistical analyses on their data, researchers typically organize and examine them. In the following sections, we will discuss why and how data are organized and examined.

2.2 Why Examine Data?

There are a variety of reasons why researchers examine data they’ve collected before analyzing the data. These reasons include the following:

To gain an initial sense of the data

To identify outliers

To evaluate research methodology

To determine whether data meet statistical criteria and assumptions

Each of these reasons is discussed in the following sections.

Table 2.1 The Lottery Ticket Choices of 22 Nonproblem Gamblers

Participant	Lottery Ticket
1	**Random**
2	**Random**
3	**Nonequilibrated**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Participant	Lottery Ticket
4	**Pattern**
5	**Random**
6	**Sequence**
7	**Random**
8	**Random**
9	**Pattern**
10	**Pattern**
11	**Random**
12	**Pattern**
13	**Random**
14	**Nonequilibrated**
15	**Random**
16	**Pattern**
17	**Nonequilibrated**
18	**Random**
19	**Random**
20	**Pattern**
21	**Random**
22	**Random**

Gaining an Initial Sense of the Data

Because researchers spend a great amount of time and energy designing their research studies, they're eager to analyze the data they've collected. Students taking courses in statistics, on the other hand, face a different challenge: They're given sets of data with which they have no prior experience. Although both researchers and students may be motivated to analyze their data as quickly as possible, it's important to first examine the data in order to gain an initial sense of them.

As an example, imagine you conduct a study regarding the fuel efficiency of automobiles and collect data on the miles per gallon (MPG) of different types of cars. You might expect the scores in your data to range from 10 to 50 MPG, with the majority between 25 and 35 MPG. Examining data helps researchers gain an initial sense of whether the data are in line with their expectations. Examining data also provides a way of confirming the results of later statistical analyses. We've found that students' errors in their calculations could have been avoided had they first looked at the set of data. For example, examining data will help you avoid making statements such as, "The average MPG in my sample was 127 MPG."

Identifying Outliers

Researchers also examine their data to identify **outliers** , defined as rare, extreme scores that lie outside of the range of the majority of scores in a set of data. Imagine, for example, that a group of 15 students is asked to take a 20-item test of mathematical calculations; each student's score is the number of test items answered correctly. In scoring the tests, a researcher finds that 14 students

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

correctly answered between 8 and 17 questions. However, one student provided the wrong answer to every question for a score of 0. In this case, the score of 0 would be considered an outlier.

If not identified, outliers may distort conclusions researchers draw about their sample and their data, particularly when the sample is relatively small. In the previous example, the single score of 0 might lead an observer to conclude the class as a whole performed more poorly on the test than they actually did.

Evaluating Research Methodology

Examining data is also useful in assessing the effectiveness of the methods used to collect the data. For the 20-item math test, for example, it might be reasonable to expect the majority of scores to fall in the middle of the 0 to 20 range. But what if every student answered at least 17 of the 20 questions correctly? You might conclude the test was too easy and that more difficult questions are needed to accurately assess students' math skills. By comparing the collected data for a variable with the range of possible values, researchers are able to evaluate and modify their measurement tools.

Determining Whether Data Meet Statistical Criteria and Assumptions

Researchers also examine their data to see whether they meet certain statistical criteria and assumptions. Many of the statistics in this book are based on assumptions regarding the shape and nature of a set of data that has been collected. For example, these procedures assume that in most data sets, the majority of scores will fall in the middle of the range of possible values, with a small amount of scores at the highest and lowest possible values (you may have heard this referred to as "the bell curve"). In the example of the 20-item math test, it might be assumed that most scores will be between 7 and 13, with smaller number of scores less than 7 or greater than 13. When assumptions such as this are not met, it's possible to draw inappropriate conclusions from statistical analyses that are conducted.

2.3 Examining Data Using Tables

One part of examining data that have been collected for a variable involves answering a series of questions:

> What are the values of the variable?
>
> How many participants in the sample have each value of the variable?
>
> What percentage of the sample has each value of the variable?

As it may be difficult to answer questions such as these just by looking at the data in Table 2.1 , researchers typically organize their data by creating a table known as a **frequency distribution table** , a table that summarizes the number and percentage of participants for the different values of a variable. This section discusses the steps involved in creating frequency distribution tables using these questions.

What Are the Values of the Variable?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The first step in organizing the data for a variable is to identify all of the possible values for the variable. Table 2.2 begins the construction of a frequency distribution table by identifying the four values for the lottery ticket variable: sequence, pattern, nonequilibrated, and random.

How Many Participants in the Sample Have Each Value of the Variable?

Once the values of the variable have been identified, the next step in organizing a set of data is to determine the **frequency** of each value, which is the number of participants in a sample corresponding to a value of a variable. These frequencies help indicate the nature and shape of the distribution of scores for a variable in a sample.

Table 2.2 Identifying the Values of the Lottery Ticket Variable

Number	Lottery Ticket
1	Sequence
2	Pattern
3	Nonequilibrated
4	Random

Table 2.3 Using the Tick Mark Method to Determine the Frequencies of the Lottery Ticket Variable

Lottery Ticket	Tick Marks	*f*
Sequence	|	1
Pattern	||||— |	6
Nonequilibrated	|||	3
Random	||||— ||||— ||	12
Total	||||— ||||— ||||—||||—||	22

One way to determine the frequencies for a variable is to sort the data into the different values and then count the number of participants having each value using a “tick mark method,” where a tick mark is placed next to a value of the variable each time that value appears in the sample. As shown in Table 2.3 , the tick marks are then counted to determine the frequency for each value. For example, the number of tick marks and therefore the frequency (f) for the sequence ticket is 1 because it was picked by one participant (Participant 6 in Table 2.1). Similarly, the tick marks for each of the other three types of tickets can be recorded and counted until all of the data in the sample are accounted for.

What Percentage of the Sample Has Each Value of the Variable?

Although it’s important to determine the frequency for each value of a variable, these frequencies by themselves are not always particularly informative. For example, how you interpret “There are 35 Democrats in my sample” is different when the sample consists of 40 voters rather than 400 voters.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Because the interpretation of frequencies depends on the size of the sample, it's useful to calculate the *percentage* of the sample having each value of a variable. These percentages can be calculated using the following formula:

$$\% = \frac{f}{total\,number\,of\,scores} * 100$$

where *f* is the frequency for a value of a variable.

Looking at Table 2.3 , 1 of the 22 nonproblem gamblers chose the sequence ticket. Therefore, the percentage of the sample choosing this ticket is

$$\begin{aligned} \% &= \frac{f}{total\,number\,of\,scores} * 100 \\ &= \frac{1}{22} * 100 = .05 * 100 \\ &= 5\% \end{aligned}$$

After calculating the percentage for each of the four tickets, Table 2.4 provides the frequency distribution table for the lottery ticket variable.

Once a frequency distribution table has been constructed for a variable, researchers can begin to draw conclusions about their sample. For the lottery ticket variable, the random ticket had the highest frequency and percentage (*f* = 12, 54%) compared to the pattern (*f* = 6, 27%), nonequilibrated (*f* = 3, 14%), and sequence (*f* = 1, 5%) tickets. As a result, the researchers in the lottery study made the following observation:

"The results of the present study indicated that, for the entire sample . . . the greatest percentage of tickets selected by the entire sampl. . . . were random tickets." (Hardoon et al., 2001, p. 760)

Table 2.4 Frequency Distribution Table for the Lottery Ticket Variable

Lottery Ticket	***f***	**%**
Sequence	1	**5%**
Pattern	6	**27%**
Nonequilibrated	3	**14%**
Random	12	**54%**
Total	22	**100%**

Learning Check 1

Reviewing What You've Learned So Far

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



1. Review questions
    a. What are some reasons for examining data before conducting statistical analyses?
    b. Why are outliers of concern to researchers?
    c. What questions do you address about a set of data for a variable in constructing frequency distribution tables?
    d. Why is it useful to calculate percentages in addition to frequencies for each value of a variable?
2. For each of the following situations, create a frequency distribution table.
    a. A college counselor asks a group of 116 seniors what their plans are after graduating from college; 71 say they are going to “work,” 34 are planning on going to “graduate school,” and 11 say they are “not sure” what their plans are.
    b. A high school instructor teaching a course finds that of the students she teaches, 69 are Freshmen, 18 are Sophomores, 12 are Juniors, and 9 are Seniors.
    c. A pollster stops 12 people and asks if they are either “in favor” or “against” a local proposition in Table 2.5 :

Table 2.5 Example of People’s Position

Person	Position
1	**In favor**
2	**In favor**
3	**Against**
4	**In favor**
5	**Against**
6	**Against**
7	**In favor**
8	**Against**
9	**In favor**
10	**Against**
11	**In favor**
12	**In favor**

    d. The company that makes M&Ms candy (www.mms.com) conducted a survey asking people which new color they would like to have added: purple, aqua, or pink. Table 2.6 are the votes of a hypothetical sample:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Table 2.6
Hypothetical Sample of New M&M Color

Person	Color
1	**Pink**
2	**Purple**
3	**Pink**
4	**Aqua**
5	**Purple**
6	**Purple**
7	**Purple**
8	**Aqua**
9	**Purple**
10	**Pink**
11	**Aqua**
12	**Pink**
13	**Pink**
14	**Purple**
15	**Aqua**
16	**Pink**
17	**Purple**
18	**Purple**
19	**Aqua**
20	**Purple**
21	**Purple**
22	**Pink**
23	**Pink**
24	**Purple**

2.4 Grouped Frequency Distribution Tables

The frequency distribution table in Table 2.4 was sufficient for presenting the frequencies for the lottery ticket variable. However, other situations may involve variables with a large number of possible values. Consider, for example, the variable grade point average (GPA). The values for GPA typically range from 0.00 to 4.00, with hundreds of possible values in between. Creating a frequency distribution table for GPA could result in an extremely large table, with many values having frequencies of zero.

To examine a variable with a large number of values, it is useful to construct a **grouped frequency distribution table** , a table that groups the values of a variable measured at the interval or ratio level

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

of measurement into a small number of intervals and then provides the frequency and percentage for each interval. To illustrate how grouped frequency distribution tables are constructed, Table 2.7 (a) provides a sample of 42 hypothetical GPAs. In Table 2.7 (b), each of these 42 GPAs has been placed into one of eight intervals that comprise a grouped frequency distribution table.

To illustrate how to read and interpret a grouped frequency distribution table, the lowest interval in Table 2.7 (b) consists of students with GPAs less than (<) 2.00. The frequency in this interval (*f* = 1) consists of one student: Student 29 (GPA = 1.96). The next interval consists of students with GPAs ranging from 2.00 to 2.29 and contains three students (4, 22, and 38). This grouping continues until all of the GPAs in the sample have been accounted for.

To create a grouped frequency distribution table, the real limits of each interval should be identified. The **real limits** of an interval are the values of the variable that are halfway between the top of one interval and the bottom of the next interval. For example, because of how values of GPA are rounded, the lowest value of GPA that would lead someone to be placed in the interval 2.30–2.59 is 2.295, which is halfway between the top of the 2.00–2.29 interval and the bottom of the 2.30–2.59 interval. The **real lower limit** is the smallest value of a variable that would be grouped into a particular interval. In this case, 2.295 represents the real lower limit for the 2.30–2.59 interval. On the other hand, the **real upper limit** is the largest value of a variable that would be grouped into a particular interval. For the 2.30–2.59 interval, the real upper limit is 2.594.

Once the grouped frequency table for the GPA example has been created (Table 2.7(b)), what preliminary conclusions might be drawn about these students? First, because the 2.90–3.19 interval has the greatest frequency (*f* = 10), you might conclude that the typical student in this sample has a GPA close to 3.00. Second, moving in both directions away from the 2.90–3.19 interval, the frequencies of the different intervals grow progressively smaller, with fewer and fewer students having either very high or very low GPAs. Table 2.8 provides a list of guidelines for creating grouped frequency distribution tables.

Cumulative Percentages

In addition to determining the percentage of the sample that has each individual value or interval for a variable, it is sometimes useful to combine these percentages. For the GPA example, you may be interested in knowing, “What percentage of the sample had GPAs less than 2.60?” To answer this question, the percentages in the “%” column of the frequency distribution table must be combined. This type of percentage is referred to as a **cumulative percentage** , defined as the percentage of a sample at or below a particular value of a variable.

Table 2.7a Grade Point Average (GPA), 42 Students: Individual GPAs

Student	GPA
1	**3.05**
2	**2.83**
3	**3.26**
4	**2.19**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Student	GPA
5	**3.52**
6	**3.34**
7	**3.02**
8	**2.97**
9	**3.71**
10	**2.50**
11	**2.77**
12	**3.10**
13	**3.06**
14	**3.83**
15	**2.39**
16	**3.16**
17	**3.92**
18	**2.37**
19	**3.65**
20	**3.70**
21	**3.00**
22	**2.26**
23	**2.65**
24	**3.40**
25	**2.87**
26	**2.92**
27	**2.89**
28	**3.41**
29	**1.96**
30	**3.77**
31	**3.20**
32	**2.86**
33	**2.94**
34	**2.71**
35	**3.43**
36	**3.95**
37	**3.32**
38	**2.28**
39	**3.08**
40	**3.25**
41	**2.40**
42	**3.22**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Table 2.7b Grade Point Average (GPA), 42 Students: Grouped Frequency Distribution Table

GPA	*f*	%
3.80+	3	7%
3.50−3.79	5	12%
3.20−3.49	9	21%
2.90−3.19	10	24%
2.60−2.89	7	17%
2.30−2.59	4	10%
2.00−2.29	3	7%
<2.00	1	2%
Total	42	100%

Table 2.8 Guidelines for Creating Grouped Frequency Distribution Tables

Guideline	Guideline Description
1. *Variables are grouped into approximately 10 intervals.*	The exact number of intervals depends on the data in each sample. For example, data that fall within a small range of values may be accurately represented with a relatively small number of intervals, whereas data extending across a wide range may require a larger number of intervals.
2. *The number of intervals should accurately represent the data.*	The intervals should represent the nature of the data as accurately as possible. For example, there should not be so large a number of intervals that many intervals have frequencies of zero (e.g., 3.00−3.05) or so few intervals that a large majority of the sample falls into only one or two intervals (e.g., 3.00−3.99).
3. *Intervals should be of equal size.*	To make the intervals comparable, they should be of equal size or width. For example, you would not want one interval of 2.01−2.50 (a width of .50) and another of 2.51−2.70 (a width of .20). One exception to this guideline is the lowest or highest interval (e.g., less than 2.00 or greater than 3.70), which is typically wider than the other intervals because it contains a relatively small frequency.
4. *Intervals should not overlap.*	The intervals should be distinct from each other, such that each score may be placed in only one interval. For example, intervals of 2.30− *2.60* and *2.60*− 2.90 are a problem because the GPA 2.60 is in both intervals leading to confusion regarding within which interval to place someone with a GPA of 2.60.

Table 2.9 includes the cumulative percentages for the GPA example. As an example of a cumulative percentage, the cumulative percentage for the 2.30–2.59 interval (19%) has been calculated by combining the percentages for three intervals: < 2.00, 2.00–2.29, and 2.30–2.59: 2% + 7% + 10% = 19%. As a result, we could conclude that 19% of this sample had GPAs less than 2.60. The accumulating of percentages continues until all (100%) of the scores in the sample are combined. As

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

shown in Table 2.9 , cumulative percentages are typically located in the final column of a frequency distribution table and labeled “Cum %.”

Table 2.9 Frequency Distribution Table With Cumulative Percentages for the GPA Variable

GPA	*f*	%	Cum %
3.80+	3	7%	**100%**
3.50−3.79	5	12%	**93%**
3.20−3.49	9	21%	**81%**
2.90−3.19	10	24%	**60%**
2.60−2.89	7	17%	**36%**
2.30−2.59	4	10%	**19%**
2.00−2.29	3	7%	**9%**
< 2.00	1	2%	**2%**
Total	42	100%	

2.5 Examining Data Using Figures

In addition to organizing data into frequency distribution tables, researchers also examine data visually using figures such as charts or graphs. There are four types of figures often used to visually portray data for variables: bar charts, pie charts, histograms, and frequency polygons. Which type of figure a researcher may use depends on the variable’s level of measurement (nominal, ordinal, interval, ratio). Data for variables measured at the nominal or ordinal level of measurement are typically displayed using bar charts or pie charts. The values of interval and ratio variables are graphed using histograms and frequency polygons. Each of these types of visual illustration is discussed in the following sections.

Displaying Nominal and Ordinal Variables: Bar Charts and Pie Charts

The values of variables measured at the *nominal* level of measurement differ in category or type; one example of a nominal variable is the lottery ticket variable discussed earlier, which consisted of four categories: sequence, pattern, nonequilibrated, and random. The values of variables measured at the *ordinal* level of measurement can be placed in an order relative to the other values; age (young, middle age, elderly) is one example of an ordinal variable.

One way to display nominal and ordinal variables is to use a bar chart. A **bar chart** is a figure that uses bars to represent the frequency or percentage of a sample corresponding to each value of a variable, with the different bars not touching each other. Figure 2.1 shows a bar chart for the data from the lottery ticket variable. In this figure, each bar corresponds to one of the four types of lottery tickets; the height of each bar corresponds to the frequency (*f*) of nonproblem gamblers who selected that type of ticket.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

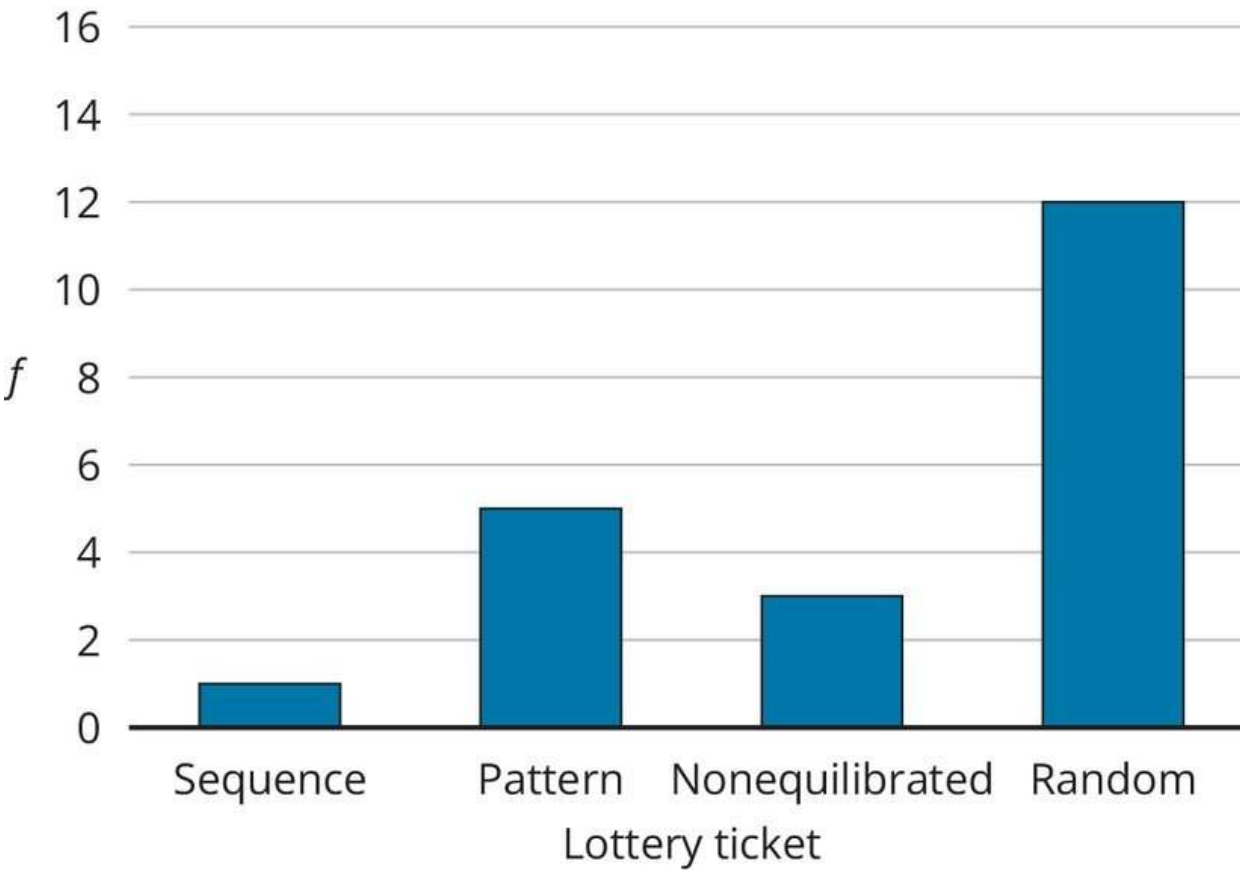


Figure 2.1 Bar Chart Showing the Frequencies of the Lottery Ticket Variable for the Nonproblem Gamblers

When creating a bar chart, the horizontal (X) axis contains the name of the variable and the values of the variable; the vertical (Y) axis contains the frequency (*f*) or percentage (%) for each value. Notice that the bars in the bar chart in Figure 2.1 do not touch each other; the gap between the bars indicates that the values of the variable represent distinct groups or categories that can't be placed along a numeric continuum.

A second type of figure used to display nominal or ordinal variables is a pie chart. A **pie chart** is a figure that uses a circle divided into proportions (slices) that represent the percentage of the sample for each value of a variable. Pie charts are created by dividing the 360 degrees of a circle into proportions corresponding to the percentage of each value.

Figure 2.2 shows a pie chart for the data for the lottery ticket variable. To illustrate how pie charts are constructed, the frequency distribution table in Table 2.4 shows that 5% of nonproblem gamblers chose the sequence ticket. Given that a pie chart consists of 360 degrees, 5% of the 360 degrees are needed to represent the sequence ticket. As 5% * 360 = 18, 18 degrees of the pie is assigned to the sequence ticket. Similarly, 97 degrees of the pie chart (27% of 360) are used to represent the pattern ticket. This process continues until all 360 degrees of the pie chart are accounted for.

The decision whether to use a bar chart or pie chart to illustrate a variable may depend on the number of values for the variable. For example, a pie chart for a variable that has 12 values may

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

result in many of the 12 “slices” being very thin and small, making them hard to see, meaning that a bar chart may be a better choice to illustrate this variable.

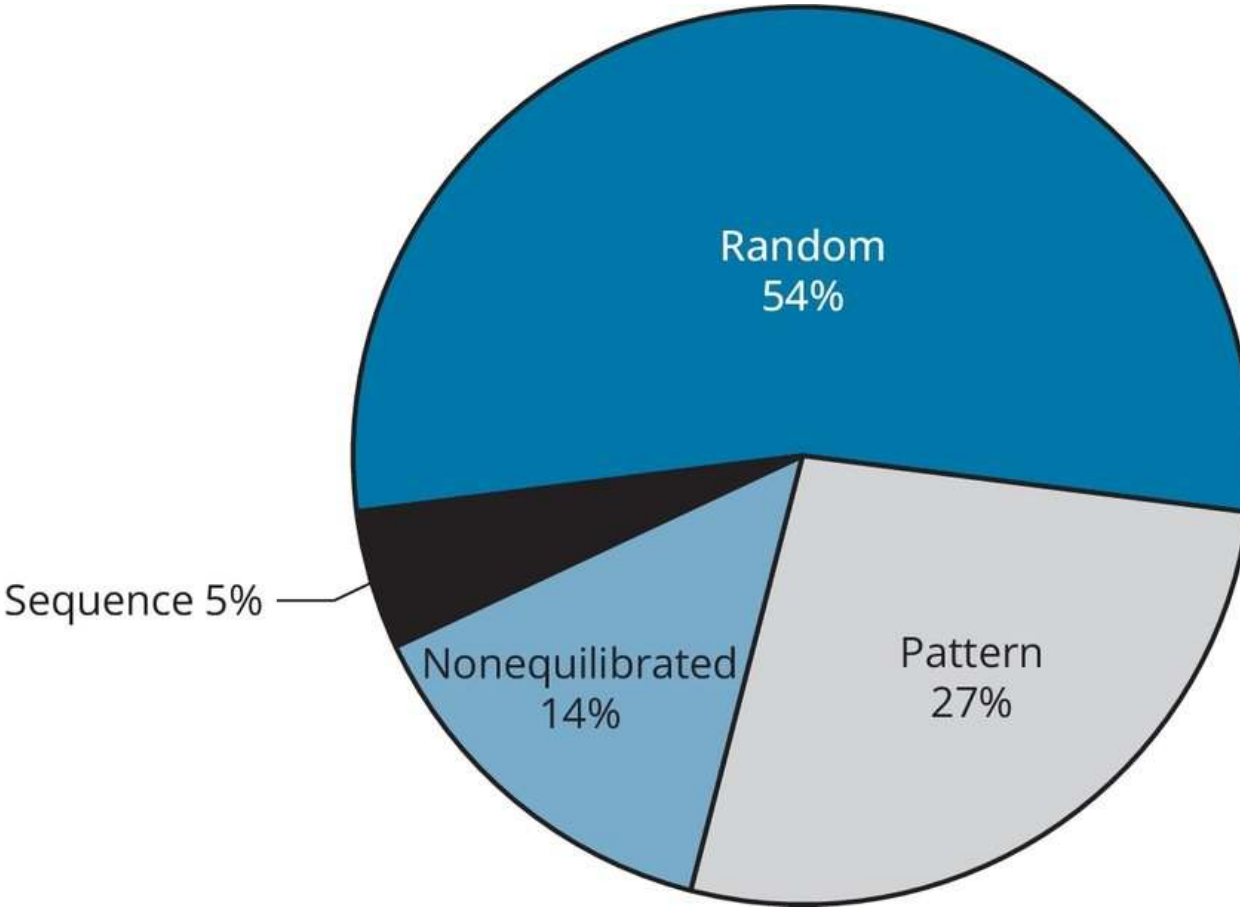


Figure 2.2 Pie Chart Showing the Percentages of the Lottery Ticket Variable for the Non−Problem Gamblers

Displaying Interval and Ratio Variables: Histograms and Frequency Polygons

Interval and ratio variables have values equally spaced along a numeric continuum and are identical to each other with one exception: Ratio scales possess a true zero point, for which the value of zero (0) represents the complete absence of the variable. The values of interval and ratio variables may be illustrated using histograms and frequency polygons, the choice of which depends on the number of values for the variable.

A **histogram** is a figure in which bars are used to represent the frequency of each value of a variable, with the different bars touching each other. The name and values of the variable are located on the X-axis, and the frequency or percentage for each value is on the Y-axis. Histograms are similar to bar charts, with one important difference: The bars in a histogram touch each other, indicating that the values of the variable are connected to each other along a numeric continuum.

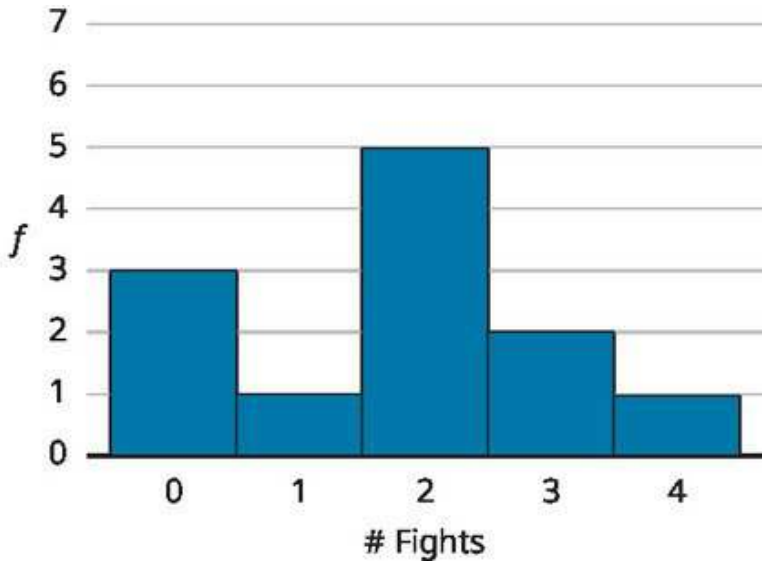
join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

As an example, imagine a teacher is interested in studying aggressive behavior in children and measures this variable by counting the number of fights each of her students is involved in over a 3-day period. “Number of fights” is a ratio variable rather than an interval variable because the value of 0 represents the complete absence of fights. Figure 2.3 provides a frequency distribution table and histogram for this variable.

A histogram is used when an interval or ratio variable consists of a relatively small number of values. But what if, for example, the number of fights could be anywhere from 0 to 20 rather than 0 to 5? In order to have the histogram fit on one page, the 21 bars would need to be very thin and crowded together, making the histogram confusing and difficult to read. In situations such as this, the values of the variable may be more clearly illustrated using a figure known as a **frequency polygon**, which is a line graph that uses data points to represent the frequency of each value of a variable, with straight lines connecting the data points. In a frequency polygon, the X-axis represents the name and values of the variable, while the Y-axis represents the frequency for each value.

# Fights	*f*
4	1
3	2
2	5
1	1
0	3
Total	12



▶ Long Description: Figure 2.3 Frequency Distribution Table and Histogram for the Number of Fights Variable

Figure 2.3 Frequency Distribution Table and Histogram for the Number of Fights Variable

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

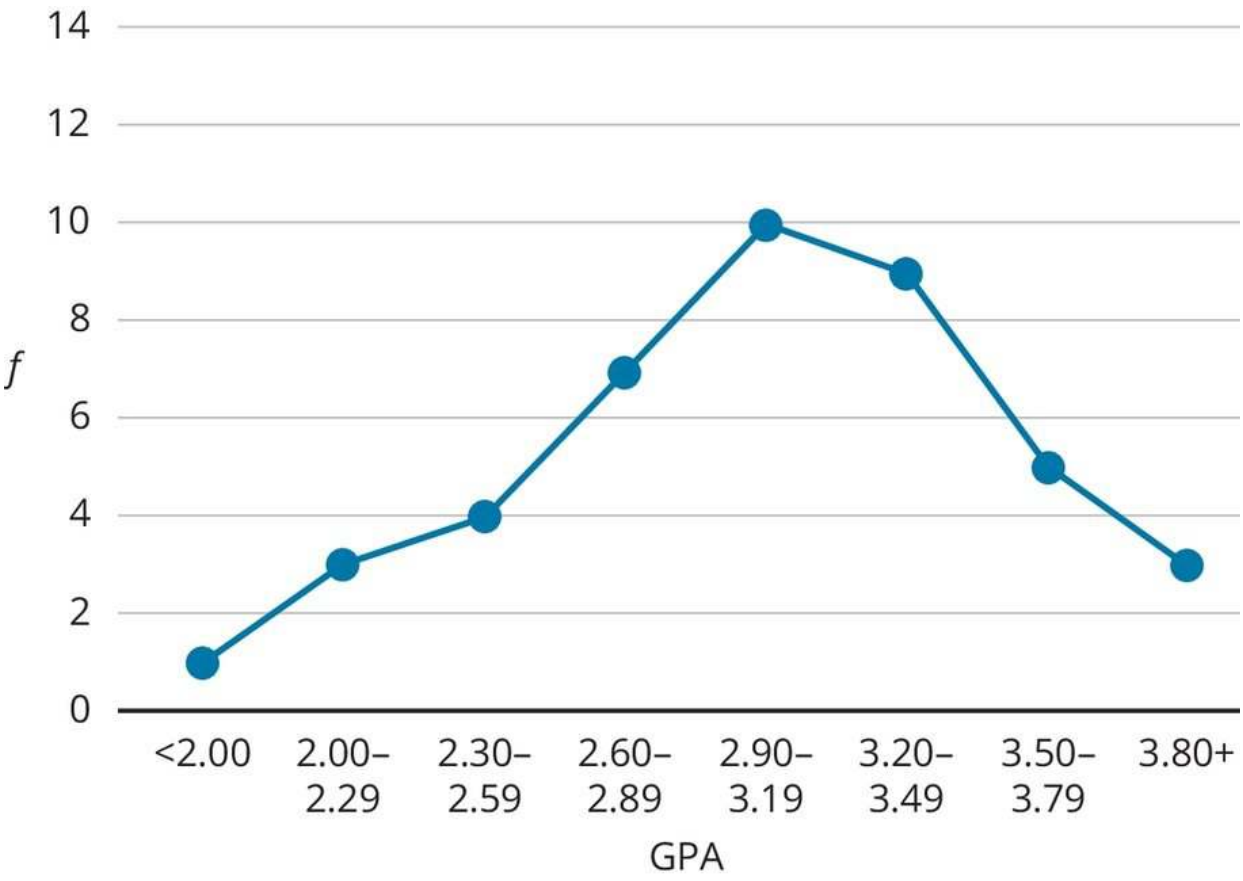


Figure 2.4 Frequency Polygon Showing the Frequency of Students in Each Interval of the Grade Point Average (GPA) Variable

Figure 2.4 shows a frequency polygon based on the grouped frequency distribution table of the earlier example of students' GPA (Table 2.7(b)). In this frequency polygon, the data points representing the frequencies of the eight GPA intervals are connected using straight, sharp lines. These lines indicate that the values of the variable can be connected together along a numeric continuum.

Although a frequency polygon like the one in Figure 2.4 may look jagged as it moves from one data point to the next, it provides a general sense of the shape of the distribution of scores for a variable. For example, looking at Figure 2.4 , you might conclude that the majority of GPAs in this sample are in the middle of the 1.00 to 4.00 range, with relatively low frequencies of students with either low or high GPAs.

Learning Check 2

Reviewing What You've Learned So Far

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



1. Review questions
    a. What is the difference between a frequency distribution table and a grouped frequency distribution table?
    b. For what levels of measurement would you create a pie chart, bar chart, histogram, or a frequency polygon?
    c. What is the difference between a bar chart and a histogram?
    d. Why would it be inappropriate to create a histogram or frequency polygon for a nominal or ordinal variable?
    e. Under what circumstances might you use a frequency polygon rather than a histogram to graph an interval or ratio variable?
2. For each of the following variables, identify whether you could use a bar chart, pie chart, histogram, and/or a frequency polygon to visually display the data.
    a. Miles per gallon (MPG)
    b. Living situation (on-campus, off-campus)
    c. Number of brothers and sisters
    d. Marital status (single, married, divorced)

Drawing Inappropriate Conclusions From Figures

A well-constructed figure can greatly facilitate the understanding of complex information. However, a poorly designed figure may not only be confusing, it can lead to inappropriate or misleading conclusions. Box 2.1 provides guidelines for constructing figures based on recommendations from the American Psychological Association (APA).

A simple example may demonstrate how altering the features of a figure can lead to different interpretations of the same set of data. Imagine a pollster assesses how voters with different political affiliations plan to vote on a ballot measure. She asks a sample of Democrats, Republicans, and Independents the question, “Do you plan on voting in favor of Measure A?” In examining her data, the pollster finds 56% of Democrats, 47% of Republicans, and 52% of Independents intend to vote “yes” on the measure.

Figure 2.5 graphs these three percentages using two bar charts; bar charts are used to display the data because the variable of interest (political affiliation) is a nominal variable (Democrat, Republican, or Independent). The bars don’t touch each other because the three political affiliations are distinct from each other and can’t be placed along a numeric continuum.

Although the two bar charts are similar, there is one important difference in how they are constructed. Following the guidelines in Box 2.1, the bar chart in Figure 2.5 (a) starts the Y-axis with the value 0% and ends with 100%. However, the bar chart in Figure 2.5 (b) starts with the value 45% and ends with 65%. By limiting the range of values of the percentages, the bar chart in Figure 2.5 (b) exaggerates differences in the heights of the three bars. As a result, someone looking at Figure 2.5 (b) may interpret the differences between the three political affiliations as being larger than someone looking at Figure 2.5 (a).

The saying “a picture is worth a thousand words” is particularly relevant for the presentation of statistical data. Because figures can provide a vivid representation of data, it’s the responsibility of researchers to portray their data in an accurate manner and to avoid creating images that may mislead or confuse.

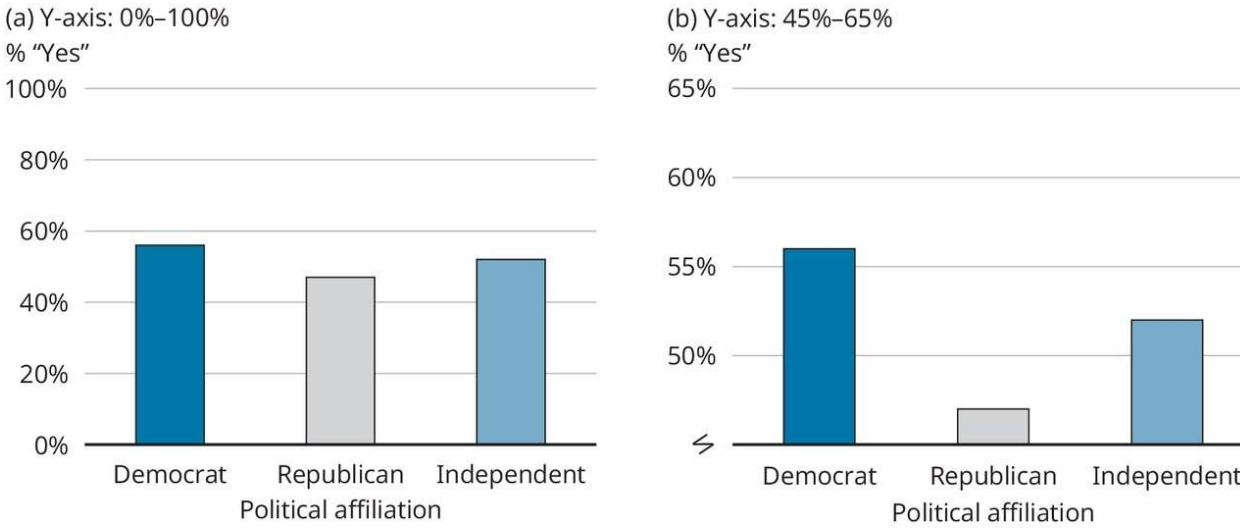
join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Box 2.1 Some APA Guidelines for Creating Figures

1. The figure should be formatted in a portrait rather than landscape page orientation (the reader should not turn the page sideways to read the table).
2. The figure is numbered using Arabic numbers (Figure 1, not Figure *I* or Figure *A*).
3. The figure includes a title that describes the contents of the figure. For example, it is not enough to name a figure simply " Figure 2.1 " or " Figure 2.1 . Bar chart."
4. Figures have a rectangular shape, with the height of the vertical (Y) axis two thirds to three fourths the length of the horizontal (X) axis.
5. The horizontal (X) axis contains the values of the variable; the vertical (Y) axis contains the frequency (*f*) or percentage (%) of the sample having each value of the variable.
6. The *X* - and *Y* -axes include descriptive labels written in both upper- and lowercase letters (as opposed to all capital letters).
7. The labels for the *X* - and *Y* -axes are short enough to fit on one line along the axes.
8. The label of the Y-axis is often placed parallel to the Y-axis (the letters face sideways rather than vertically) but can sometimes be seen on top of the y-axis for accessibility.
9. The Y-axis starts with the value zero (0) and is divided into equally spaced values or intervals.
10. The Y-axis is divided into 6 to 10 values (just enough values to accommodate bar length). The bars should neither be flattened along the bottom of the Y-axis nor bumped up against the top of the Y-axis.

Political affiliation	% "Yes"
Democrat	56%
Republican	47%
Independent	52%



▶ Long Description: Figure 2.5 Different Displays of the Same Data

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Figure 2.5 Different Displays of the Same Data

2.6 Examining Data: Describing Distributions

We've discussed how researchers construct tables and figures to organize their data. Looking at a table or figure enables researchers to describe the distribution of values or scores for a variable. Describing a distribution involves addressing questions such as, "Which values of the variable occur most often?," "How do the frequencies of the values of the variable change in relation to the most common values?," and "What is the amount of differences among the different values?"

Questions such as these pertain to three aspects of distributions: modality, symmetry, and variability. These three aspects of distributions are introduced in the following sections; they'll also be discussed in greater detail in the next two chapters of this book.

Modality

The first feature of a distribution of data for a variable is its modality. The **modality** of a distribution refers to the value or values of the variable that have the highest frequency or occur most often in a set of data. Using the lottery study as an example, identifying the modality of the lottery ticket variable involves answering the question, "Which of the four lottery tickets was chosen the most often?"

The modality of a distribution can be identified by looking at a frequency distribution table; the modality is the value(s) of the variable with the highest frequency (*f*) or percentage (%). The modality may also be determined by examining a figure and locating the tallest bar in a bar chart or histogram, the highest point in a frequency polygon, or the biggest "slice" in a pie chart. For the lottery ticket variable, the modality is the random ticket.

In identifying the modality of a distribution, it may be that one value or a limited number of values adjacent to one another clearly have the highest frequency. In such cases, the distribution is referred to as a unimodal distribution. A **unimodal distribution** is a distribution where one value of a variable occurs with the greatest frequency. Looking at the frequency distribution table (Table 2.4) or figure (Figure 2.1) for the lottery ticket variable, this distribution would be considered unimodal as one value, the random ticket, clearly occurred more often than the other three ticket choices.

In contrast to a unimodal distribution, it's possible that two distinct values may both have relatively high frequencies; if so, the distribution is referred to as a **bimodal distribution** , which is a distribution where two values clearly distinct from each other have the greatest frequency. A bimodal distribution sometimes resembles the two humps on a camel's back. As an example of a bimodal distribution, in April 2024 the Pew Research Center (www.pewresearch.org) asked a sample of 7,000 voters whether they felt Donald Trump had the mental fitness needed to be president, with five responses ranging from "extremely" to "not at all." The majority of voters chose one of the two extreme choices, either "extremely" (23%) or "not at all" (35%), with only 14% choosing the middle response of "somewhat."

And, in addition to unimodal and bimodal distributions, it's possible for a distribution to have more than two values with relatively equally high frequencies; this type of distribution is known as a **multimodal distribution** —a distribution where more than two values have the greatest frequency. In examining people's attitudes toward gun control, in May 2023, the news network CNN asked 1,227 people the question, "Do you think allowing gun owners to carry their guns in public places makes those places safer, less safe, or makes no difference in how safe a public place is?" They found people differed greatly in their responses, with 32% saying "makes public places safer," 36% saying

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

"makes public places less safe," and 32% saying "makes no difference," Figure 2.6 presents an example of a unimodal, bimodal, and multimodal distribution.

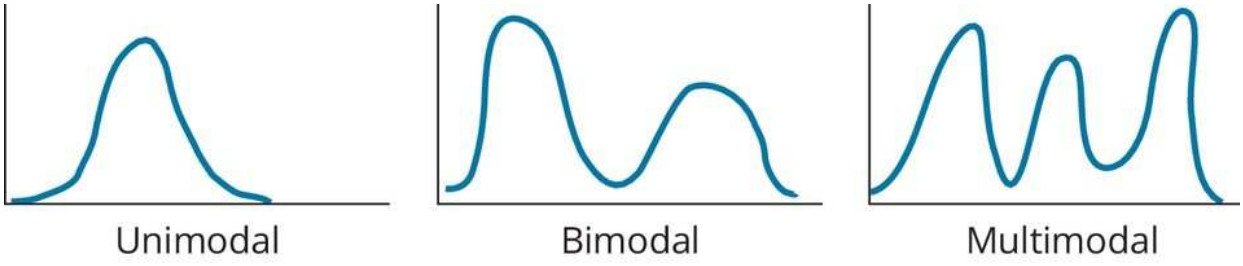


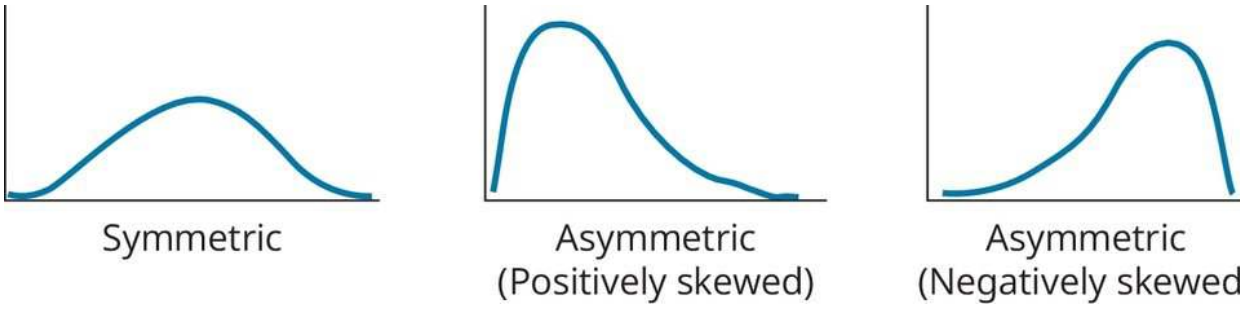
Figure 2.6 Distributions With Different Modalities (Unimodal, Bimodal, and Multimodal)

Symmetry

The second feature of a distribution of data is its symmetry. The **symmetry** of a distribution refers to how the frequencies of values of a variable change in relation to the most common or frequently occurring values. Describing the symmetry of a distribution means answering the question, "What is the shape of the distribution?"

The symmetry of a distribution can be determined by looking at the frequencies in a frequency distribution table or the bars or points in a figure. By doing so, a distribution may be described as being either symmetric or asymmetric. A **symmetric distribution** is one in which the frequencies change in a similar manner moving away in both directions from the most frequently occurring values. An example of a symmetric distribution is illustrated in Figure 2.7 . One way to determine whether a distribution is symmetric is to draw the distribution on a piece of paper and then fold the distribution vertically down the middle so that the two halves are on top of each other. A distribution is symmetric when the two halves are similar in shape such that they resemble mirror images of each other.

However, a distribution may also be asymmetric. An **asymmetric distribution** (also known as a *skewed* distribution) is one in which the frequencies change in a different manner moving away in both directions from the most frequently occurring values. In asymmetric distributions, the highest frequencies are located at one end of the distribution rather than in the middle, such that if you fold an asymmetric distribution down the middle, the shapes of the two halves do not look the same.



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Figure 2.7 Distributions With Different Symmetries (Symmetric, Positively Skewed, and Negatively Skewed)

In an asymmetric distribution, the frequencies take on the shape of a long tail as the values move toward the end of the distribution; scores in the tail of the distribution are often considered to be outliers. For example, in the 2022 Current Population Survey, the U.S. Census Bureau (www.census.gov) estimated that most Americans (63%) had annual incomes less than $50,000. However, a very small percentage (3%) were greater than $200,000; these higher incomes would be considered outliers.

Depending on the location of its tail, asymmetric distributions are referred to either as positively skewed or negatively skewed. A **positively skewed distribution** is one in which the higher frequencies are at the lower end of the distribution, with the tail on the upper (right) end of the distribution; a **negatively skewed distribution** is one in which the higher frequencies are at the upper end of the distribution, with the tail on the lower (left) end of the distribution. Figure 2.7 provides an illustration of a symmetric distribution and two asymmetric distributions (one positively skewed and one negatively skewed). Annual income, illustrated by the U.S. Census Bureau example mentioned earlier, is an example of a positively skewed distribution.

Variability

The third way in which distributions may be described is in terms of the amount or nature of differences among the different values for a variable. The **variability** of a distribution refers to the amount of differences in a distribution of data for a variable. The variability in a distribution is described by answering the question, “To what extent are the scores in the distribution different from each other?”

The amount of variability in a distribution can be estimated by looking at a frequency distribution table or figure. Imagine, for example, an instructor gives an 11-item quiz to her class and finds almost all of the students answer six, seven, or eight of the items correctly. In this situation, the distribution would resemble the shape of a tall mountain: thin and tall, with many of the scores residing in a few values of the variable. A distribution with a small amount of variability such as this one is referred to as a **peaked distribution** , defined as a distribution where much of the data are in a small number of values of a variable.

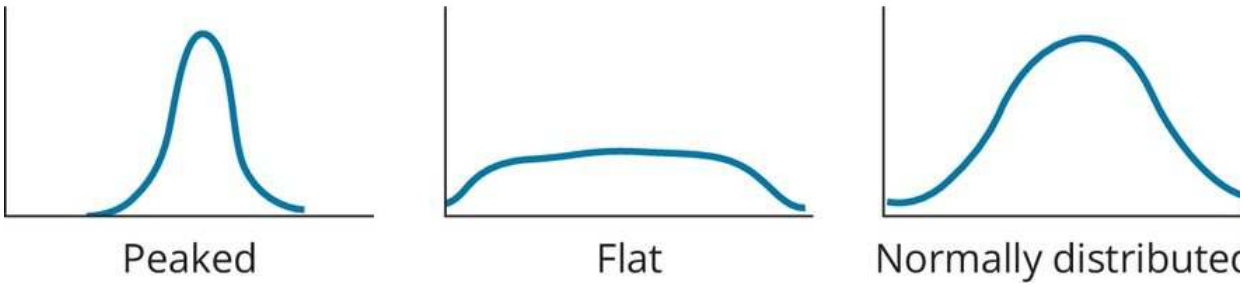


Figure 2.8 Distributions With Different Variability (Peaked, Flat, and Normally Distributed)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

On the other hand, what would the distribution look like if she had found her students' scores were distributed relatively evenly across the entire range of possible scores? In this situation, the distribution would not be tall and thin like a mountain but rather would be low and wide like a plateau. This type of distribution is referred to as a **flat distribution** , which is a distribution where the data are spread evenly across the values of a variable. The CNN gun control example described earlier is an example of a flat distribution.

In discussing modality, symmetry, and variability, distributions may have more than one mode, may be asymmetric (skewed), and may be peaked or flat. But what if a distribution has none of these qualities? A distribution that is unimodal, symmetric, and neither peaked nor flat may be described as "bell shaped." If you look at a bell, you'll notice that it is the highest in the middle, and as you move away from the top of a bell in both the left and right directions, the shape of the bell slopes downward in the same manner. A bell-shaped distribution for a variable is given a special name: a normally distributed variable.

A **normally distributed variable** is a distribution for a variable that is considered unimodal, symmetric, and neither peaked nor flat. Note that the word *normal* is a statistical concept and not a value judgment. In other words, asymmetric, peaked, or flat distributions are not considered "abnormal." The concept of normal distributions will be discussed in greater detail in Chapter 5. Figure 2.8 provides an illustration of three distributions with different amounts of variability: peaked, flat, and normally distributed.

Learning Check 3

Reviewing What You've Learned So Far



1. Review questions
    a. What are the three aspects used to describe distributions of variables?
    b. What is the difference between unimodal, bimodal, and multimodal distributions?
    c. What is the difference between symmetric and asymmetric (skewed) distributions?
    d. What is the difference between positively and negatively skewed distributions?
    e. What is the difference between peaked and flat distributions?
    f. What are the characteristics of a normally distributed variable?
2. For each of the following sets of data, determine whether the distribution of scores is unimodal, bimodal, or multimodal.
    a. 4, 2, 11, 6, 10, 2, 6, 13, 2, 10, 6, 10, 1
    b. 15, 17, 13, 18, 17, 14, 20, 18, 17, 14, 17
    c. 20, 10, 50, 10, 80, 30, 70, 10, 60, 10, 80, 0, 30, 80, 10, 100, 80
3. For each of the following sets of data, determine whether the distribution of scores is symmetric or asymmetric (skewed).
    a. 4, 5, 1, 4, 2, 4, 3, 5, 3, 4, 4, 3
    b. 220, 230, 250, 230, 200, 240, 210, 230, 270, 220, 240, 230, 240, 230
    c. 3, 1, 9, 2, 3, 6, 2, 7, 3, 5, 2, 6, 3, 2, 2
4. For each of the following sets of data, determine whether the distribution of scores is peaked or flat.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

a. 16, 10, 19, 12, 13, 15, 14, 17, 15, 20, 11, 13, 19, 16, 15
b. 12, 15, 12, 11, 13, 11, 12, 12, 14, 10, 12, 11, 16, 12, 13, 12, 11
c. 14, 12, 16, 14, 13, 15, 12, 14, 14, 15, 14, 14, 13, 14

2.7 Looking Ahead

This chapter has discussed why and how researchers examine data. Examining data by creating tables and figures requires a number of different skills, including organizational skills (to prepare a set of data for statistical analysis), communication skills (to present data in tables or figures), and conceptual skills (to describe a distribution). The next two chapters focus on mathematical skills related to the statistical analysis of data, describing how researchers describe distributions numerically rather than visually. Given that this is a statistics textbook, we will focus on calculating statistics to describe and analyze data. However, and we cannot say this strongly enough, it is *extremely* important to examine a set of data before conducting statistical analyses as this will prevent errors both in your calculations of statistics as well as your interpretation of the results of these calculations.

2.8 Summary

Before conducting statistical analyses, researchers typically examine their data to gain an initial sense of the data, detect data coding or data entry errors, identify ***outliers*** (rare, extreme scores that lie outside of the range of the majority of scores in a set of data), evaluate research methodology, and determine whether the data meet statistical criteria and assumptions.

One common way of examining variables involves creating tables known as ***frequency distribution tables*** (tables that summarize the number and percentage of participants that express the different values for a variable) and ***grouped frequency distribution tables*** (tables that group the values of a variable measured at the interval or ratio level of measurement into a small number of intervals and then provide the frequency and percentage within each interval). To create a grouped frequency table, the ***real limits*** of each interval (the values of the variable that fall halfway between the top of one interval and the bottom of the next interval) need to be identified.

Data for a variable may also be examined by creating figures such as bar charts, pie charts, histograms, and frequency polygons. ***Bar charts*** and ***pie charts*** are used for variables measured at the nominal or ordinal level of measurement; a bar chart uses bars to represent the frequency or percentage of a sample corresponding to each value of a variable, with the different bars not touching each other; a pie chart is a circle divided into proportions that represent the percentage of the sample corresponding to each value of a variable. ***Histograms*** and ***frequency polygons*** are used for variables measured at the interval or ratio level of measurement; a histogram is a figure in which bars are used to represent the frequency of each value of a variable, with the different bars touching each other; a frequency polygon is a line graph that uses data points to represent the frequency of each value of a variable, with lines connecting the data points.

From examining their data, researchers can describe the distribution of values or scores for a variable. There are three main aspects of distributions: modality, symmetry, and variability.

The ***modality*** of a distribution is the value(s) of the variable that have the highest frequencies or occur most often in a set of data. ***Unimodal distributions*** are distributions where one value occurs with the greatest frequency, ***bimodal distributions*** have two values with the greatest frequency, and ***multimodal distributions*** have more than two values with the greatest frequencies.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The ***symmetry*** of a distribution refers to how the frequencies of values of a variable change in relation to the most common or frequently occurring values of the variable. In ***symmetric distributions*** , the frequencies change in a similar manner, moving away in both directions from the most common values of the variable. In ***asymmetric (skewed) distributions*** , the frequencies change in a different manner, moving away in both directions from the most common values of the variable.

The ***variability*** of a distribution refers to the amount of differences in a distribution of data. In ***peaked distributions*** , much of the data are in a few values of a variable; in ***flat distributions*** , the data are spread evenly across the values of a variable. A ***normally distributed variable*** has a symmetric distribution that is neither peaked nor flat.

2.9 Important Terms

asymmetric (skewed) distribution

bar chart

bimodal distribution

cumulative percentage

flat distribution

frequency (*f*)

frequency distribution table

frequency polygon

grouped frequency distribution table

histogram

modality

multimodal distribution

negatively skewed distribution

normally distributed variable

outlier

peaked distribution

pie chart

positively skewed distribution

real limits

real lower limit

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

real upper limit

symmetric distribution

symmetry

unimodal distribution

variability

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

2.10 Formula Introduced in This Chapter

Percentage (%)

$$\% = \frac{f}{total\ number\ of\ scores} * 100$$

2.11 Using IBM SPSS Software *

Creating Data Files: Defining Variables and Entering Data

* IBM SPSS software. SPSS is a registered trademark of International Business Machines Corporation.

1. Open a new data file (See Figure 2.9).

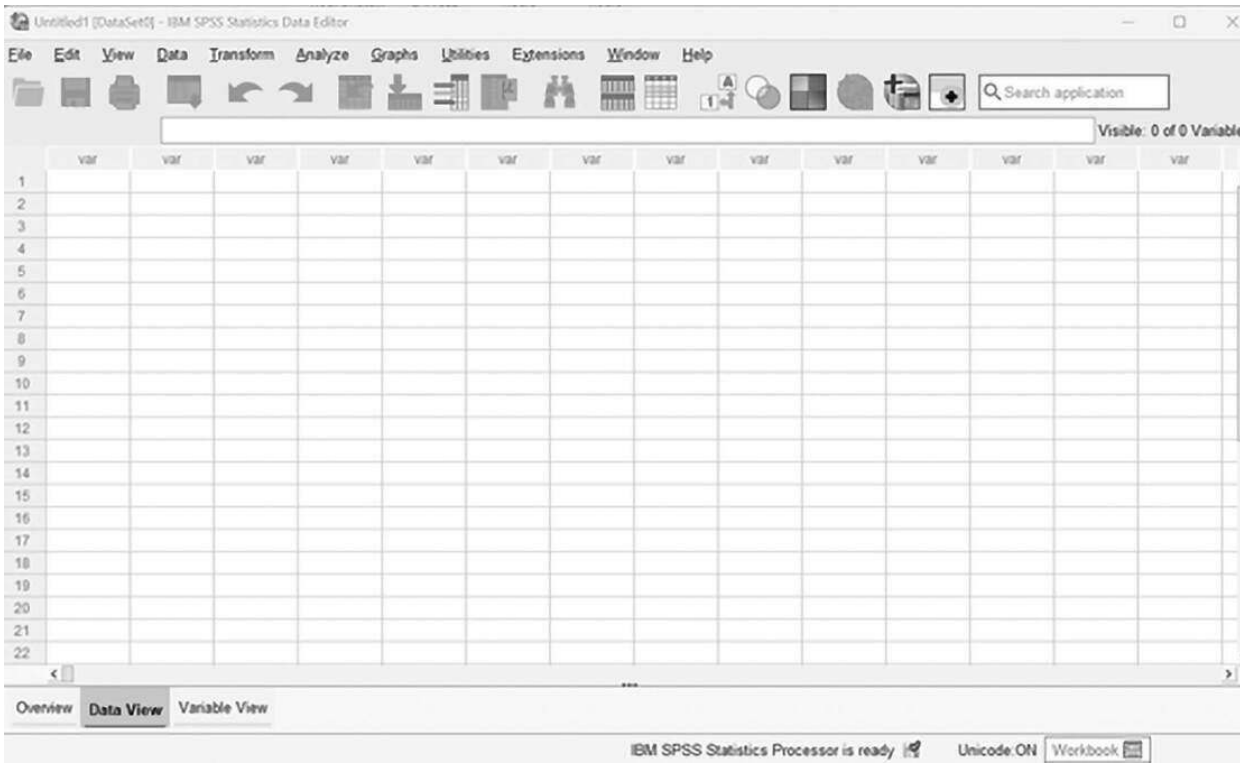


Figure 2.9 Data View

2. Define and name variable(s) (See Figure 2.10).

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

How? (1) Click **Variable View** tab, (2) click below **Name** , and (3) type name of variable in the cell.

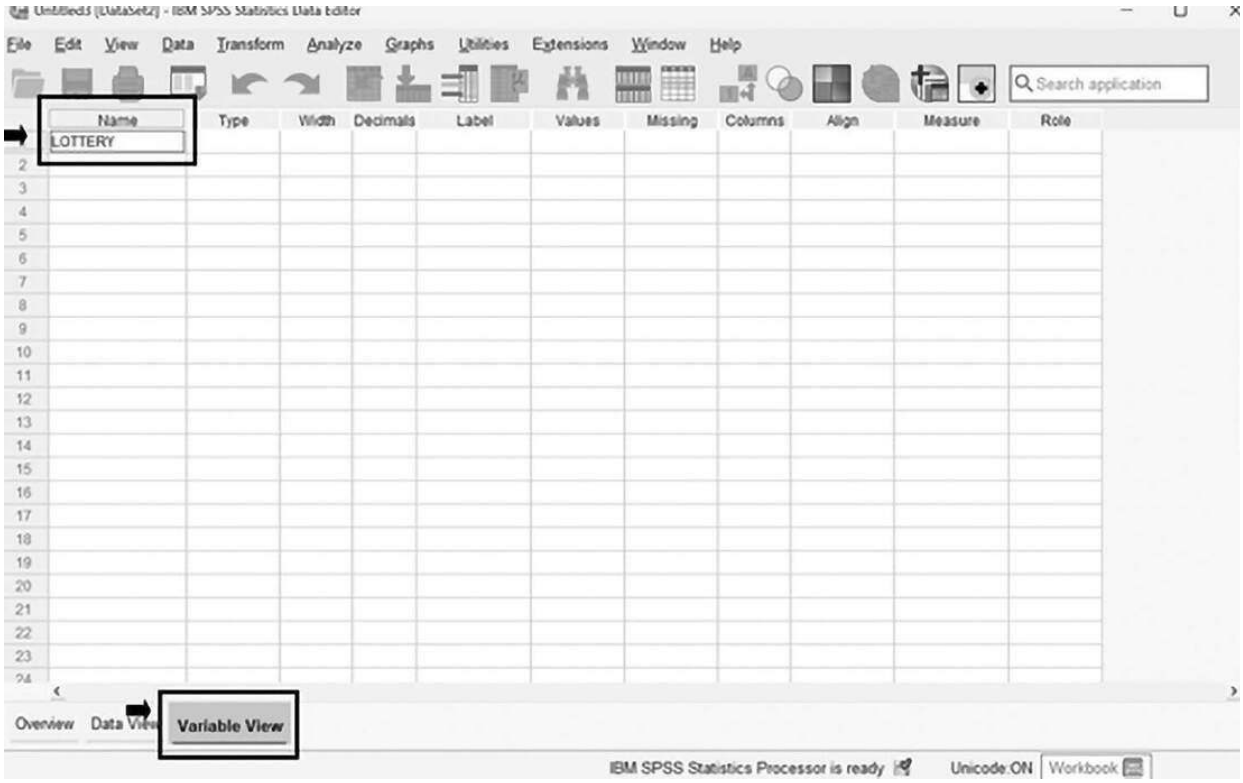


Figure 2.10 Variable View

3. Determine number of decimals for your variable(s) (See Figure 2.11).

How? Click desired number of decimal places in the **Decimals** box.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

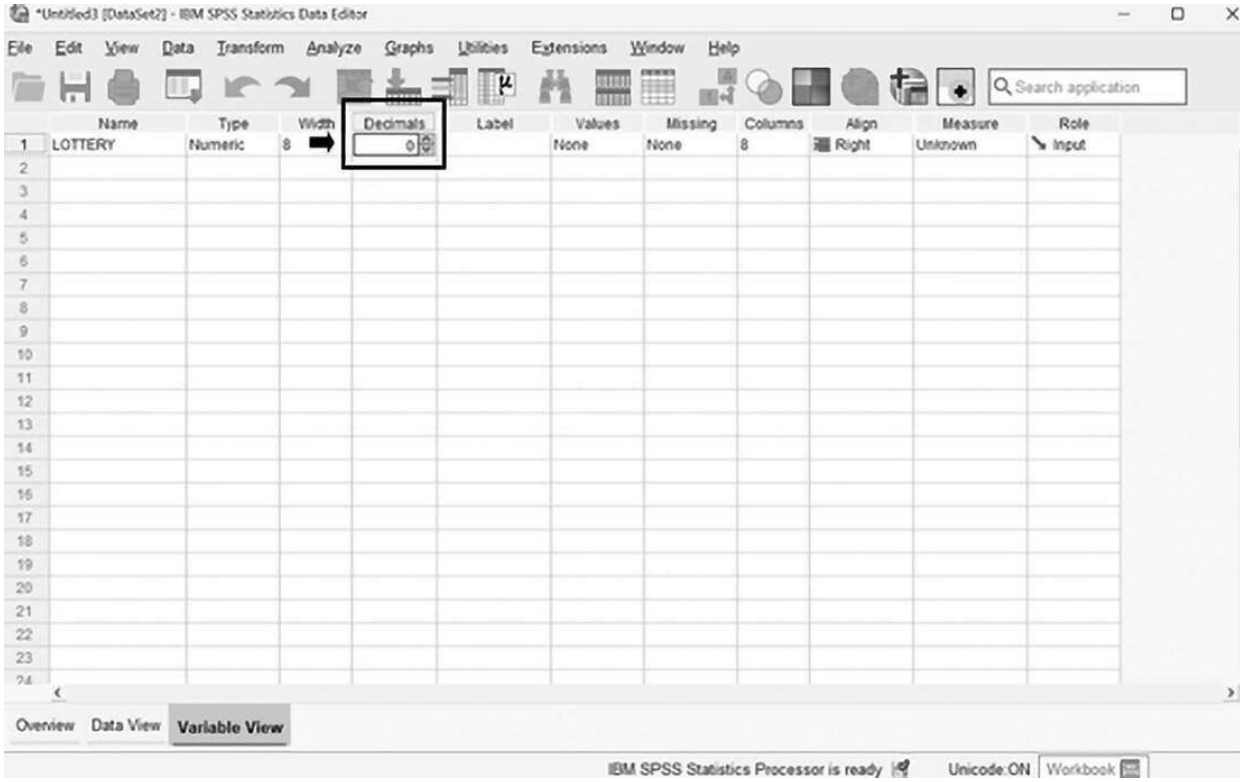


Figure 2.11 Decimals Box

4. Provide labels for your variable(s) (See Figure 2.12).

How? (1) Click below **Label** , and (2) type the label for the variable in the cell.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

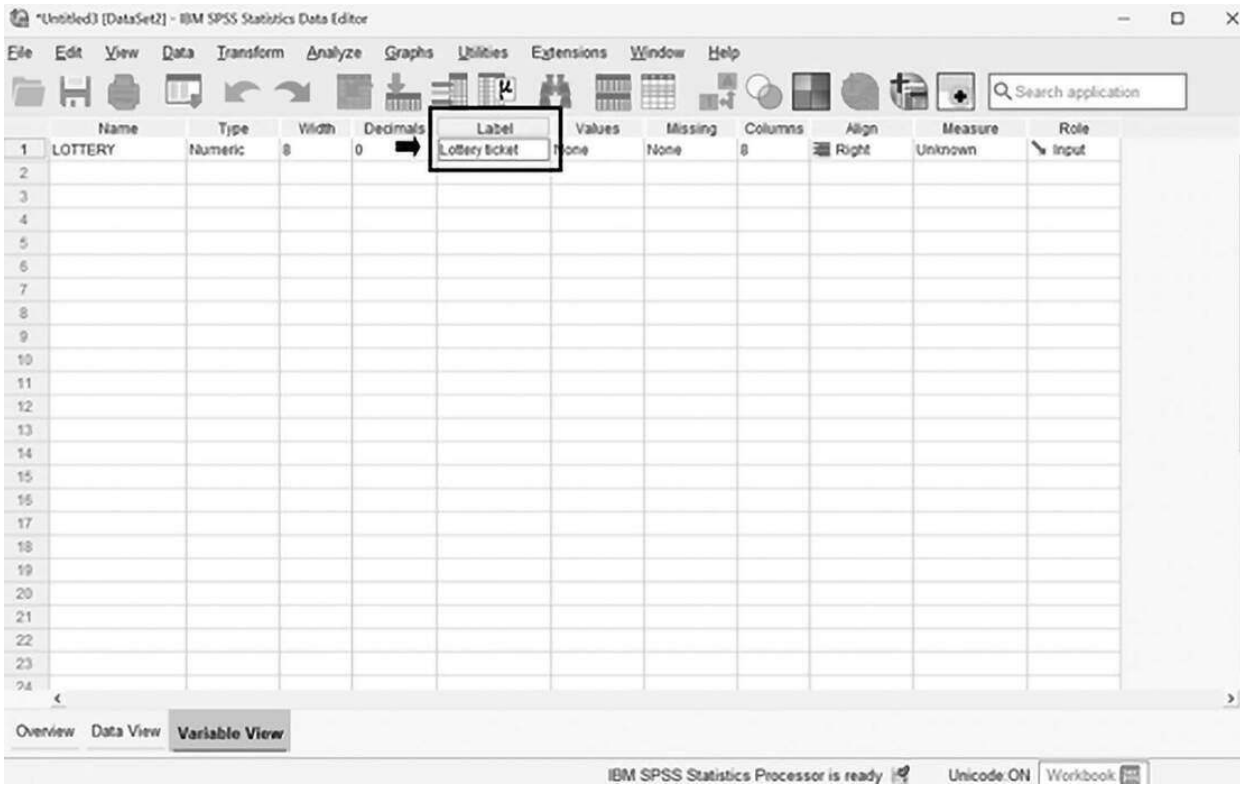


Figure 2.12 Label Box

5. Provide labels for values of nominal (categorical) variable(s) (See Figure 2.13).

How? (1) Click below **Values** , and (2) type labels for coded values of the variable.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

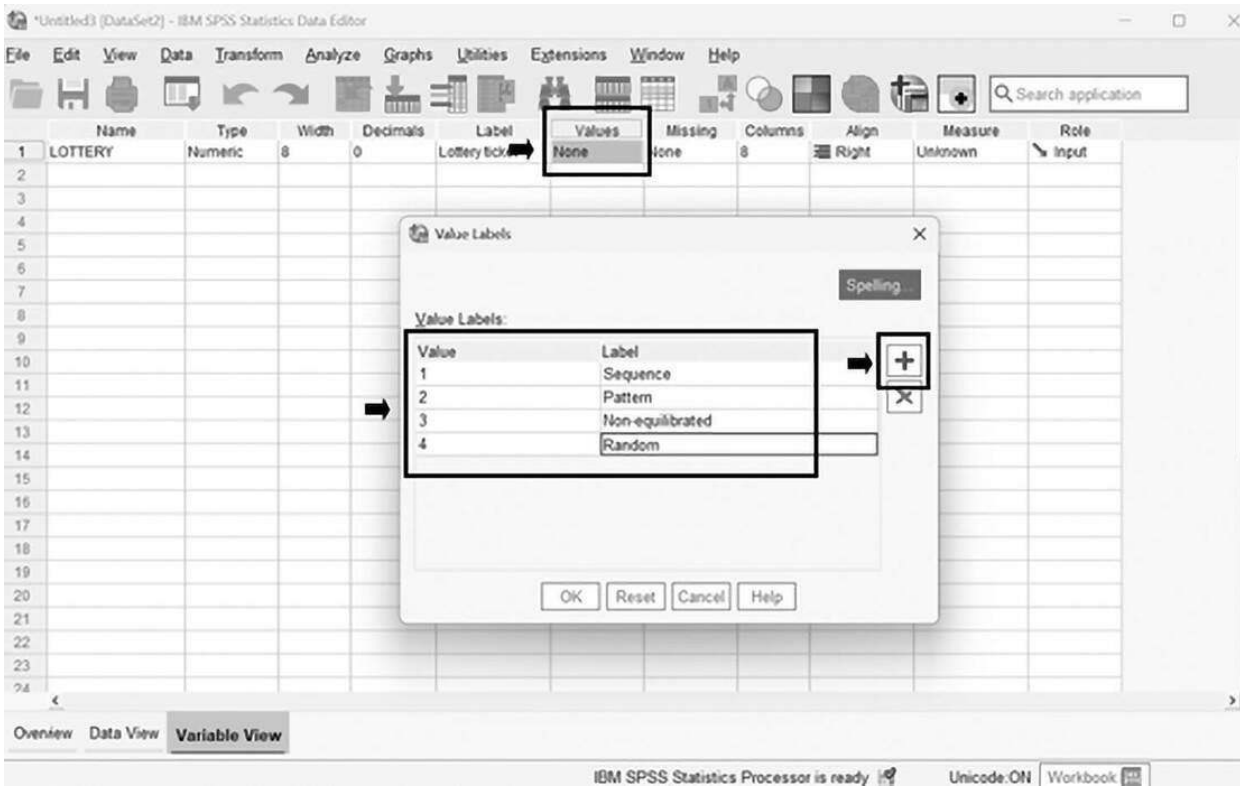


Figure 2.13 Nominal Variables

6. Enter data (See Figure 2.14).

How? (1) Click **Data View** tab and (2) enter data for the variable in the cells.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

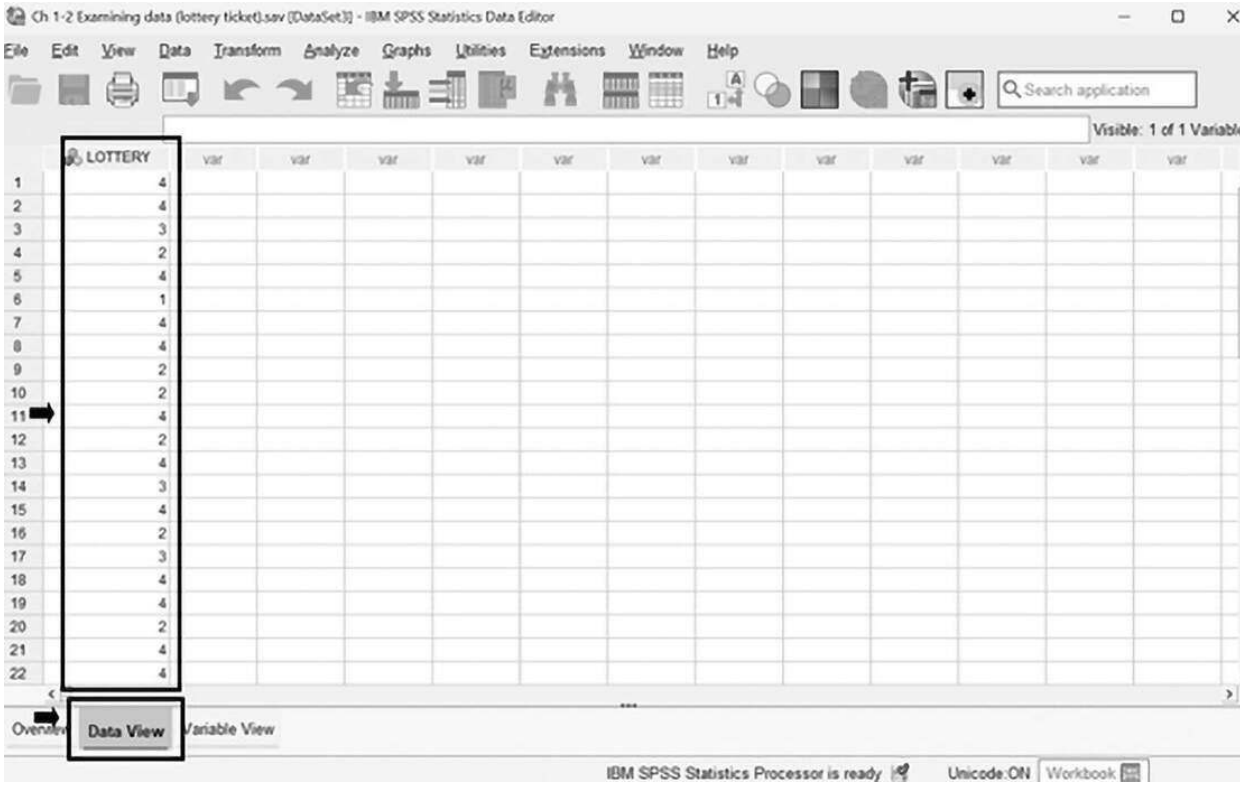


Figure 2.14 Enter Data View

Examining Data: Frequency Distribution Tables and Figures: The Lottery Study (2.1)

1. Select the frequency distribution procedure within SPSS (See Figure 2.15).

   How? (1) Click **Analyze** menu, (2) click **Descriptive Statistics** , and (3) click **Frequencies** .

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

To Jennifer , my kids Meagan and Eika , and my parents, Katsumi and Grayce Tokunaga.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

As a global academic publisher, Sage is driven by the belief that research and education are critical in shaping society. Our mission is building bridges to knowledge—supporting the development of ideas into scholarship that is certified, taught, and applied in the real world.

Sage's founder, Sara Miller McCune, transferred control of the company to an independent trust, which guarantees our independence indefinitely. This enables us to support an equitable academic future over the long term by building lasting relationships, championing diverse perspectives, and co-creating social and behavioral science resources that transform teaching and learning.

▶ Long Description:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

FUNDAMENTAL STATISTICS FOR THE SOCIAL AND BEHAVIORAL SCIENCES

Third Edition

Howard T. Tokunaga

San Jose State University



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

eISBN: 978-1-0718-6160-8

Copyright © 2027 by Sage.

All rights reserved. Except as permitted by U.S. copyright law, no part of this work may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without permission in writing from the publisher.

All third party trademarks referenced or depicted herein are included solely for the purpose of illustration and are the property of their respective owners. Reference to these trademarks in no way indicates any relationship with, or endorsement by, the trademark owner.



FOR INFORMATION:

2455 Teller Road, Thousand Oaks, California 91320, E-mail: orders@sagepub.com

3rd floor, HYLO, 103-105 Bunhill Row, London, EC1Y 8LZ, United Kingdom

10th Floor, Emaar Capital Tower 2, MG Road Sikanderpur, Sector 26, Gurugram, Haryana – 122002, India

18 Cross Street #10-10/11/12, China Square Central, Singapore 048423

Library of Congress Control Number: 2026943277

ISBN: 978-1-0718-6144-8

Printed in the United States of America

This book is printed on acid-free paper.

26 27 28 29 30 10 9 8 7 6 5 4 3 2 1

Acquisitions Editor: Anna Villarruel

Content Development Editor: Kenzie Offley

Editorial Assistant: Daria Terry

Production Editor: Syeda Aina Rahat Ali

Copy Editor: Michelle Ponce

Typesetter: Lumina Datamatics

Cover Designer: Candice Harman

Marketing Manager: Victoria Velasquez

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CONTENTS

Preface
Acknowledgments
About the Author
Chapter 1 Introduction to Statistics
Chapter 2 Examining Data: Tables and Figures
Chapter 3 Measures of Central Tendency
Chapter 4 Measures of Variability
Chapter 5 Normal Distributions
Chapter 6 Probability and Introduction to Hypothesis Testing
Chapter 7 Testing One Sample Mean
Chapter 8 Estimating the Mean of a Population
Chapter 9 Testing the Difference Between Two Means
Chapter 10 Errors in Hypothesis Testing, Statistical Power, and Effect Size
Chapter 11 One-Way Analysis of Variance (ANOVA)
Chapter 12 Two-Way Analysis of Variance (ANOVA)
Chapter 13 Correlation
Chapter 14 Linear Regression and Multiple Correlation
Chapter 15 Chi-Square
Appendix A: Tables
Appendix B: Review of Basic Mathematics
Glossary
References
Index

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

DETAILED CONTENTS

Preface

Acknowledgments

About the Author

Chapter 1 Introduction to Statistics

- 1.1 What Is Statistics?
- 1.2 Why Learn Statistics?
- 1.3 Introduction to the Stages of the Research Process
  - Developing a Research Hypothesis to Be Tested
    - Identifying a Question or Issue to Be Examined
    - Reviewing and Evaluating Relevant Theories and Research
    - Stating a Research Hypothesis: Independent and Dependent Variables
  - Collecting Data
    - Drawing a Sample From a Population
    - Determining How Variables Will Be Measured: Levels of Measurement
    - Selecting a Method to Collect the Data: Experimental and Nonexperimental Research Methods
  - Analyzing the Data
    - Calculating Descriptive Statistics
    - Calculating Inferential Statistics
  - Drawing a Conclusion Regarding the Research Hypothesis
  - Communicating the Findings of the Study
- 1.4 Plan of the Book
- 1.5 Looking Ahead
- 1.6 Summary
- 1.7 Important Terms
- 1.8 Exercises

Answers to Learning Checks

Learning Check 1

Learning Check 2

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Odd-Numbered Exercises

Chapter 2 Examining Data: Tables and Figures

2.1 An Example From the Research: Winning the Lottery

2.2 Why Examine Data?

Gaining an Initial Sense of the Data

Identifying Outliers

Evaluating Research Methodology

Determining Whether Data Meet Statistical Criteria and Assumptions

2.3 Examining Data Using Tables

What Are the Values of the Variable?

How Many Participants in the Sample Have Each Value of the Variable?

What Percentage of the Sample Has Each Value of the Variable?

2.4 Grouped Frequency Distribution Tables

Cumulative Percentages

2.5 Examining Data Using Figures

Displaying Nominal and Ordinal Variables: Bar Charts and Pie Charts

Displaying Interval and Ratio Variables: Histograms and Frequency Polygons

Drawing Inappropriate Conclusions From Figures

2.6 Examining Data: Describing Distributions

Modality

Symmetry

Variability

2.7 Looking Ahead

2.8 Summary

2.9 Important Terms

2.10 Formula Introduced in This Chapter

Percentage (%)

2.11 Using IBM SPSS Software *

Creating Data Files: Defining Variables and Entering Data

Examining Data: Frequency Distribution Tables and Figures: The Lottery Study (2.1)

2.12 Exercises

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 3 Measures of Central Tendency
3.1 An Example From the Research: The 10% Myth
3.2 Understanding Central Tendency
3.3 The Mode
Determining the Mode
3.4 The Median
Calculating the Median: Odd Number of Scores
Calculating the Median: Even Number of Scores
Calculating the Median When Multiple Scores Have the Median Value
3.5 The Mean
Calculating the Mean
Reporting Calculated Numbers
Rounding Calculated Numbers
Calculating the Mean From Frequency Distribution Tables
The Mean as a Balancing Point in a Distribution
The Sample Mean vs. the Population Mean: Statistics vs. Parameters
3.6 Comparison of the Mode, Median, and Mean
Strengths and Weaknesses of the Mode
Strengths and Weaknesses of the Median
Strengths and Weaknesses of the Mean
Choosing a Particular Measure of Central Tendency
3.7 Measures of Central Tendency: Drawing Conclusions
3.8 Looking Ahead
3.9 Summary
3.10 Important Terms
3.11 Formulas Introduced in This Chapter
Median (Odd Number of Scores)
Median (Even Number of Scores)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Mean
Mean (Frequency Distribution Table)
Mean (Balancing Point in a Distribution)
3.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 4 Measures of Variability
4.1 An Example From the Research: How Many Sometimes" in an Always"?
4.2 Understanding Variability
4.3 The Range
Strengths and Weaknesses of the Range
4.4 The Interquartile Range
Strengths and Weaknesses of the Interquartile Range
4.5 The Variance (s^2)
Definitional Formula for the Variance
Computational Formula for the Variance
Why Not Use the Absolute Value of the Deviation in Calculating the Variance?
Why Divide by N − 1 Rather Than N in Calculating the Variance?
4.6 The Standard Deviation (s)
Definitional Formula for the Standard Deviation
Computational Formula for the Standard Deviation
Eyeball Estimating" the Mean and Standard Deviation
4.7 Measures of Variability for Populations
The Population Variance (σ^2)
The Population Standard Deviation (σ)
Measures of Variability for Samples vs. Populations
4.8 Measures of Variability for Nominal Variables
The Index of Qualitative Variation
4.9 Measures of Variability: Drawing Conclusions

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

4.10 Looking Ahead
4.11 Summary
4.12 Important Terms
4.13 Formulas Introduced in This Chapter
Range
Interquartile Range
Variance (s^2) (Definitional Formula)
Variance (s^2) (Computational Formula)
Standard Deviation (s) (Definitional Formula)
Standard Deviation (s) (Computational Formula)
Population Variance (σ^2) (Definitional Formula)
Population Standard Deviation (σ) (Definitional Formula)
Index of Qualitative Variation (IQV)
4.14 Using SPSS
Calculating Measures of Central Tendency and Variability: The Frequency Rating Study (4.1)
4.15 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 5 Normal Distributions
5.1 Normal Distributions
What Are Normal Distributions?
Why Are Normal Distributions Important?
Is There Only One Normal Distribution?
5.2 The Standard Normal Distribution
What Is the Standard Normal Distribution?
Interpreting z-Scores: Distance From the Mean
Interpreting z-Scores: Relative Position Within the Distribution
Finding the Area Between the Mean and a z-Score

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Finding the Area Less or Greater Than a z-Score
Finding the Area Between Two z-Scores
5.3 Applying z-Scores to Normal Distributions
Transforming Normal Distributions Into the Standard Normal Distribution
Understanding the z-Score Formula
Example: Transforming SAT Scores Into z-Scores
Interpreting z-Scores: Distance From the Mean
Interpreting z-Scores: Relative Position Within Distribution
Finding the Area Between the Mean and a Score
Finding the Area Less or Greater Than a Score
Finding the Area Between Two Scores
Transforming Scores Into z-Scores: Drawing Conclusions
5.4 Standardizing Frequency Distributions
Example: Measuring Politicians' Behavior
Standardizing a Frequency Distribution
Interpreting Standardized Scores
Using Standardized Scores to Compare or Combine Variables
Characteristics of Standardized Distributions
When Can You Apply the Normal Curve Table to a Standardized Distribution?
5.5 Looking Ahead
5.6 Summary
5.7 Important Terms
5.8 Formulas Introduced in This Chapter
z-Score
Standardized Score
5.9 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 6 Probability and Introduction to Hypothesis Testing

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

6.1 A Brief Introduction to Probability
- What Is Probability?
- Why Is Probability Important to Researchers?
- Applying Probability to Normal Distributions
- Applying Probability to Binomial Distributions

6.2 Example: Making Heads or Tails of the Super Bowl

6.3 Introduction to Hypothesis Testing
- State the Null and Alternative Hypotheses (H_0 and H_1)
  - The Null Hypothesis (H_0)
  - The Alternative Hypothesis (H_1)
- Make a Decision About the Null Hypothesis
  - Defining a Low" Probability of a Statistic: Setting Alpha (α)
  - Identifying the Values of the Statistic With a Low Probability
  - Stating a Decision Rule
  - Calculating a Value of a Statistic
  - Making the Decision Whether to Reject the Null Hypothesis
- Draw a Conclusion From the Analysis
- Relate the Result of the Analysis to the Research Hypothesis
- Summary of Steps in Hypothesis Testing

6.4 Issues Related to Hypothesis Testing: An Introduction
- The Issue of Proof" in Hypothesis Testing
- Errors in Decision Making
- Factors Influencing the Decision About the Null Hypothesis
  - Sample Size
  - Alpha (α)
  - Directionality of the Alternative Hypothesis

6.5 Looking Ahead

6.6 Summary

6.7 Important Terms

6.8 Formula Introduced in This Chapter

Probability

6.9 Exercises

Answers to Learning Checks

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 7 Testing One Sample Mean
7.1 An Example From the Research: Do You Read Me?
7.2 The Sampling Distribution of the Mean
Characteristics of the Sampling Distribution of the Mean
7.3 Inferential Statistics: Testing One Sample Mean (σ Known)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: z-Test for One Mean
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the z-Test for One Mean
Summary
7.4 A Second Example From the Research: Unique Invulnerability
7.5 Introduction to the t-Distribution
Characteristics of the Student t-Distribution
7.6 Inferential Statistics: Testing One Sample Mean (σ Not Known)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: t-Test for One Mean
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the t-Test for One Mean

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Summary
7.7 Factors Affecting the Decision About the Null Hypothesis
Sample Size
Alpha (α)
Directionality of the Alternative Hypothesis
7.8 Looking Ahead
7.9 Summary
7.10 Important Terms
7.11 Formulas Introduced in This Chapter
z-Test for One Mean
Population Standard Error of the Mean $\left(\mathbf{i}t\sigma_{\mathbf{i}t\overline{X}}\right)$
Degrees of Freedom (df) for One Sample
t-Test for One Mean
Standard Error of the Mean $\left(\mathbf{i}ts_{\mathbf{i}t\overline{X}}\right)$
7.12 Using SPSS
Testing One Sample Mean (σ Not Known): The Unique Invulnerability Study (7.6)
7.13 Exercises
Answers to Learning Checks
Learning Check 2
Learning Check 3
Learning Check 5
Learning Check 6
Answers to Odd-Numbered Exercises
Chapter 8 Estimating the Mean of a Population
8.1 An Example From the Research: Salary Survey
8.2 Introduction to the Confidence Interval for the Mean
Using the Sampling Distribution of the Mean to Create the Confidence Interval for the Mean
8.3 The Confidence Interval for the Mean (σ Not Known)
State the Desired Level of Confidence
Calculate the Confidence Interval and Confidence Limits
Calculate the Standard Error of the Mean $\left(s_{\overline{X}}\right)$

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Calculate the Degrees of Freedom (df) and Identify the Critical Value of t

Calculate the Confidence Interval and Confidence Limits

Draw a Conclusion About the Confidence Interval

Probability of an Interval or the Probability of a Population Mean?

Summary

8.4 The Confidence Interval for the Mean (σ Known)

State the Desired Level of Confidence

Calculate the Confidence Interval and Confidence Limits

Calculate the Population Standard Error of the Mean $\left(\sigma_{\overline{X}}\right)$

Identify the Critical Value of z

Calculate the Confidence Interval and Confidence Limits

Draw a Conclusion About the Confidence Interval

Summary

8.5 Factors Affecting the Width of the Confidence Interval for the Mean

Sample Size

Estimating the Sample Size Needed for a Desired Confidence Interval

Level of Confidence

Choosing Between Different Levels of Confidence: Probability Versus Precision

8.6 Interval Estimation and Hypothesis Testing

Concerns About Hypothesis Testing

The Interpretation of Statistical Significance and Nonsignificance

The Influence of Sample Size on the Decision About the Null Hypothesis

Interval Estimation as an Alternative to Hypothesis Testing

8.7 Looking Ahead

8.8 Summary

8.9 Important Terms

8.10 Formulas Introduced in This Chapter

Confidence Interval for the Mean (σ Not Known)

Confidence Interval for the Mean (σ Known)

Estimated Sample Size for the Confidence Interval for the Mean

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

8.11 Using SPSS
Calculating the confidence interval for the mean (The salary survey study (8.1))
8.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 9 Testing the Difference Between Two Means
9.1 An Example From the Research: You Can Just Wait
9.2 The Sampling Distribution of the Difference
Characteristics of the Sampling Distribution of the Difference
9.3 Inferential Statistics: Testing the Difference Between Two Sample Means
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: t-Test for Independent Means
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the t-Test for Independent Means
Summary
9.4 Inferential Statistics: Testing the Difference Between Two Sample Means (Unequal Sample Sizes)
An Example From the Research: Kids' Motor Skills and Fitness
Inferential Statistics: Testing the Difference Between Two Sample Means (Unequal Sample Sizes)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Relate the Result of the Analysis to the Research Hypothesis

Assumptions of the t-Test for Independent Means (Unequal Sample Sizes)

Summary

9.5 Inferential Statistics: Testing the Difference Between Paired Means

Research Situations Appropriate for Within-Subjects Research Designs

An Example From the Research: Web-Based Family Interventions

Inferential Statistics: Testing the Difference Between Paired Means

Calculate the Difference Between the Paired Scores

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Draw a Conclusion From the Analysis

Relate the Result of the Analysis to the Research Hypothesis

Assumptions of the t-Test for Dependent Means

Summary

9.6 Looking Ahead

9.7 Summary

9.8 Important Terms

9.9 Formulas Introduced in This Chapter

Degrees of Freedom (df), Difference Between Two Sample Means

t-Test for Independent Means

Standard Error of the Difference $\left(s_{\overline{X}_1-\overline{X}_2}\right)$

Standard Error of the Difference (Unequal Sample Sizes)

Degrees of Freedom (df), Difference Between Paired Means

t-Test for Dependent Means

Standard Error of Difference Scores (s D)

9.10 Using SPSS

Testing the Difference Between Two Sample Means: The Parking Lot Study (9.1)

Testing the Difference Between Paired Means: The Web-Based Intervention Study (9.5)

9.11 Exercises

Answers to Learning Checks

Learning Check 1

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Answers to Odd-Numbered Exercises
Chapter 10 Errors in Hypothesis Testing, Statistical Power, and Effect Size
10.1 Hypothesis Testing vs. Criminal Trials
10.2 An Example From the Research: Truth or Consequences
Correct Decisions and Mistakes in Criminal Trials
10.3 Two Errors in Hypothesis Testing: Type I and Type II Error
Type I Error: The Risk in Rejecting the Null Hypothesis
What Is the Probability of Making a Type I Error?
What Is the Probability of Not Making a Type I Error?
Why Is Type I Error a Concern?
Type II Error: The Risk in Not Rejecting the Null Hypothesis
What Is the Probability of Making a Type II Error?
What Is the Probability of Not Making a Type II Error?
Why Is Type II Error a Concern?
Why Does Type II Error Occur?
Type I and Type II Error: Summary
10.4 Controlling Type I and Type II Error
Controlling Type I Error
Concerns About Controlling Type I Error
Controlling Type II Error
Increasing Sample Size
Raising Alpha (α)
Using a Directional Alternative Hypothesis (H1)
Increasing Between-Group Variability
Decreasing Within-Group Variability
Controlling Type I and Type II Error: Summary
10.5 Measures of Effect Size
Interpreting Inferential Statistics: The Influence of Sample Size
The Inappropriateness of “Highly Significant”

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Effect Size as Variance Accounted for

Measure of Effect Size for the Difference Between Two Sample Means: r^2

Interpreting Measures of Effect Size

Presenting Measures of Effect Size

Reasons for Calculating Measures of Effect Size

A Second Measure of Effect Size for the Difference Between Two Sample Means: Cohen's d

10.6 Looking Ahead

10.7 Summary

10.8 Important Terms

10.9 Formulas Introduced in This Chapter

Probability of Type I Error

Probability of Type II error

r Squared (r^2)

Cohen's d

10.10 Exercises

Answers to Learning Checks

Learning Check 1

Learning Check 2

Learning Check 3

Learning Check 4

Answers to Odd-Numbered Exercises

Chapter 11 One-Way Analysis of Variance (ANOVA)

11.1 An Example From the Research: It's Your Move

11.2 Introduction to Analysis of Variance (ANOVA)

Review: Testing the Difference Between the Means of Two Samples

Understanding Between-Group and Within-Group Variability

Relating Between-Group and Within-Group Variability

Introduction to Analysis of Variance (ANOVA) and the F-Ratio

Characteristics of the F-Ratio Distribution

11.3 Inferential Statistics: One-Way Analysis of Variance (ANOVA)

State the Null and Alternative Hypotheses (H_0 and H_1)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Value, and State a Decision Rule
Calculate a Statistic: F-Ratio for the One-Way ANOVA
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (R^2)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Summary
11.4 A Second Example: The Parking Lot Study Revisited
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Value, and State a Decision Rule
Calculate a Statistic: F-Ratio for the One-Way ANOVA
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (R^2)
Draw a Conclusion From the Analysis
The Relationship Between the t-Test and the F-Ratio
11.5 Analytical Comparisons Within the One-Way ANOVA
Planned Versus Unplanned Comparisons
Between-Group and Within-Group Variance in Analytical Comparisons
Conducting Planned Comparisons
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Conducting Unplanned Comparisons
Concerns Regarding Unplanned Comparisons
Methods of Controlling Familywise Error

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Controlling Familywise Error: The Tukey Test

11.6 Looking Ahead

11.7 Summary

11.8 Important Terms

11.9 Formulas Introduced in This Chapter

Degrees of Freedom for Between-Group Variance, One-Way ANOVA (df_{BG})

Degrees of Freedom for Within-Group Variance, One-Way ANOVA (df_{WG})

Between-Group Variance, One-Way ANOVA (MS_{BG})

Total Sample Mean, One-Way ANOVA $\left(\overline{X}_T\right)$

Within-Group Variance, One-Way ANOVA (MS_{WG})

F-Ratio, One-Way ANOVA (F)

Measure of Effect Size, One-Way ANOVA (R^2)

Relationship Between t-Test and F-Ratio for the One-Way ANOVA

F-Ratio, Analytical Comparison (F_{comp})

Between-Group Variance, Analytical Comparison $\left(MS_{\mathrm{BG}_{\mathrm{comp}}}\right)$

Measure of Effect Size, Analytical Comparison $\left(R^2_{\mathrm{comp}}\right)$

Critical Value, Tukey Test for Unplanned Comparisons (F_T)

11.10 Using SPSS

One-Way Analysis of Variance (ANOVA): The Chess Study (11.1)

Analytical Comparisons: Planned Comparisons—Explanation vs. Prediction (11.5)

11.11 Exercises

Answers to Learning Checks

Learning Check 2

Learning Check 3

Learning Check 4

Answers to Odd-Numbered Exercises

Chapter 12 Two-Way Analysis of Variance (ANOVA)

12.1 An Example From the Research: Vote—or Else!

12.2 Introduction to Factorial Research Designs

Testing Interaction Effects

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

An Example of Interaction Effect

An Example of No Interaction Effect

Testing Main Effects

The Relationship Between Main Effects and Interaction Effects

Factorial Designs vs. Single-Factor Designs: Advantages and Disadvantages

12.3 The Two-Factor (A × B) Research Design

Notational System for Components of the Two-Factor Research Design

Analyzing the Two-Factor Research Design: Descriptive Statistics

Notational System for Means in the Two-Factor Research Design

Examining the Descriptive Statistics

12.4 Introduction to Analysis of Variance (ANOVA) for the Two-Factor Research Design

F-Ratios in the Two-Factor Research Design

12.5 Inferential Statistics: Two-Way Analysis of Variance (ANOVA)

State the Null and Alternative Hypotheses (H_0 and H_1)

Main Effect of Factor A

Main Effect of Factor B

A × B Interaction Effect

Make Decisions About the Null Hypotheses

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Values, and State Decision Rules

Calculate Statistics: F-Ratios for the Two-Way ANOVA

Make Decisions Whether to Reject the Null Hypotheses

Determine the Levels of Significance

Calculate Measures of Effect Size (R^2)

Draw Conclusions From the Analyses

Relate the Results of the Analyses to the Research Hypothesis

12.6 Investigating a Significant A × B Interaction Effect: Analysis of Simple Effects

12.7 Looking Ahead

12.8 Summary

12.9 Important Terms

12.10 Formulas Introduced in This Chapter

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Degrees of Freedom for Main Effect of Factor A, Two-Way ANOVA (df_A)
Degrees of Freedom for Main Effect of Factor B, Two-Way ANOVA (df_B)
Degrees of Freedom for A × B Interaction Effect, Two-Way ANOVA ($df_{A\times B}$)
Degrees of Freedom for Within-Group Variance, Two-Way ANOVA (dfWG)
Between-Group Variance for Main Effect of Factor A, Two-Way ANOVA (MS_A)
Between-Group Variance for Main Effect of Factor B, Two-Way ANOVA (MS_B)
Sums of Squares for Between-Group Variance, Two-Way ANOVA (SS_{BG})
Between-Group Variance for A × B Interaction, Two-Way ANOVA ($MS_{A\times B}$)
Within-Group Variance, Two-Way ANOVA (MS_{WG})
F-Ratio for Main Effect of Factor A, Two-Way ANOVA (F_A)
F-Ratio for Main Effect of Factor B, Two-Way ANOVA (F_B)
F-Ratio for A × B Interaction, Two-Way ANOVA ($F_{A\times B}$)
Measure of Effect Size for Main Effect of Factor A, Two-Way ANOVA(R^2_A)
Measure of Effect Size for Main Effect of Factor B, Two-Way ANOVA(R^2_B)
Measure of Effect Size for A × B Interaction, Two-Way ANOVA(R^2_{AB})
12.11 Using SPSS
Two-Way Analysis of Variance (ANOVA): The Voting Message Study (12.1)
12.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 13 Correlation
13.1 An Example From the Research: It's Good for You!
13.2 Introduction to the Concept of Correlation
Describing the Relationship Between Variables
Nature of the Relationship
Direction of the Relationship
Strength of the Relationship

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Measuring the Relationship Between Variables: Correlational Statistics
Introduction to the Pearson Correlation Coefficient (r)
Calculating Correlational Statistics: The Role of Variance and Covariance
13.3 Inferential Statistics: Pearson Correlation Coefficient
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: Pearson Correlation Coefficient (r)
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (r^2)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the Pearson Correlation
A Computational Formula for the Pearson r
Represent the Covariance Between X and Y (SPXY)
Represent the Variance of Variable X (SSX)
Represent the Variance of Variable Y (SSY)
Calculate the Pearson Correlation Coefficient (r)
13.4 Correlating Two Sets of Ranks: The Spearman Rank-Order Correlation
Introduction to Ranked Variables
An Example From the Research: Do Students and Teachers Think Alike?
Inferential Statistics: The Spearman Rank-Order Correlation Coefficient
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
13.5 Correlational Statistics vs. Correlational Research
Does ANOVA = Causality?
“Correlation Is Not Causality”—What Does That Mean?
13.6 Looking Ahead

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

13.7 Summary
13.8 Important Terms
13.9 Formulas Introduced in This Chapter
Degrees of Freedom for Pearson Correlation (df)
Sum of Products (Definitional Formula) (SPXY)
Sum of Squares for the X Variable (Definitional Formula) (SSX)
Sum of Squares for the Y Variable (Definitional Formula) (SSY)
Pearson Correlation (Definitional Formula) (r)
Measure of Effect Size for the Pearson Correlation Coefficient (r^2)
Sum of Products (Computational Formula) (SPXY)
Sum of Squares for the X Variable (Computational Formula) (SSX)
Sum of Squares for the Y Variable (Computational Formula) (SSY)
Pearson Correlation (Computational Formula) (r)
Spearman Rank-Order Correlation (rs)
Measure of Effect Size for the Spearman Rank-Order Correlation $\left(r_s^2\right)$
13.10 Using SPSS
Pearson Correlation: The Picky Eater Study (13.1)
Spearman Rank-Order Correlation: The Business Student Study (13.4)
13.11 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Answers to Odd-Numbered Exercises
Chapter 14 Linear Regression and Multiple Correlation
14.1 Predicting One Variable From Another: Linear Regression
The Linear Regression Equation
Calculating the Linear Regression Equation
Calculate the Slope of the Equation (b)
Calculate the Y-Intercept of the Equation (a)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Report the Linear Regression Equation

Drawing the Linear Regression Equation

A Second Example: Put Down Your Pencils

Calculating the Linear Regression Equation

Drawing the Linear Regression Equation

14.2 Correlation With Two or More Predictors: Introduction to Multiple Correlation and Regression

The Limitations of Using Only One Predictor

Introduction to Multiple Correlation

Multiple Correlation With Two Predictors: An Example

Measuring the Relationship Between Two Predictors and a Criterion: The Multiple Correlation (R)

Testing the Multiple Correlation for Statistical Significance

Predicting a Criterion From Two Predictors: The Multiple Regression Equation

Conclusion

14.3 Looking Ahead

14.4 Summary

14.5 Important Terms

14.6 Formulas Introduced in This Chapter

Linear Regression Equation

Estimate of Slope (b)

Estimate of Y-Intercept (a)

Multiple Correlation Coefficient (R)

F-Ratio for the Multiple Correlation Coefficient (F)

Multiple Regression Equation

Regression Coefficient

Estimate of Y-Intercept (a) (Multiple Regression)

14.7 Using SPSS

Linear Regression: The Picky Eater Study (14.1)

Multiple Regression: The College Admission Example (14.2)

14.8 Exercises

Answers to Learning Checks

Learning Check 1

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 2

Answers to Odd-Numbered Exercises

Chapter 15 Chi-Square

15.1 An Example From the Research (One Categorical Variable): Are You My Type?

15.2 Introduction to the Chi-Square Statistic

Observed and Expected Frequencies: The Basis of the Chi-Square Statistic

Characteristics of the Distribution of Chi-Square Statistics

15.3 Inferential Statistic: Chi-Square Goodness-of-Fit Test

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Value, and State a Decision Rule

Calculate a Statistic: Chi-Square (χ^2)

Make a Decision Whether to Reject the Null Hypothesis

Determine the Level of Significance

Calculate a Measure of Effect Size (Cramér's φ)

Draw a Conclusion From the Analysis

Relate the Result of the Analysis to the Research Hypothesis

Assumptions Underlying the Chi-Square Statistic

Conducting the Chi-Square Goodness-of-Fit Test With Unequal Hypothesized Proportions

Calculate Expected Frequencies (f_e)

Calculate the Chi-Square Statistic (χ^2)

15.4 An Example From the Research (Two Categorical Variables): Seeing Red

15.5 Inferential Statistic: Chi-Square Test of Independence

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Value, and State a Decision Rule

Calculate a Statistic: Chi-square (χ^2)

Make a Decision Whether to Reject the Null Hypothesis

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Determine the Level of Significance
Calculate a Measure of Effect Size (Cramér's φ)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
15.6 Parametric and Nonparametric Statistical Tests
Parametric vs. Nonparametric Statistical Tests
Reasons for Using Nonparametric Statistical Tests
Examples of Nonparametric Statistical Tests
15.7 Looking Ahead
15.8 Summary
15.9 Important Terms
15.10 Formulas Introduced in This Chapter
Degrees of Freedom for the Chi-Square Goodness-of-Fit Test
Expected Frequency for the Chi-Square Goodness-of-Fit Test
Chi-Square Statistic (χ^2)
Cramér's φ for the Chi-Square Goodness-of-Fit Test
Degrees of Freedom for the Chi-Square Test of Independence
Expected Frequency for the Chi-Square Test of Independence
Cramér's φ for the Chi-Square Test of Independence
15.11 Using SPSS
Chi-Square Goodness-of-Fit Test—Equal Hypothesized Proportions: The Personality Study (15.1)
Chi-Square Goodness-of-Fit Test—Unequal Hypothesized Proportions: The Personality Study (15.3)
Chi-Square Test of Independence: The Color Preference Study (15.4)
Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Learning Check 6

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Odd-Numbered Exercises

Appendix A: Tables

Appendix B: Review of Basic Mathematics

I. Symbols

II. Terms

III. Addition, Subtraction, Multiplication, Division

A. All Positive Numbers

B. Adding All Negative Numbers

C. Adding a Combination of Positive and Negative Numbers

D. Subtracting Negative Numbers

E. Multiplying Two Negative Numbers

F. Multiplying a Combination of Positive and Negative Numbers

G. Dividing Two Negative Numbers

H. Dividing One Positive Number and One Negative Number

IV. Order of Operations

V. Fractions, Decimals, Percentages

A. Fraction

B. Decimal

C. Percentage

VI. Solving Simple Equations With One Unknown (X)

A. X Has a Value Subtracted From It (Solve for X Using Addition)

B. X Has a Value Added to It (Solve for X Using Subtraction)

C. X Is Multiplied by a Value (Solve for X Using Division)

D. X Is Divided by a Value (Solve for X Using Multiplication)

VII. Solving More Complex Equations With One Unknown (X)

A. X Is Multiplied by a Value and Then Has a Value Either Added to It or Subtracted From It (Solve for X Using Either Subtraction or Addition and Then Division)

B. X Is Divided by a Value and Then Has a Value Either Added to It or Subtracted From It (Solve for X Using Either Addition or Subtraction and Then Multiplication)

C. X Has a Value Added to It or Subtracted From It and Is Then Multiplied by a Value (Solve for X Using Division and Then Either Subtraction or Addition)

D. X Has a Value Either Added to It or Subtracted From It and Is Then Divided by a Value (Solve for X Using Multiplication and Then Either Subtraction or

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Addition)
Glossary
References
Index

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

PREFACE

A Few (?) Words About Me

When I entered college, I thought my path would take me to law school; however, taking an Intro to Psych class made me realize I enjoy the challenge of trying to understand the human mind. The college I attended, U.C. Santa Cruz, was a fairly unconventional university, but somehow in the midst of a sea of humanistic psychologists I became attracted to the empirical and methodological aspects of psychology by taking classes from instructors such as David Harrington and Dane Archer, who showed me that statistics could appeal to students if taught with a gentle, guiding approach that addresses questions relevant to students' lives. After graduating from college, I was able to get a job as a research assistant for a human resource consulting firm. Despite my lack of work experience, I was hired primarily as a result of having taken statistics and research methods courses, which taught me that learning statistics has benefits both inside and outside of the classroom.

Several years later, I started grad school at U.C. Berkeley, where two events critical to this book took place. First, serving as a teaching assistant, I found I really enjoyed helping students, particularly in statistics and research methods classes often approached with fear and suspicion. Second, I took graduate classes from Geoff Keppel, who had developed his own unique system for analyzing experimental research designs. His lectures and books showed me that statistics can be taught in a systematic way that highlights similarities rather than differences between different research situations. Geoff managed to transform something as daunting-sounding as a "3 x 2 x 4 research design" into the mathematical equivalent of playing with wooden toy alphabet blocks labeled "A," "B," and "C."

After leaving Cal, I was hired by San Jose State, where I teach an introductory statistics course aimed at students with a wide range of backgrounds, ability, and motivation. When I went to select a textbook for the course, for the first time I looked carefully at the wide range of offerings. What I found striking (and still find striking) was that the majority of books focused on providing formulas and very small sets of data designed to demonstrate how to correctly calculate the correct numbers from these formulas. Little emphasis, however, was given to what these numbers meant or implied. Given my own experiences learning and using statistics, I thought a book was needed that focusses on how statistics may be used to answer questions and test ideas within the larger research process.

Goal of this Book

This textbook, aimed primarily at students majoring in the behavioral and social sciences, includes many of the same topics and chapters as the majority of existing introductory statistics books. However, what I believe distinguishes this book is its thematically-based approach that places statistics within the research process. The primary purpose of this

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

book is to not just teach students how to calculate statistics but how to interpret the results of statistical analyses in light of a study's research hypothesis, and to communicate one's results and interpretations to a broader audience. Hopefully, this book will not only help students understand the purpose and use of statistics, but it will also give them a greater appreciation of how research studies are conceived, conducted, and communicated.

The book is intended for students with a limited knowledge of mathematics, statistics, and research design and methodology. Some common obstacles faced by students are their math anxiety, fear of conducting arithmetic calculations, and an inability to see the relevance of the material to their daily lives. To address these concerns, this books attempts to judiciously use mathematical terminology, write in a conversational manner, and utilize actual research situations rather than contrived sets of numbers.

Organization of the Book

The 15 chapters of this book may be placed into three general categories. Chapters 1 to 4 are designed to introduce students to the research process and how data that have been collected may be organized, presented, and summarized:

Chapter 1 (Introduction to Statistics) defines the concept of statistics, explains why statistics are important to learn, and describes the steps in the research process that are the framework for the rest of the book.

Chapter 2 (Examining Data: Tables and Figures) discusses the importance of examining data that have been collected and describes how tables and figures may be created to represent data.

Chapter 3 (Measures of Central Tendency) presents measures of central tendency (the mode, median, and the mean) and compares them in terms of their relative strengths and weaknesses.

Chapter 4 (Measures of Variability) presents measures of variability (range, interquartile range, variance, standard deviation, index of qualitative variation) and the concept of parameters (population variance and standard deviation).

Chapters 5 through 10 discuss the process of conducting statistical analyses to test research questions and hypotheses, as well as issues and controversies regarding this process:

Chapter 5 (Normal Distributions) introduces theoretical distributions, starting with normal distributions, before moving on to the standard normal distribution (normal curve).

Chapter 6 (Probability and Introduction to Hypothesis Testing) presents the structure for later chapters by introducing, defining, and illustrating steps in hypothesis testing.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Chapter 7 (Testing One Sample Mean) tests one sample mean using the *z* -test and *t* -test and introduces sampling distributions, the central limit theorem, and the Student *t* distribution.

Chapter 8 (Estimating the Mean of a Population) introduces interval estimation by developing the confidence interval for the mean and discusses interval estimation as an alternative to null hypothesis significance testing (NHST).

Chapter 9 (Testing the Difference Between Two Means) covers the *t* -test for independent means for both equal and unequal sample size and introduces within-subjects research designs using the *t* -test for dependent means.

Chapter 10 (Errors in Hypothesis Testing, Statistical Power, and Effect Size) discusses Type I and Type II error and introduces measures of effect size as a way of drawing appropriate conclusions about an analysis.

Chapters 11 to 15 discuss different statistical procedures used in research situations that vary in the number of independent variables as well as how the variables have been measured:

Chapter 11 (One-way Analysis of Variance (ANOVA)) discusses the one-way ANOVA, the relationship between the *t* -statistic and the F-ratio, and planned and post-hoc comparisons.

Chapter 12 (Two-way Analysis of Variance (ANOVA)) introduces factorial research designs, the calculation of F-ratios for main effects and interaction effects, and illustrates the use of simple effects analyses to test research hypotheses.

Chapter 13 (Correlation) defines and illustrates the concept of correlation, discusses the Pearson and Spearman correlation coefficients, and makes the distinction between correlational statistics and correlational research methods.

Chapter 14 (Linear Regression and Multiple Correlation) introduces the concept of regression, discusses linear regression, and works through an example of multiple regression with two predictor variables.

Chapter 15 (Chi-Square) describes the analysis of categorical variables using the chi-square goodness-of-fit test and the chi-square test of independence before ending the book with a brief introduction to nonparametric statistics.

In most cases, the chapters in this book share a uniform format and structure, centered on the research process, based on examples of actual published research studies addressing topics relevant to students' lives. For each example, the research hypothesis is clearly identified; after a brief description of how the data in the study were collected, the calculation and interpretation of both descriptive and inferential statistics are presented,

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ending with the extent to which the results of the statistical analyses support the research hypothesis and how these findings may be communicated to a broader audience.

Pedagogical Features

The chapters in this textbook contain several reoccurring features designed to enhance students' comprehension of the material. These include the following:

Periodic learning checks within chapters. Short learning checks, in the form of conceptual review questions and computational exercises, are provided throughout each chapter.

Explicit presentation of formulas and calculations. This book provides both definitional formulas to highlight their relation to the concept being measured and computational formulas to facilitate calculations. In applying these formulas, all of the steps in the calculation of statistics such as the *t* -test and F-ratio are included to help students in completing exercise and homework assignments.

Illustration of the presentation and communication of analyses. Every chapter includes a discussion and illustration of how statistical analyses may be reported as well how they may be communicated to a larger audience in both written and visual form.

Highlighted definition of terms. Every term and concept introduced in this book is boldfaced, immediately followed by an explicit definition.

End of chapter summaries and exercises. The end of each chapter contains a summary of critical points, a list of terms and formulas introduced in the chapter, and a large number of exercises, many of which refer to published research studies.

Calculation of statistics using SPSS. Many chapters include end-of-chapter content that provides an annotated presentation of how statistics may be calculated using the SPSS software. Screen shots of data files, menu selections, and output files are included and annotated to locate and identify relevant aspects of the analysis.

What's New to the 3rd Edition

I'd like to start by expressing my most heartfelt appreciation to the instructors and reviewers of the first two editions of this book as their evaluation, insights, criticisms, and suggestions were absolutely vital to the changes made in this third edition. I'd like to particularly thank the instructors who've adopted the book who were kind enough to spontaneously send me their thoughts and suggestions. It is my sincere hope that all of you see the incorporation of your input in this edition.

There were two main goals in this edition. The first goal was to make the writing as brief and concise as possible, manifested in the goal of reducing the length of each chapter by 5% to

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

10%. This goal was achieved by simplifying the language and shortening or eliminating what I, the book's adopters, and the book's reviewers felt was unnecessary detail or discussion. The second goal was to increase and update the exercises at the end of many of the chapters, including the use of recent published research as the basis of these exercises. Another goal was to update the demonstration of the use of SPSS to analyze data using the most recent version (Version 31 at the time of this book's publication).

More specifically, in terms of content, the following changes were made to these chapters:

Chapter 1 (Introduction to Statistics): The published research studies used to illustrate concepts in the text, learning checks, and exercises were added to and updated using topics relevant to students (e.g., fear of missing out (FOMO), cryptocurrency, legalization of cannabis, Gen Z, COVID, fake news).

Chapter 6 (Probability and Introduction to Hypothesis Testing): The Super Bowl example used to introduce and illustrate the process of hypothesis testing was updated.

Chapter 8 (Estimating the Mean of a Population): The main research example, which involves data from an actual national salary survey, has been updated throughout this chapter with more current salary data.

As with the first two editions, any and all questions, concerns, suggestions, and complaints you have about the book are most welcome and can be sent to me at howard.tokunaga@sjsu.edu .

A Few Tips for Students

To you, the college student about to read this book as part of taking a statistics course: "Welcome!" and "Great job!" We welcome you because you're about to embark on a semester-long journey that we hope enhances your skills and widens your perspective; we congratulate you because it's a journey not everyone is willing to take. We say to our students that by taking just one statistics course they are ahead of 95% of the population in their knowledge of statistics.

We also tell our students that our primary goal for them is not to become a statistics genius but just to become comfortable with the process of analyzing and interpreting data. In that regard, we agree with a 2012 *Forbes* magazine article entitled *The Six Classes That Will Make Any College Grad Employable* that noted "a good statistics class will help the student become comfortable pushing numbers around and avoiding false interpretations."

At the present moment, we know our encouragement may be of little comfort because you might be anxious about having to learn statistics. Some of you might be anxious about having to learn statistical concepts and formulas; others might be anxious because the research process seems pretty complicated. Talking with students who have taken our courses over the years has helped us assemble the following advice:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Master the Material Presented in the Early Chapters

Chapters 1 through 6 discuss how research is typically conducted, how data are summarized and described, and the process by which researchers conduct statistical analyses to test their ideas. It is absolutely critical for you to have a firm understanding of these chapters as they lay the groundwork for later chapters that discuss a variety of statistical procedures used by researchers to test hypotheses.

What does it mean to master this material? First, *read the chapters both before and after they're discussed in class* . Reading the chapters beforehand helps you identify anything that's unclear and have your questions addressed by your instructor; re-reading the chapters afterward helps confirm your understanding of the material. Next, *be able to define and explain key conceptspresented in these chapters* . These concepts are highlighted in bold-faced letters, and it's important to learn them when they're first introduced because they'll appear throughout the remainder of the book. Next, *do the learning checks within the chapters and the exercises at the end of the chapters* , because they include exercises and review questions you can ask yourself to assess your understanding of the material. Most importantly, *do not miss class during the early part of the semester* ! It's been our experience that students who miss critical lectures at the beginning of the term often have difficulty keeping up as the semester continues.

Review High School Algebra

If you happen to move beyond the introductory statistics course for which you're reading this book, you'll find that statistics are often calculated using computers and statistical software rather than hand calculations. Consequently, some statistics textbooks no longer include mathematical formulas but instead have students use statistical software or Microsoft Excel. However, this book takes a different approach for two main reasons. First, we believe that in order to understand the results of statistical analyses, one must understand their underlying foundation. Learning statistics via computer software can lead to a "brain-dead" approach in which students are at the mercy (rather than control) of their computers. We refer to this as the "I'm only as smart as my printout" method of learning statistics, also known as "Because this is what my computer told me."

Second, this book contains mathematical formulas and calculations for one simple reason: The statistics in this book are not hard to calculate. To assess whether you're prepared to read this book, take the following test:

1. Do you know how to add, subtract, multiply, and divide?
2. Do you know how to calculate the square or square root of a number?
3. Do you understand the order of operations and parentheses within mathematical formulas?

If your answers to these questions are "Yes," congratulations! You possess the ability to conduct *every* mathematical calculation in this book because *none* of the formulas involve geometry, trigonometry, or calculus. However, if you're unsure of your mathematical ability, we suggest you review your high school algebra. Also, there's an appendix at the end of this

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

book that includes the mathematical concepts and operations needed to conduct the statistical analyses in this book. We highly recommend you read this appendix to review and assess your mathematical skills.

Learn to Use a Statistical Calculator

Although learning the statistics in this book doesn't require the use of computers or software, you will need a hand calculator. Calculators with statistical capabilities are typically identified as "statistical" or "scientific" calculators. Although there's a wide selection of brands and models, know that the calculators most appropriate for this book cost only $10 to $15. We do *not* recommend you use or purchase a more expensive calculator! We prefer these simpler calculators because students are able to do their calculations more quickly and with fewer errors. We do suggest you purchase a calculator similar to the one your instructor uses as they will be more familiar with its features and idiosyncrasies.

In selecting a calculator, look closely at the keyboard; you'll need a calculator whose keys have labels such as $\bar{X}$, σ, and ΣX (Chapters 3 and 4 discuss what these symbols represent). When you go calculator shopping, you'll find calculators that have keys with labels such as ΣY and XY. Although calculators with these Y symbols are only slightly more expensive than calculators with just the X symbols, we do *not* recommend their purchase because these features are not needed to perform the calculations in this book. In fact, we've found students with Y calculators have a more difficult time performing simpler calculations.

Ask Questions

Conducting research is the process of asking and answering questions. Accordingly, throughout this book issues are often framed in the form of questions, which we hope encourages you to ask questions as well. The first person you should direct any questions to is yourself. As you work your way through the chapters, get into the habit of asking yourself, "What does this mean?" A good test of your level of comprehension is whether you can explain the material to yourself in a meaningful way.

You should also be sure to work with your instructor to confirm your level of understanding and clarify unanswered questions. Instructors often say to students, "The only bad question is the one you don't ask." Don't be hesitant about asking questions in class. Instructors will tell you they use one student's question to assess an entire class's level of understanding. This is what we call the "pencil test"—when one student asks a question, we look to see how many other pencils get raised, which reflects the number of students who had the same question. Asking questions enhances the learning experience for yourself, your classmates, and your instructor.

Form a Study Group

No matter how much we encourage students to ask questions in class, it seems they hesitate for fear of drawing attention to themselves. As a result, we turn to another well-

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

traveled saying: “Misery loves company.” Forming a study group with other students will help you keep up with reading assignments, confirm your comprehension of the material, clarify any confusion you have, prepare for exams, and provide social support. We do suggest that your group contain at least one person who will raise the group’s questions and concerns to the instructor during class or office hours.

Practice, Practice, Practice

The best way to learn the topics covered by this book is to obtain as much practice as possible. Throughout the chapters are examples in which the calculations have been worked out for you; in the middle and end of each chapter are a number of exercises. The answers to some of the exercises are provided at the back of this book, and your instructor has access to the answers to the other exercises. Do the end-of-chapter exercises even if your instructor does not assign them as homework in order to identify any reoccurring mistakes you make. We find that the vast majority of errors students make are simple computational errors, committed when students do the calculations too quickly.

Also note that the publisher of this book, Sage Publications, has developed a webpage for this book that contains a huge amount of supplemental tools such as videos and quizzes for each chapter. Find it by going to the Sage Publishing website **collegepublishing.sagepub.com** , searching for my name (Tokunaga), and looking for the **Resources** tab.

Enough cheerleading and preaching – let’s get started!

WELCOME!

GREAT JOB!

Additional Teaching Resources

Visit **collegepublishing.sagepub.com** , and navigate to the Resources tab on your book’s page to find the teaching materials designed to accompany this textbook. On this site you will find an array of materials that will save you time and to help you keep students engaged, including the following:

Learning management system cartridges that easily integrate with your course management system so student test results and graded assignments seamlessly flow into your gradebook

Test banks , aligned to Bloom’s Taxonomy, that provide a diverse range of test items, including multiple choice, true/false, and essay questions

Lecture notes that provide an outline and the key concepts in each chapter to aid in lecture preparation

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

PowerPoint ® slides that offer a flexible, accessible, and customizable solution for creating multimedia lectures

Tables and Figures from the book to support lecture preparation and class discussions

Sample course syllabi , including suggested models for structuring your course

Answers to even-numbered questions from the text help facilitate grading

SPSS datasets for use with exercises from the text.

Multimedia content , including original SAGE videos featuring tutorials with author Howard T. Tokunaga that bring concepts to life and appeal to diverse learners

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ACKNOWLEDGMENTS

This book has traveled a long journey, with many people playing both transient and persevering roles in its development. First and foremost, I'd like to thank my kids, Meagan and Will, for providing me perspective and inspiration amid all of the twists and turns we faced while this book was being written. I'm so very blessed to have you as my "burden." Away from home, I'm grateful to my graduate students at San Jose State—working on this book has given me much empathy for their thesis-related travails. However, they should be aware that the completion of this book ends the unspoken agreement that I not ask, "So how's your thesis?" if they don't ask, "So how's your book?" A number of friends have provided most welcome relief after long days staring at a computer screen, with particular thanks going to Ellie, Eileen, Susan, Bonnie, Ann, Lisa, and Terry. I would like to thank Charles Linsmeier and David Shirley for their many helpful comments and advice on earlier versions of this book, Julie West for her help in developing chapter exercises, and Youn-Jeong Choi and Naomi Ran for their careful examination of the chapters and exercises.

I would like to express my appreciation to everyone at SAGE who has played a role in shepherding this book to completion. First and foremost, I must express my gratitude to my editor, Vicki Knight. It has been a joy and an honor to work with someone with such a truly exceptional combination of experience, knowledge, perspective, and humor. In production I was fortunate enough to work with others on the SAGE team who made the process a pleasure: Yvonne McDuffee, editorial assistant; Katie Guarino, assistant editor; Nicole Elliott, executive marketing manager; Candice Harman, cover designer; and the copyeditor, Michelle Ponce. Furthermore, I am grateful to the reviewers who took the time to read the chapters, raise questions, and provide constructive comments and suggestions to assist me in clarifying and strengthening the manuscript. The following are the reviewers for third edition:

Erin M. Fekete, University of Indianapolis

Brooke Magnus, Boston College

Loretta Neal McGregor; Arkansas State University

Michael T. Moore, Adelphi University

Megan M. Ringel, St. Edward's University

Bettina Viereck, University of Hartford

Janice Weaver, Community College of Allegheny County

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ABOUT THE AUTHOR

Howard T. Tokunaga is professor of Psychology at San Jose State University, where he serves as coordinator of the MS Program in Industrial/Organizational (I/O) Psychology and teaches undergraduate and graduate courses in statistics, research methods, and I/O psychology. He received his bachelor's degree in psychology at U.C. Santa Cruz and his PhD in psychology at U.C. Berkeley. In addition to his teaching, he has consulted with a number of public sector and private sector organizations on a wide variety of management and human resource issues. He is author of *Moving from IBM SPSS to R and RStudio: A Statistics Companion* , and coauthor (with G. Keppel) of *Introduction to Design and Analysis* : *A Student's Handbook* . In his spare time, he enjoys the outdoors, museums, live theater, and being as physically, socially, and politically active as possible.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CHAPTER 1 INTRODUCTION TO STATISTICS

Learning Objectives

1.1 Define the concept of statistics.

1.2 Describe different reasons for learning statistics.

1.3 Describe the main steps in the research process and the tasks and concepts within each step.

In introducing this book to you, we assume you're a college student taking what is perhaps your first course in statistics to fulfill a requirement for your major or a general education requirement. If so, you may be asking yourself two questions:

What is statistics?

Why learn statistics?

The ultimate goal of this book is to help you begin to answer these two questions.

1.1 What Is Statistics?

You may not know it, but you encounter a variety of "statistics" in your day-to-day activities: the typical cost of going to college, the yearly income of the average college graduate, the average price of a car, and so on. So what exactly is "statistics"?

The Merriam-Webster dictionary defines **statistics** as *a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data* . When most people think about statistics, they focus on the "analysis" aspect of this definition—that is to say, they focus on numbers that result from analyzing data. However, statistics not only is concerned about how data are *analyzed* but also how data are *collected* and how the results of statistical analyses are *interpreted* and *communicated* . The main purpose of this book is to introduce, describe, and illustrate the role of statistics within the larger research process.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1.2 Why Learn Statistics?

There are many reasons why you should learn statistics. First, not only do you currently encounter statistics in your daily activities, but *throughout your life you are affected by the results of research and statistical analyses* . Which college you attend is based in part on test scores developed by researchers. You may have to take a personality or intelligence test to get a job. The choices of drugs and medicines available to you are based on medical research. Learning about statistics will help you become a more informed and aware consumer of research and statistical analyses that affect many aspects of your life.

A second reason for learning statistics is that *you may be required to interpret the results of statistical analyses* . Many college courses require students to read academic research journal articles. Evaluating published research is complicated by the fact that different people studying the same topic may come up with diverse or even opposing conclusions. Understanding statistics and their role in the research process will help you decide whether conclusions drawn in research articles are appropriate and justified.

Another reason for learning statistics is that *you might be required to do your own research* . College courses sometimes have students design and conduct mini-research studies; undergraduate majors might encourage students to do senior honors theses; graduate programs often require masters' theses and doctoral dissertations. Learning to collect and analyze data will help you address your own questions in an objective, systematic manner.

A final reason for learning statistics is that it may *help you in your future career* . The website Careercast.com conducts an annual survey in which they evaluate jobs on four dimensions: environmental factors, income, employment outlook, and stress. Recently, their #1 rated job was "data scientist," and the #2 rated job was "statistician." They noted, "A statistician's skill set can be used to break down and analyze large quantities of data. The demand for these skills spans a variety of industries, including marketing, banking, government, sports, retail, and even healthcare."

It's generally a good idea for students to maintain a healthy level of curiosity or even skepticism in regard to their education. However, we find that when it comes to learning statistics, students' feelings may be characterized as fear or anxiety. We understand these feelings, and we hope the benefits of learning statistics will become clear to you and help overcome any concerns you have.

1.3 Introduction to the Stages of the Research Process

Much of scientific research involves asking questions. Throughout this book, we'll examine a broad range of questions regarding human attitudes and behavior that contemporary researchers have asked and attempted to answer. The following are research questions we'll address in this chapter to introduce the stages of the research process:

> Does appealing to one's fear of missing out (FOMO) affect the intention to invest in cryptocurrencies?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Do pro-cannabis messages on social media lead to cannabis use?

Do parents differ in their awareness of their children's online activities?

Do nonbinary college students have the same experiences as cis men or cis women?

How might you try to answer questions such as these? You could base your answers on your personal beliefs, or you could adopt the answers given to you by others. But rather than relying on subjective beliefs and feelings, researchers in many academic disciplines test their ideas using the scientific method. The **scientific method** is a method of investigation that uses the objective and systematic collection and analysis of empirical data to test theories and hypotheses.

At its simplest, this book will portray the scientific method as consisting of five main steps or phases:

developing a research hypothesis to be tested,

collecting data,

analyzing the data,

drawing a conclusion regarding the research hypothesis, and

communicating the findings of the study.

Accomplishing each of these five steps requires completing a number of tasks, as shown in Figure 1.1 . Because this sequence of steps serves as the model for the research process we'll discuss throughout this book, each step is briefly introduced in the following sections. It's important to understand that the research process depicted in Figure 1.1 represents an ideal way of doing research. The "real" way, as you may discover in your own efforts, is often anything but a smooth ride but rather is filled with false starts, dead ends, and wrong turns.

Developing a Research Hypothesis to Be Tested

The initial stage—and the first step—of the research process is to develop a research hypothesis to be tested. A **research hypothesis** may be defined as a statement regarding an expected or predicted relationship between variables. A **variable** is a property or characteristic of an object, event, or person that can take on different values. One example of a variable is "U.S. state," a variable with 50 possible values (Alabama, Arkansas, etc.).

Research hypotheses are usually developed through the completion of several tasks:

identifying a question or issue to be examined,

reviewing and evaluating relevant theories and research, and

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

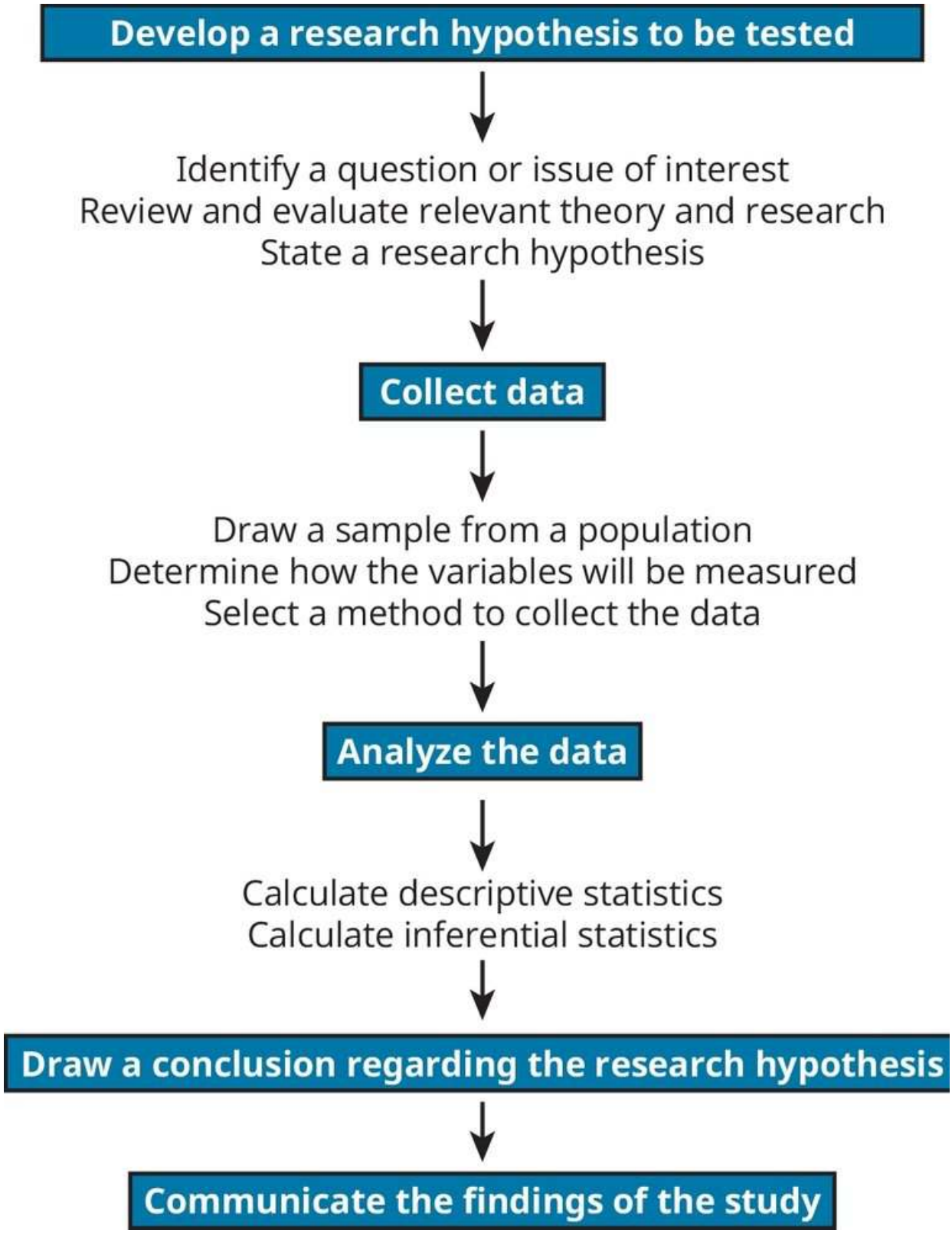
stating a research hypothesis.

Identifying a Question or Issue to Be Examined

Most research starts with a question posed by the researcher that often comes from the researcher's own ideas and daily observations. Although this may not seem terribly scientific, there's an advantage in using one's own experience as a starting point: People are generally much more motivated to explore issues that concern them personally. In teaching statistics, we advise students developing their own research projects to study something of interest to them. Conducting research can be tedious, difficult, and frustrating. At various points, you may ask yourself, "Why am I doing this?" Being able to provide a satisfactory answer to this question will help you overcome obstacles you encounter along the way.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

▶ Long Description: Figure 1.1 Steps in the Research Process Within the Scientific Method

Figure 1.1 Steps in the Research Process Within the Scientific Method

Reviewing and Evaluating Relevant Theories and Research

Beyond your own curiosity, research questions often arise from studying the theories, ideas, and research of others. A **theory** is a set of propositions used to describe or explain a phenomenon. The purpose of a theory is to summarize and explain specific facts using a logically consistent framework. Placing a question within a theoretical framework provides guidance and structure to research.

Reviewing and evaluating existing theories and research helps you decide whether it's worth the time and energy required to conduct a study. By seeing what others have done, you may conclude that the current way of thinking on a topic is incomplete or mistaken. By doing this review, you're able to ensure that the research you conduct adds to and improves upon an existing body of knowledge.

Stating a Research Hypothesis: Independent and Dependent Variables

Understanding and evaluating an existing literature not only helps develop a question of interest, but it may also lead to a predicted answer to that question. Within the scientific method, this answer is stated as a research hypothesis, defined earlier as a statement regarding an expected or predicted relationship between variables. Table 1.1 lists the research questions and research hypotheses for the studies listed at the beginning of this section. For example, the first research hypothesis states that "consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos."

One characteristic of research hypotheses such as those listed in Table 1.1 is that they identify the variables that are the focus of their research. As mentioned earlier, a variable is a property or characteristic with different values. Variables can be classified in several ways. In specifying a research hypothesis, researchers often speak in terms of "independent" and "dependent" variables. An **independent variable** may be defined as a variable manipulated or controlled by the researcher. A **dependent variable** , on the other hand, is a variable measured by the researcher that is expected to change or vary as a function of the independent variable. The independent variable is seen as the "cause" or "treatment," whereas the dependent variable is the "effect" or "outcome."

Researchers are interested in examining the effect of the independent variable on the dependent variable. Consider the first research hypothesis provided in Table 1.1 : "Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos." Here, the independent variable is the type of appeal consumers are exposed to: FOMO or non-FOMO. The dependent variable is the intention to invest in cryptos. In this study, the effect of the independent variable on the dependent variable is that

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

differences in the intention to invest in cryptos (the dependent variable) may “depend” upon which type of appeal (FOMO or non-FOMO) a consumer receives. Table 1.2 lists the independent and dependent variables for each of the research hypotheses in Table 1.1 .

In addition to identifying variables, a second characteristic of research hypotheses is that they specify the nature and direction of the relationship between variables. For example, the first research hypothesis in Table 1.1 states that “Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *higher* intention to invest in cryptos.” The word *higher* indicates the nature and direction of the relationship between the independent variable (type of appeal) and the dependent variable (intention to invest in cryptos). The direction of the relationship would not have been stated if the hypothesis had included the less specific phrase “will have a *different* intention,” which simply indicates that the two groups of consumers are not expected to be the same. Table 1.3 provides directional and nondirectional research hypotheses for the research studies from Table 1.1 .

Table 1.1 Examples of Research Questions and Hypotheses

Research Question	Research Hypothesis
Does appealing to one’s fear of missing out (FOMO) affect the intention to invest in cryptocurrencies? (Friederich et at., 2024)	Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.
Do pro-cannabis messages on social media lead to cannabis use? (Willoughby et al., 2024)	Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.
Do parents differ in their awareness of their children’s online activities? (Symons et al., 2017)	Mothers have more accurate knowledge of their children’s online activities than do fathers.
Do nonbinary college students have the same experiences as cis men or cis women? (Marx et al., 2024)	Nonbinary college students experience higher levels of victimization than do cis men or cis women.

The ability to form a directional research hypothesis is dependent on the state of the literature on the question of interest. If little or perhaps conflicting research has been conducted, researchers may not be able to form a directional hypothesis. For example, one study examining the relationship between exercise deprivation (not allowing people to exercise) and depression. reported, “We did not have a directional hypothesis, and when participants asked what we expected to find in this study, we replied: ‘We really are not sure since the results of earlier work on exercise deprivation are mixed’” (Mondin et al., 1996, p. 1200).

Table 1.2 Research Hypotheses and Their Independent and Dependent Variables

Research Hypothesis	Independent Variable	Dependent Variable
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher	Type of appeal	Intention to invest in cryptos

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Independent Variable	Dependent Variable
intention to invest in cryptos.		
Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.	Exposure to pro-cannabis messages	Intention to use cannabis
Mothers have more accurate knowledge of their children's online activities than do fathers.	Parent	Knowledge of children's online activities
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	Gender identity	Amount of victimization

Collecting Data

Once a research hypothesis has been formulated, researchers are ready to proceed to the second stage in the research process: collecting data. This step is seen as being composed of three tasks:

drawing a sample from a population,

determining how the variables will be measured, and

selecting a method by which to collect the data.

Table 1.3 Directional and Nondirectional Research Hypotheses

Directional Research Hypothesis	Nondirectional Research Hypothesis
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *higher* intention to invest in cryptos.	Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *different* intention to invest in cryptos.
Exposure to pro-cannabis messages on social media is related to *increased* intentions to use cannabis.	Exposure to pro-cannabis messages on social media is *related to* intentions to use cannabis.
Mothers have *more* accurate knowledge of their children's online activities than do fathers.	Mothers have *different* knowledge of their children's online activities than do fathers.
Nonbinary college students experience *higher* levels of victimization than do cis men or cis women.	Nonbinary college students experience *different* levels of victimization than do cis men or cis women.

Drawing a Sample From a Population

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The first step in collecting data is to identify the group of participants to which the research hypothesis applies. The group to which the results of a study may be applied or generalized is called a population. A **population** is the total number of possible units or elements that could potentially be included in a study. Although people often think the word "population" refers to everyone in the world, researchers typically define their populations more narrowly. For example, the population for the first research hypothesis in Table 1.1 could be "consumers," "U.S. consumers," "Georgia consumers," or "University of Georgia consumers." Researchers typically try to define their populations as broadly as possibly (e.g., "consumers" rather than "University of Georgia consumers") to maximize the applications or implications of their research.

Learning Check 1

Reviewing What You've Learned So Far



1. Review questions
    a. What are the five steps involved in the research process?
    b. Why is it useful to review and evaluate theories and research before conducting a study?
    c. What are the two main characteristics of research hypotheses?
    d. What is the difference between an independent variable and a dependent variable?
2. The following are several research hypotheses from published studies. For each research hypothesis, identify the independent and dependent variable.
    a. "College students will rate instructors who dress formally (i.e., business suit and tie) as having more expertise than instructors who dress casually (i.e., slacks and shirt)" (Sebastian & Bristow, 2008, p. 197).
    b. "It was expected tha. . . . greater amounts of television viewin. . . . would predict greate. . . . posttraumatic stress symptoms" (McLeish & Del Ben, 2008, p. 417).
    c. "Do borrowers who did not complete college have greater financial distress compared with . . . non-borrowers who did not complete college [or] borrowers who earned a degree?" (Jabbari et al., 2023, p. 6).
    d. "We expected that drivers would believe that texting while driving is more detrimental to their driving performance than browsing social media" (Hashash et al., 2019, p. 70).

It's extremely difficult to collect data from all members of a population. Imagine, for example, the time and money needed to collect information from every consumer in the United States! For this reason, researchers typically draw conclusions about populations based on information collected from a sample of the population. A **sample** is a subset or portion of a

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

population. Table 1.4 describes the samples used in the four studies introduced in Table 1.1 . As you can see from Table 1.4 , samples vary greatly in terms of their targeted population and size (the number of participants). The extent to which the results of a study can be generalized to the target population depends on the extent to which the sample is representative of the population.

Determining How Variables Will Be Measured: Levels of Measurement

The research hypotheses described in Table 1.1 involve variables such as type of appeal, intentions to use cannabis, and gender identity. To conduct a research study, the researcher must decide how to measure the variables stated in the research hypothesis. **Measurement** is the assignment of numbers or categories to objects or events according to rules.

For example, to measure the variable “height,” a researcher might use “number of inches from the ground in bare feet” as a form of measurement. To measure “success in college,” a student’s grade point average (GPA) may be obtained from school transcripts. “Self-esteem” might be measured by having people complete a questionnaire and be categorized as having either “low” or “high” self-esteem. As these examples demonstrate, different types of variables have different forms of measurement. Consequently, researchers have identified four distinct levels for measuring variables: nominal, ordinal, interval, and ratio.

The values of variables measured at the **nominal level of measurement** differ in category or type. The word *nominal* implies having to do with “names,” such that we use first and last names to distinguish between people. Gender identity is an example of a nominal variable, in that it consists of categories or types (such as cis male and cis female) rather than numeric values. In the first research hypothesis in Table 1.1 , the independent variable of “type of appeal” is measured at the nominal level with two categories: FOMO and non-FOMO.

Variables measured at the **ordinal level of measurement** have values that can be placed in an order relative to the other values. Rankings (such as finishing first, second, or third in a race), size (small, medium, large, or extra-large), and ratings (below average, average, or above average) are examples of ordinal variables. The values of ordinal variables indicate that one value represents relatively more or less of the variable than other values, but they don’t specify the precise difference between values. For example, although you can say that a “medium” drink is smaller than a “large” drink which is smaller than an “extra-large” drink, you can’t specify the exact difference in size between the three drinks.

Table 1.4 Research Hypotheses and Study Samples

Research Hypothesis	Study Sample
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.	“We recruited 150 participants from Amazon MTurk” (Friederich et at., 2024, p. 105).
Exposure to pro-cannabis messages on social media is	“Teens (N = 350) . . . needed to be between 13 and 17 years old and from Washington state” (Willoughby et

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Study Sample
related to increased intentions to use cannabis.	al., 2024, p. 185).
Mothers have more accurate knowledge of their children's online activities than do fathers.	"triadic data including reports from children aged 13 to 18, mothers and fathers . . . recruited in Flanders, the northern, Dutch speaking region of Belgium . . . a total of 357 valid triads was achieved" (Symons et al., 2017, p. 405).
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	"undergraduate college students aged 18–30 who were enrolled at 4-year institutions. This yielded a sample of 46,510 students" (Marx et al., 2024, p. 334).

The values of variables measured at the **interval level of measurement** are numbers equally spaced along a continuum. One example of an interval variable is the Fahrenheit scale of temperature. Here, a difference of 5 degrees has the same meaning anywhere along the scale; for example, the difference between 45°F and 50°F is the same as the difference between 65°F and 70°F. Many variables studied in the behavioral sciences (e.g., personality characteristics or attitudes) are considered to be measured at the interval level of measurement.

Variables at the **ratio level of measurement** are identical to interval variables, with one exception: Ratio scales possess what is known as a true zero point, for which the value of zero (0) represents the complete absence of the variable. Variables that describe a physical dimension (such as height, weight, distance, and time duration) typically have a true zero point. In the last research hypothesis in Table 1.1 , the variable "number of times victimized in last 12 months" is a ratio variable because it has a true zero point, where zero would indicate the complete absence of victimizations.

One advantage of ratio measurement versus interval is that ratio variables allow for a greater number of comparisons among values. Consider, for example, the value 6 for the ratio variable "inches." Not only is the difference between 4 and 6 inches the same as the difference between 6 and 8 inches (thereby involving addition and subtraction, which also can be used with interval variables), 6 inches is twice as long as 3 inches and half as long as 12 inches (involving multiplication and division). Multiplication and division comparisons cannot be made with interval variables. You cannot, for example, say that a temperature of 90°F is three times as much temperature as 30°F.

The boldfaced and italicized text in Table 1.5 illustrates how the four studies described in the earlier tables measured their independent and dependent variables. As you can see, the variables in these studies have been measured in a variety of ways at different levels of measurement.

Table 1.5 Research Hypotheses and Measurement of Variables

Research Hypothesis	Independent Variable	Dependent Variable

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Independent Variable	Dependent Variable
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.	Type of appeal **FOMO, non-FOMO (nominal)**	Intention to invest in cryptos **Likelihood of investing** **1-7 scale (1 = Highly unlikely to 7 = Highly likely) (ordinal)**
Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.	Exposure to pro-cannabis messages on social media **How often saw pro-cannabis messages (ratio)**	Intentions to use cannabis **Likelihood of using cannabis in the future (ratio)**
Mothers have more accurate knowledge of their children's online activities than do fathers.	Type of parent **Mother, Father (nominal)**	Accuracy of online activities **Accurate, somewhat accurate, inaccurate (nominal)**
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	Gender identity **Nonbinary, cis man, cis woman** **(nominal)**	Number of times victimized in last 12 months **(ratio)**

What difference does the level of measurement make?

A variable's level of measurement has important implications for researchers in that it influences how research hypotheses are stated and how data are analyzed. For example, imagine you study the variable "success in college." You could measure this variable at the *nominal* level of measurement by classifying students into one of two groups: successful or unsuccessful. Researchers typically use nominal variables when they are interested in studying differences between groups, such as, "Do successful and unsuccessful students differ in their study habits?"

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

You could instead measure success in college as an *ordinal* variable by ranking students from top to bottom, which enables a researcher to study relative differences with less concern for the precise size of these differences. For example, using ranked data, you could ask, “How similar are younger and older faculty members’ rankings of their students?”

To study success in college at the *interval* level, you could rate each student on a scale from 1 (low) to 100 (high). You could then use these ratings to address a research question such as, “Is there a relationship between faculty ratings of their students and students’ ratings of themselves?”

Finally, to measure success in college using the *ratio* level of measurement, you could use students’ salary after graduation (in dollars) as an indicator of success. Salary is a ratio variable because it contains a true zero point. In practice, ratio variables can be used to ask many of the same questions as those involving interval variables. For example, you could ask, “Is there a relationship between faculty members’ ratings of students and students’ salaries after graduation?”

As you see, how researchers measure their variables influences how they state their research questions and hypotheses. The level of measurement has another important implication: It helps determine the statistics researchers use to analyze their data. Certain statistics can be applied to variables measured at the interval or ratio levels of measurement, whereas others are appropriate for nominal or ordinal variables. As we move from chapter to chapter of this book, the level of measurement of the variables being analyzed will be noted.

Selecting a Method to Collect the Data: Experimental and Nonexperimental Research Methods

In addition to drawing a sample from a population and determining how the variables will be measured, the third step in collecting data is to determine the type of research method to use to collect the data on these variables. Research methods may be classified into two main types:

experimental research methods, and

nonexperimental research methods.

Experimental research methods.

Experimental research methods are methods designed to test causal relationships between variables—more specifically, whether changes in independent variables produce or cause changes in dependent variables. To make inferences about cause-effect relationships, researchers conducting an experiment must first eliminate other possible causes or explanations for changes in the dependent variable besides the independent variable. If it can be claimed that a variable other than the independent variable created the changes in the dependent variable, the research results are said to be confounded. A **confounding variable** is a variable related to an independent variable that provides an alternative explanation for the relationship between the independent and dependent variables.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

To illustrate the concept of confounding variables, consider the following question: Is there a relationship between a mother's ethnicity and the birth weight of her baby? Although research has shown differences in the birth weights of babies of different ethnicities, one study noted that some of this research did not eliminate the possibility that these differences may be partly due to differences between ethnic groups on factors associated with low birth weights, such as the mother's average age at the time of pregnancy and behaviors such as smoking and drinking (Ma, 2008). These other factors are considered confounding variables in that they provide alternative explanations for the relationship between ethnicity and babies' birth weights. Confounding variables make it difficult for researchers to draw firm conclusions regarding the relationship between the independent and dependent variable.

Learning Check 2

Reviewing What You've Learned So Far



1. Review questions
    a. What is the difference between a population and a sample?
    b. Within the research process, what is the relationship between populations and samples?
    c. What are the four levels of measurement? How do they differ?
2. For each of the following variables, name the scale of measurement (nominal, ordinal, interval, or ratio).
    a. Type of school attended (public, private)
    b. Probability of graduating college in 4 years (0% to 100%)
    c. Rating of television (good, better, best)
    d. Number of computers in home

Researchers minimize the influence of confounding variables in two main ways. First, researchers exercise experimental control by making the research setting (i.e., characteristics of the research participants, the instruments or measures administered, the methods used to collect data) the same for all participants. In the birth weight study, for example, the research design might control for the effect of parental smoking on birth weight by only including nonsmokers in the study sample, thereby excluding those who smoke.

Another way researchers exert control over the research setting is by including a condition known as a **control group** , a group of participants in an experiment not exposed to the independent variable. For example, if you conducted an experiment examining the effects of caffeine on one's health, one group (the experimental group) might be instructed to drink coffee while a second group (the control group) drinks decaffeinated coffee. By comparing the results of the two groups, researchers can then assess the impact of caffeinated coffee on the health of those who drink it.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Because researchers cannot possibly identify and control for all potential confounding variables, a second strategy used to minimize the influence of confounding variables is **random assignment** , which is assigning participants to each category of an independent variable in such a way that each participant has an equal chance of being assigned to each category. For example, a researcher could randomly assign a participant to the experimental or control condition by flipping a coin: heads for experimental, tails for control. The purpose of random assignment is to equalize or neutralize the effects of confounding variables by distributing them equally over the levels of the independent variable.

Nonexperimental (correlational) research methods.

Experimental research designs are one of the best tools researchers have for making causal inferences. However, the ability to make causal inferences requires a great deal of control that may not always be possible or desirable. For this reason, researchers often employ **nonexperimental research methods** , which are research methods designed to measure relationships between variables without having the ability to infer cause-effect relationships. These research methods are sometimes referred to as *correlational* research methods. It is important to understand that correlational research methods are different from correlational statistics we'll introduce in Chapter 13. Correlational research methods pertain to how data are *collected* ; correlational statistics pertain to how data are *analyzed* . Some of the most common types of nonexperimental research designs include quasi-experiments, survey research, observational research, and archival research.

Quasi-experimental research compares naturally formed or preexisting groups rather than randomly assigning participants to conditions. One example of this type of research was a study examining the effects of the COVID-19 pandemic on three groups of students: those living on campus, those living off campus but within commuting distance to school, and those living far away in another city, state, or country (Bisconer & McGill, 2023). This study is a quasi-experiment because the students were not randomly assigned to one of the three living conditions before collecting data.

Survey research methods obtain information directly from a group of people regarding their opinions, beliefs, or behavior. The goal of survey research, which can include using questionnaires and interviews, is to obtain information from a sample to represent or estimate the views of a larger population. One example of survey research examined college students' mental health by having them fill out a survey in which they reported their spirituality (the desire for truth, meaning, and significance of one's life) and their self-acceptance (knowledge and acceptance of one's self and limitations) (Nonis et al., 2024). Survey research is nonexperimental in that the variables being measured are not manipulated by the researcher.

Observational research is the systematic and objective observation of naturally occurring behavior or events. Observational research is often used to study phenomena that the researcher cannot or should not deliberately manipulate. For example, researchers studying interpersonal conflict among children would never force children to fight each other. One study resolved this issue by videotaping groups of adolescents discussing topics such as competition and peer pressure and then counting the number of times the adolescents teased or mocked each other (Connolly et al., 2015).

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Rather than observing or measuring behavior directly, **archival research** is the use of archives (records or documents of the activities of individuals, groups, or organizations) to examine research questions or hypotheses. One example of an archival research study studied the question of whether men are perceived to be more creative than women (Proudfoot et al., 2015). In this study, the 100 most viewed talks on TED.com were identified, and the researchers compared viewers' posted ratings of the talks given by men versus women.

Labeling variables in nonexperimental research.

One issue that often confuses students when learning about nonexperimental research is the practice of referring to the variables as the "independent variable" and the "dependent variable." For example, in the study described previously that involved biological sex and TED.com ratings, biological sex may be called the "independent variable." However, because biological sex isn't manipulated by the researcher, it would be inappropriate to say that biological sex "causes" differences in TED.com ratings. To avoid confusion, researchers sometimes refer to the independent variable and the dependent variable in a nonexperimental study as the **predictor variable** and **criterion variable**, respectively.

Analyzing the Data

Once data collection has been completed, the next step in the research process involves data analysis. This part of the research process addresses the primary topic of this book: statistics. Because this is the first chapter of this book, we won't describe analyzing data in detail. Instead, we introduce the notion that there are two main purposes of analyzing data: (1) to organize, summarize, and describe data that has been collected and (2) to test and draw conclusions about ideas and hypotheses. These two purposes are met by calculating two main types of statistics: descriptive statistics and inferential statistics.

Calculating Descriptive Statistics

Descriptive statistics are statistics used to summarize and describe a set of data for a variable. For example, you've probably heard of statistics used to describe or summarize aspects of our society, such as crime statistics and unemployment statistics. Using a research-related example, Caitlin Abar, a researcher at Brown University, conducted a study examining increases in students' alcohol-related behaviors after entering college (Abar, 2012). To measure students' level of alcohol use, she created a variable called "typical weekend drinking," which "was measured as the sum of drinks consumed on a typical Friday and Saturday within the past 30 days" (p. 22). One way to summarize students' responses to the typical weekend drinking variable (0, 1, 2, etc.) would be to calculate the mean, which is the mathematical average of a set of scores. The mean, described in Chapter 3, is an example of a descriptive statistic. The first part of this book describes a variety of descriptive statistics used by researchers to summarize data they've collected.

Calculating Inferential Statistics

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Besides summarizing and describing data, a second purpose of statistics is to test hypotheses. **Inferential statistics** are statistical procedures used to test hypotheses and draw conclusions from data collected during research studies. Using inferential statistics, researchers make inferences about what's believed to exist in populations based on data gathered from samples of the population.

As an example of an inferential statistic, let's return to the example of students' alcohol-related behaviors (Abar, 2012). This study was interested in seeing whether these behaviors were related to aspects of these students' relationships with their parents. One aspect was "alcohol communications," defined as "how often they discussed alcohol related topics with their parents" (p. 22). To test a hypothesis regarding the relationship between students' alcohol use and their alcohol communications, we could use an inferential statistic known as the Pearson correlation coefficient, which will be discussed in Chapter 13. The chapters in the last half of this book discuss a broad variety of inferential statistics, along with detailed instructions for analyzing and drawing conclusions from statistical data and presenting research findings.

Drawing a Conclusion Regarding the Research Hypothesis

Once statistical analyses have been completed, the next step is to interpret the results of the analyses as they relate to the study's research hypothesis. More specifically, do the findings of the analyses support or not support the research hypothesis? The word *support* in the previous sentence is very important. Students and researchers are sometimes tempted to say their findings "prove" their hypotheses are either "true" or "false." However, as we'll discuss in Chapter 5, because researchers typically don't collect data from the entire population, they can't know with 100% certainty whether their hypothesis has been proven to be true or false. A *huge* part of this book is learning how to draw appropriate conclusions from statistical analyses.

Communicating the Findings of the Study

Conducting research requires a variety of different skills: conceptual skills to develop research hypotheses, methodological skills to collect data, and mathematical skills to analyze these data. Another integral part of the research process is the communication skills needed to inform others about a study. Researchers must be not only skilled scientists but also effective writers.

Learning Check 3

Reviewing What You've Learned So Far



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1. Review questions
    a. What are the main differences between experimental and nonexperimental research methods?
    b. What are confounding variables? What do researchers do to minimize the effects of confounding variables?
    c. What are the main types of nonexperimental research methods?
    d. What are the relative strengths and weaknesses of experimental and nonexperimental research?
    e. What are the main purposes of descriptive and inferential statistics?

For many years, the American Psychological Association (APA), the professional association for psychologists, has provided researchers guidance on how to communicate the results of their research. In 1929, the APA published a seven-page article entitled, “Instructions in Regard to Preparation of Manuscript.” In contrast, the seventh edition of the *Publication Manual of the American Psychological Association* , published in 2020, consists of 428 pages! This increase in length highlights the challenge of communicating research effectively and appropriately.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1.4 Plan of the Book

The primary purpose of this chapter was to introduce statistics and place it within the larger research process. The remainder of this book discusses different statistics used by researchers to examine various questions of interest. Although it's critical for you to learn to correctly calculate statistics, it is equally important to understand the role of statistics within the research process and appreciate issues related to the use (and sometimes misuse) of statistics.

The first half of the book (Chapters 2 through 6) introduces conceptual and mathematical issues that are the foundation of statistical analyses. Chapters 2, 3, and 4 discuss how data may be examined, described, summarized, and presented in numeric, visual, and graphic form. Descriptive statistics are introduced and discussed in these chapters. Chapters 5 and 6 introduce characteristics of the inferential statistics used to test research hypotheses. The chapters in the second half of the book cover a number of different inferential statistics used with different types of variables.

In most cases, the chapters in this book share a uniform format and structure and are centered on the research process, using examples of actual published research studies that address questions with which you may be familiar. For each example, we'll clearly state the research hypothesis, briefly describe how the data in the study were collected, illustrate the calculation and interpretation of descriptive and inferential statistics, discuss the extent to which the statistical analyses support the research hypothesis, and show how to communicate one's research to a broader audience.

1.5 Looking Ahead

We began this chapter by defining statistics and providing reasons why learning about them may be of value to you. Next, we placed statistics within the stages of the larger research process, a process that will be the foundation of this textbook. As we've explained, the research process is centered on a research hypothesis, which is a predicted relationship between variables. Once a hypothesis has been stated, the next step is to collect data to be analyzed. However, before conducting statistical analyses on a set of data, there is preliminary work that must be completed. The next chapter will focus on methods used by researchers to examine data.

1.6 Summary

Statistics may be defined as a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data. As such, statistics not only is concerned about how data are analyzed but also recognizes the importance of understanding how data are collected and how the results of analyses are interpreted and communicated. There are a variety of reasons why students should learn statistics: Statistics are encountered in a wide variety of daily activities, students may be asked or required to read and interpret the results of statistical analyses in their courses, students may use statistics in conducting their own research, and statistics may help one's career.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Researchers conduct research using the **scientific method** of inquiry, a method of investigation that uses the objective and systematic collection and analysis of empirical data to test theories and hypotheses. The research process used within the scientific method of inquiry consists of five main steps: developing a research hypothesis to be tested, collecting data, analyzing the data, drawing a conclusion regarding the research hypothesis, and communicating the findings to the study.

A **research hypothesis** is a statement regarding an expected or predicted relationship between variables. Developing a research hypothesis involves identifying a question or issue to be examined, reviewing and evaluating relevant theories and research, and stating the research hypothesis to be tested in the study. A **theory** is a set of propositions used to describe or explain a phenomenon. A research hypothesis contains **variables** , which are properties or characteristics of some object, event, or person that can take on different values. More specifically, a research hypothesis states the nature and direction of a proposed relationship between an **independent variable** (a variable manipulated or controlled by the researcher) and a **dependent variable** (a variable measured by the researcher that is expected to change or vary as a function of the independent variable).

Collecting data involves drawing a sample from a population, determining how the variables will be measured, and selecting a method to collect the data. A **population** is the total number of possible units or elements that could potentially be included in a study; a **sample** is a subset or portion of a population. Variables can be measured at one of four levels of measurement: **nominal** (values differing in category or type), **ordinal** (values placed in an order relative to the other values), **interval** (values equally spaced along a numeric continuum), or **ratio** (values equally spaced along a numeric continuum with a true zero point).

There are two main types of methods used to collect data: **experimental research methods** , designed to test whether changes in independent variables produce or cause changes in dependent variables, and **nonexperimental (correlational) research methods** , designed to examine the relationship between variables without the ability to infer cause-effect relationships.

In experimental research, researchers are concerned about possible **confounding variables** , which are variables related to independent variables that provide an alternative explanation for the relationship between independent and dependent variables. To minimize the impact of confounding variables, a research study may exercise experimental control over various aspects of the situation, including the use of a **control group** , which is a group of participants not exposed to the independent variable the research is designed to test, or **random assignment** , which involves assigning participants to each category of an independent variable in such a way that each participant has an equal chance of being assigned to each category.

Examples of nonexperimental research methods are **survey research** , in which information is directly obtained from a group of people regarding their opinions, beliefs, or behavior; **observational research** , which involves the systematic and objective observation of naturally occurring events; and **archival research** , which uses archives (records or documents of the activities of individuals, groups, or organizations) to examine research questions or hypotheses. In nonexperimental research, the independent variable and

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

dependent variable may be referred to as the **predictor variable** and **criterion variable**, respectively.

Once data have been collected, the next step is to analyze them using two main types of statistics: **descriptive statistics**, which summarize and describe a set of data, and **inferential statistics**, which are statistical techniques used to test hypotheses and draw conclusions from data collected during research studies.

Once the statistical analyses have been completed, the results of the analyses are interpreted regarding whether they support or not support the study's research hypothesis. The final step in the research process is to communicate the study to a broader audience.

1.7 Important Terms

archival research

confounding variable

control group

criterion variable

dependent variable

descriptive statistics

experimental research methods

independent variable

inferential statistics

interval level of measurement

level of measurement (nominal, ordinal, interval, ratio)

measurement

nonexperimental research methods

observational research

population

predictor variable

quasi-experimental research

random assignment

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

research hypothesis

sample

scientific method

statistics

survey research

theory

variable

1.8 Exercises

1. Identify the independent and dependent variable for each of the following research questions and hypotheses from published articles.
    a. “We hypothesize that finance majors will embody narcissistic traits more than other business majors” (Wood et al., 2021, p. 924).
    b. “Men are expected to employ physical characteristics (particularly those that are directly related to sex) more often than women in selecting dating candidates” (Hetsroni, 2000, p. 91).
    c. “It is hypothesized that perceptions of emotional pain and support needs will be higher for the community of the Black victim than those of the officers’” (Kahn et al., 2024).
    d. “Gen Z students are more likely to believe that ‘Group projects bring out the worst in people’ than Millennials” (Schlee et al., 2020, p. 145).
    e. “We hypothesized to find . . . higher stress and anxiety post COVID in the Portuguese group in contrast with the British ” (Penalba-Sanchez et al., 2022, p. 322).
    f. “We hypothesized that consumers would think that a ‘sale price’ presentation would generate a greater monetary savings than an ‘everyday low price’ presentation” (Tom & Ruiz, 1997, p. 403).
2. The following are a number of research questions and hypotheses from actual published articles. For each hypothesis, identify the independent and dependent variable.
    a. “The use of color in a Yellow Pages advertisement will increase the perception of quality of the products for a particular business when compared with noncolor advertisements” (Lohse & Rosen, 2001, p. 75).
    b. “We hypothesized that parents who use more frequent corporal and verbal punishmen. . . . will report more problem behaviors in their children” (Brenner & Fox, 1998, p. 252).
    c. “It was hypothesized that adolescents with anorexia nervosa woul. . . . be more respectful when compared with peers with bulimia nervosa” (Pryor & Wiederman, 1998, p. 292).
    d. “The purpose of the present research was to assess brand name recognition as a function of humor in advertisements. . . . It was predicted that participants would recognize product brand names that had been presented with humorous

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

advertisements more often than brand names presented with nonhumorous advertisements" (Berg & Lippman, 2001, p. 197).

e. "The purpose of this study was to explore the relation between time spent in daycare and the quality of exploratory behaviors in 9-month-old infant. . . . it was hypothesized tha. . . . infants who spent greater amounts of time in center-based care would demonstrate more advanced exploratory behaviors than infants who did not spend as much time in center-based care" (Schuetze et al., 1999, p. 269).

f. "It was hypothesized that students would score higher on test items in which the narrative contains topics and elements that resonate with their daily experiences versus test items comprised of material that is unfamiliar" (Erdodi, 2012, p. 172).

3. Identify the independent and dependent variable for each of the following research questions and hypotheses from published articles.
    a. "It was hypothesized that the number of selfies posted . . . would be positively correlated with dimensions of narcissism" (Barry et al., 2017, p. 51).
    b. "We hypothesized that preschoolers' self-reported preferences for aggressive media . . . would be positively related to parent reports of children's aggressive behaviors" (Jamnik & DiLalla, 2018, p. 159).
    c. "We hypothesized that helicopter parenting would lead to higher levels of depression" (Darlow et al., 2017, p. 3).
    d. "We predicted that weight-related shame would be associated with greater binge eating" (Craven & Fekete, 2019, p. 45).
    e. "We hypothesize that men with strong gender identities will be more sensitive to manipulated threats to their masculinity compared with men with gender identities that are less central to their identity" (Harrison & Michelson, 2019, p. 65).
    f. "After exposure to fake news story with user comments identifying the news as fake, people are less likely to share the fake news than after exposure to the fake news story with user comments supportive of the content" (Colliander, 2019, p. 204).
4. Name the scale of measurement (nominal, ordinal, interval, ratio) for each of the following variables:
    a. The amount of time needed to react to a sound
    b. Gender
    c. Score on the Scholastic Aptitude Test (SAT)
    d. Political orientation (not at all conservative, conservative, very conservative)
    e. Political affiliation (Democrat, Republican, Independent)
5. Name the scale of measurement (nominal, ordinal, interval, ratio) for each of the following variables:
    a. One's age (in years)
    b. Size of soft drink (small, medium, large, extra large)
    c. Voting behavior (in favor vs. against)
    d. IQ score
    e. Parent (mother vs. father)
6. A faculty member wishes to assess the relationship between students' scores on the SAT and their performance in college.
    a. What is a possible research hypothesis in this situation?
    b. What are the independent and dependent variables?
    c. How could you measure the variable "performance in college" at each of the four levels of measurement?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

7. A researcher hypothesizes that drivers who use cellular phones will get into a greater number of traffic accidents than drivers who do not use these phones. For this example,
    a. What are the independent and dependent variables?
    b. How could you measure the dependent variable?
    c. How could you conduct this study using an experimental research method? How could you conduct this study using a nonexperimental method?

Answers to Learning Checks

Learning Check 1

1.

    a. IV: Instructor's dress; DV: Expertise

    b. IV: Amount of television viewing; DV: Posttraumatic stress symptoms

    c. IV: Student group; DV: Financial distress

    d. IV: Activity while driving; DV: Driving performance

Learning Check 2

2.

    a. Nominal

    b. Ratio

    c. Ordinal

    d. Ratio

Answers to Odd-Numbered Exercises

1.

    a. IV: Type of business major; DV: Narcissistic traits

    b. IV: Biological sex; DV: Employment of physical characteristics

    c. IV: Community; DV: Perceptions of emotional pain and support needs

    d. IV: Generation; DV: Beliefs about group projects

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

e. IV: Country; DV: Stress and anxiety

f. IV: Price presentation; DV: Perception of savings

3.

a. IV: Number of selfies posted; DV: Narcissism

b. IV: Preference for aggressive media; DV: Parent reports of children's aggressive behaviors

c. IV: Helicopter parenting; DV: Depression

d. IV: Weight-related shame; DV: Binge eating

e. IV: Strength of gender identity; DV: Sensitivity to threat to masculinity

f. IV: Type of user comments to news story; DV: Sharing of news story

5.

a. Ratio

b. Ordinal

c. Nominal

d. Interval

e. Nominal

7.

a. IV: Cellular phone use; DV: Number of traffic accidents

b. Example: Determine the number of traffic accidents each person has experienced in the past year.

c. Example of experimental: Use a driving simulation with the "driver" talking on a phone and measure the number of potential driving mistakes or accidents.

d. Example of nonexperimental: Take a survey of the number of accidents people have been in and whether or not they were using a phone during the accident.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CHAPTER 2 EXAMINING DATA: TABLES AND FIGURES

Learning Objectives

2.1 Describe the main purpose and the research hypothesis of the “lottery” study.

2.2 Describe different reasons for examining data.

2.3 Describe the steps used to create a frequency distribution table.

2.4 Describe the steps used to create a grouped frequency distribution table.

2.5 Explain the differences between bar charts, pie charts, histograms, and frequency polygons.

2.6 Describe the three aspects and types of distributions of data.

Chapter 1 described a process by which research may be conducted, which begins by stating a research hypothesis. The next step involves collecting data that will ultimately be used in statistical analyses designed to test the research hypothesis. However, researchers don’t conduct analyses until they’ve first examined the data to make sure appropriate conclusions can be drawn from the analyses. This chapter describes *why* researchers examine data and *how* data may be examined using tables, charts, and graphs. As with many of the chapters in this book, this discussion will be guided by the findings from a published research study.

2.1 An Example From the Research: Winning the Lottery

A friend of yours flips a coin and asks you to guess whether it’s “heads” or “tails.” You guess “tails,” but the answer is “heads.” Your friend flips the coin a second time, and again you predict “tails.” But again the answer is “heads.” On the third coin flip, you guess “tails” once more—but, once again, you’re wrong. As your friend prepares to flip the coin a fourth time, you tell yourself there’s little chance the coin could be “heads” again. But the reality is that the fourth coin flip is equally likely to be “heads” or “tails” because the outcome of the first three coin flips has absolutely no impact on the fourth flip. In other words, the outcome of a coin flip is a random event that can’t be controlled or predetermined. But do you really believe this?

The impact of random events can take on greater significance than the outcome of a coin flip. For example, government-run lotteries provide the opportunity to win huge amounts of money through the random selection of numbers. In 2001, Dr. Karen Hardoon, a researcher at McGill University in Canada, studied the thought processes people use when purchasing lottery tickets (Hardoon et al., 2001). As previous research had found gamblers believe they can control random events such as the rolling of dice or the spinning of a roulette wheel, Dr. Hardoon and her colleagues were interested in

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

determining “whether individuals with gambling problems perceive the purchase of lottery tickets in a similar manner as non-problem gamblers” (p. 752). On the basis of earlier findings, Dr. Hardoon hypothesized that problem gamblers are less likely to believe lottery outcomes are random than are nonproblem gamblers.

To test their research hypothesis, Dr. Hardoon and her associates located a group of gamblers and asked whether they had ever experienced problems as a result of their gambling. Each participant was classified into one of two groups: problem gamblers and nonproblem gamblers.

The participants in the study, which we’ll refer to as the lottery study, were shown four lottery tickets and asked, “If you were to buy a ticket to play in the lottery, which one would you select?” Each ticket contained six numbers selected from among the numbers 1 through 49; the four tickets were given a particular label. The *sequence* ticket had numbers in consecutive order (30-31-32-33-34-35); numbers on the *pattern* ticket increased by 5s (5-10-15-20-25-30); the *nonequilibrated* numbers (35-37-40-43-44-49) were clustered at the upper end of the 49 numbers, and the *random* ticket (7-8-23-34-36-42) shared none of the characteristics of the other three tickets. We’ll call the variable in this study “lottery ticket,” a variable with four categories: sequence, pattern, nonequilibrated, and random.

The researchers recorded which of the four tickets each participant chose. Table 2.1 lists the data for the lottery ticket variable for the nonproblem gamblers (the data for the problem gamblers are provided in the exercises at the end of this chapter). Looking at this table, we see that the nonproblem gamblers differed in their lottery ticket choices; however, it’s difficult to describe these differences given the unorganized nature of the collected data. Consequently, before conducting statistical analyses on their data, researchers typically organize and examine them. In the following sections, we will discuss why and how data are organized and examined.

2.2 Why Examine Data?

There are a variety of reasons why researchers examine data they’ve collected before analyzing the data. These reasons include the following:

To gain an initial sense of the data

To identify outliers

To evaluate research methodology

To determine whether data meet statistical criteria and assumptions

Each of these reasons is discussed in the following sections.

Table 2.1 The Lottery Ticket Choices of 22 Nonproblem Gamblers

Participant	Lottery Ticket
1	**Random**
2	**Random**
3	**Nonequilibrated**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Participant	Lottery Ticket
4	**Pattern**
5	**Random**
6	**Sequence**
7	**Random**
8	**Random**
9	**Pattern**
10	**Pattern**
11	**Random**
12	**Pattern**
13	**Random**
14	**Nonequilibrated**
15	**Random**
16	**Pattern**
17	**Nonequilibrated**
18	**Random**
19	**Random**
20	**Pattern**
21	**Random**
22	**Random**

Gaining an Initial Sense of the Data

Because researchers spend a great amount of time and energy designing their research studies, they're eager to analyze the data they've collected. Students taking courses in statistics, on the other hand, face a different challenge: They're given sets of data with which they have no prior experience. Although both researchers and students may be motivated to analyze their data as quickly as possible, it's important to first examine the data in order to gain an initial sense of them.

As an example, imagine you conduct a study regarding the fuel efficiency of automobiles and collect data on the miles per gallon (MPG) of different types of cars. You might expect the scores in your data to range from 10 to 50 MPG, with the majority between 25 and 35 MPG. Examining data helps researchers gain an initial sense of whether the data are in line with their expectations. Examining data also provides a way of confirming the results of later statistical analyses. We've found that students' errors in their calculations could have been avoided had they first looked at the set of data. For example, examining data will help you avoid making statements such as, "The average MPG in my sample was 127 MPG."

Identifying Outliers

Researchers also examine their data to identify **outliers** , defined as rare, extreme scores that lie outside of the range of the majority of scores in a set of data. Imagine, for example, that a group of 15 students is asked to take a 20-item test of mathematical calculations; each student's score is the number of test items answered correctly. In scoring the tests, a researcher finds that 14 students

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

correctly answered between 8 and 17 questions. However, one student provided the wrong answer to every question for a score of 0. In this case, the score of 0 would be considered an outlier.

If not identified, outliers may distort conclusions researchers draw about their sample and their data, particularly when the sample is relatively small. In the previous example, the single score of 0 might lead an observer to conclude the class as a whole performed more poorly on the test than they actually did.

Evaluating Research Methodology

Examining data is also useful in assessing the effectiveness of the methods used to collect the data. For the 20-item math test, for example, it might be reasonable to expect the majority of scores to fall in the middle of the 0 to 20 range. But what if every student answered at least 17 of the 20 questions correctly? You might conclude the test was too easy and that more difficult questions are needed to accurately assess students' math skills. By comparing the collected data for a variable with the range of possible values, researchers are able to evaluate and modify their measurement tools.

Determining Whether Data Meet Statistical Criteria and Assumptions

Researchers also examine their data to see whether they meet certain statistical criteria and assumptions. Many of the statistics in this book are based on assumptions regarding the shape and nature of a set of data that has been collected. For example, these procedures assume that in most data sets, the majority of scores will fall in the middle of the range of possible values, with a small amount of scores at the highest and lowest possible values (you may have heard this referred to as "the bell curve"). In the example of the 20-item math test, it might be assumed that most scores will be between 7 and 13, with smaller number of scores less than 7 or greater than 13. When assumptions such as this are not met, it's possible to draw inappropriate conclusions from statistical analyses that are conducted.

2.3 Examining Data Using Tables

One part of examining data that have been collected for a variable involves answering a series of questions:

> What are the values of the variable?
>
> How many participants in the sample have each value of the variable?
>
> What percentage of the sample has each value of the variable?

As it may be difficult to answer questions such as these just by looking at the data in Table 2.1 , researchers typically organize their data by creating a table known as a **frequency distribution table** , a table that summarizes the number and percentage of participants for the different values of a variable. This section discusses the steps involved in creating frequency distribution tables using these questions.

What Are the Values of the Variable?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The first step in organizing the data for a variable is to identify all of the possible values for the variable. Table 2.2 begins the construction of a frequency distribution table by identifying the four values for the lottery ticket variable: sequence, pattern, nonequilibrated, and random.

How Many Participants in the Sample Have Each Value of the Variable?

Once the values of the variable have been identified, the next step in organizing a set of data is to determine the **frequency** of each value, which is the number of participants in a sample corresponding to a value of a variable. These frequencies help indicate the nature and shape of the distribution of scores for a variable in a sample.

Table 2.2 Identifying the Values of the Lottery Ticket Variable

Number	Lottery Ticket
1	Sequence
2	Pattern
3	Nonequilibrated
4	Random

Table 2.3 Using the Tick Mark Method to Determine the Frequencies of the Lottery Ticket Variable

Lottery Ticket	Tick Marks	*f*
Sequence	|	1
Pattern	||||— |	6
Nonequilibrated	|||	3
Random	||||— ||||— ||	12
Total	||||— ||||— ||||—||||—||	22

One way to determine the frequencies for a variable is to sort the data into the different values and then count the number of participants having each value using a “tick mark method,” where a tick mark is placed next to a value of the variable each time that value appears in the sample. As shown in Table 2.3 , the tick marks are then counted to determine the frequency for each value. For example, the number of tick marks and therefore the frequency (f) for the sequence ticket is 1 because it was picked by one participant (Participant 6 in Table 2.1). Similarly, the tick marks for each of the other three types of tickets can be recorded and counted until all of the data in the sample are accounted for.

What Percentage of the Sample Has Each Value of the Variable?

Although it’s important to determine the frequency for each value of a variable, these frequencies by themselves are not always particularly informative. For example, how you interpret “There are 35 Democrats in my sample” is different when the sample consists of 40 voters rather than 400 voters.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Because the interpretation of frequencies depends on the size of the sample, it's useful to calculate the *percentage* of the sample having each value of a variable. These percentages can be calculated using the following formula:

$$\% = \frac{f}{total\,number\,of\,scores} * 100$$

where *f* is the frequency for a value of a variable.

Looking at Table 2.3 , 1 of the 22 nonproblem gamblers chose the sequence ticket. Therefore, the percentage of the sample choosing this ticket is

$$\begin{aligned} \% &= \frac{f}{total\,number\,of\,scores} * 100 \\ &= \frac{1}{22} * 100 = .05 * 100 \\ &= 5\% \end{aligned}$$

After calculating the percentage for each of the four tickets, Table 2.4 provides the frequency distribution table for the lottery ticket variable.

Once a frequency distribution table has been constructed for a variable, researchers can begin to draw conclusions about their sample. For the lottery ticket variable, the random ticket had the highest frequency and percentage (*f* = 12, 54%) compared to the pattern (*f* = 6, 27%), nonequilibrated (*f* = 3, 14%), and sequence (*f* = 1, 5%) tickets. As a result, the researchers in the lottery study made the following observation:

"The results of the present study indicated that, for the entire sample . . . the greatest percentage of tickets selected by the entire sampl. . . . were random tickets." (Hardoon et al., 2001, p. 760)

Table 2.4 Frequency Distribution Table for the Lottery Ticket Variable

Lottery Ticket	***f***	**%**
Sequence	1	**5%**
Pattern	6	**27%**
Nonequilibrated	3	**14%**
Random	12	**54%**
Total	22	**100%**

Learning Check 1

Reviewing What You've Learned So Far

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



1. Review questions
    a. What are some reasons for examining data before conducting statistical analyses?
    b. Why are outliers of concern to researchers?
    c. What questions do you address about a set of data for a variable in constructing frequency distribution tables?
    d. Why is it useful to calculate percentages in addition to frequencies for each value of a variable?
2. For each of the following situations, create a frequency distribution table.
    a. A college counselor asks a group of 116 seniors what their plans are after graduating from college; 71 say they are going to “work,” 34 are planning on going to “graduate school,” and 11 say they are “not sure” what their plans are.
    b. A high school instructor teaching a course finds that of the students she teaches, 69 are Freshmen, 18 are Sophomores, 12 are Juniors, and 9 are Seniors.
    c. A pollster stops 12 people and asks if they are either “in favor” or “against” a local proposition in Table 2.5 :

Table 2.5 Example of People’s Position

Person	Position
1	**In favor**
2	**In favor**
3	**Against**
4	**In favor**
5	**Against**
6	**Against**
7	**In favor**
8	**Against**
9	**In favor**
10	**Against**
11	**In favor**
12	**In favor**

    d. The company that makes M&Ms candy (www.mms.com) conducted a survey asking people which new color they would like to have added: purple, aqua, or pink. Table 2.6 are the votes of a hypothetical sample:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Table 2.6
Hypothetical Sample of New M&M Color

Person	Color
1	**Pink**
2	**Purple**
3	**Pink**
4	**Aqua**
5	**Purple**
6	**Purple**
7	**Purple**
8	**Aqua**
9	**Purple**
10	**Pink**
11	**Aqua**
12	**Pink**
13	**Pink**
14	**Purple**
15	**Aqua**
16	**Pink**
17	**Purple**
18	**Purple**
19	**Aqua**
20	**Purple**
21	**Purple**
22	**Pink**
23	**Pink**
24	**Purple**

2.4 Grouped Frequency Distribution Tables

The frequency distribution table in Table 2.4 was sufficient for presenting the frequencies for the lottery ticket variable. However, other situations may involve variables with a large number of possible values. Consider, for example, the variable grade point average (GPA). The values for GPA typically range from 0.00 to 4.00, with hundreds of possible values in between. Creating a frequency distribution table for GPA could result in an extremely large table, with many values having frequencies of zero.

To examine a variable with a large number of values, it is useful to construct a **grouped frequency distribution table** , a table that groups the values of a variable measured at the interval or ratio level

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

of measurement into a small number of intervals and then provides the frequency and percentage for each interval. To illustrate how grouped frequency distribution tables are constructed, Table 2.7 (a) provides a sample of 42 hypothetical GPAs. In Table 2.7 (b), each of these 42 GPAs has been placed into one of eight intervals that comprise a grouped frequency distribution table.

To illustrate how to read and interpret a grouped frequency distribution table, the lowest interval in Table 2.7 (b) consists of students with GPAs less than (<) 2.00. The frequency in this interval (*f* = 1) consists of one student: Student 29 (GPA = 1.96). The next interval consists of students with GPAs ranging from 2.00 to 2.29 and contains three students (4, 22, and 38). This grouping continues until all of the GPAs in the sample have been accounted for.

To create a grouped frequency distribution table, the real limits of each interval should be identified. The **real limits** of an interval are the values of the variable that are halfway between the top of one interval and the bottom of the next interval. For example, because of how values of GPA are rounded, the lowest value of GPA that would lead someone to be placed in the interval 2.30–2.59 is 2.295, which is halfway between the top of the 2.00–2.29 interval and the bottom of the 2.30–2.59 interval. The **real lower limit** is the smallest value of a variable that would be grouped into a particular interval. In this case, 2.295 represents the real lower limit for the 2.30–2.59 interval. On the other hand, the **real upper limit** is the largest value of a variable that would be grouped into a particular interval. For the 2.30–2.59 interval, the real upper limit is 2.594.

Once the grouped frequency table for the GPA example has been created (Table 2.7(b)), what preliminary conclusions might be drawn about these students? First, because the 2.90–3.19 interval has the greatest frequency (*f* = 10), you might conclude that the typical student in this sample has a GPA close to 3.00. Second, moving in both directions away from the 2.90–3.19 interval, the frequencies of the different intervals grow progressively smaller, with fewer and fewer students having either very high or very low GPAs. Table 2.8 provides a list of guidelines for creating grouped frequency distribution tables.

Cumulative Percentages

In addition to determining the percentage of the sample that has each individual value or interval for a variable, it is sometimes useful to combine these percentages. For the GPA example, you may be interested in knowing, “What percentage of the sample had GPAs less than 2.60?” To answer this question, the percentages in the “%” column of the frequency distribution table must be combined. This type of percentage is referred to as a **cumulative percentage** , defined as the percentage of a sample at or below a particular value of a variable.

Table 2.7a Grade Point Average (GPA), 42 Students: Individual GPAs

Student	GPA
1	**3.05**
2	**2.83**
3	**3.26**
4	**2.19**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Student	GPA
5	**3.52**
6	**3.34**
7	**3.02**
8	**2.97**
9	**3.71**
10	**2.50**
11	**2.77**
12	**3.10**
13	**3.06**
14	**3.83**
15	**2.39**
16	**3.16**
17	**3.92**
18	**2.37**
19	**3.65**
20	**3.70**
21	**3.00**
22	**2.26**
23	**2.65**
24	**3.40**
25	**2.87**
26	**2.92**
27	**2.89**
28	**3.41**
29	**1.96**
30	**3.77**
31	**3.20**
32	**2.86**
33	**2.94**
34	**2.71**
35	**3.43**
36	**3.95**
37	**3.32**
38	**2.28**
39	**3.08**
40	**3.25**
41	**2.40**
42	**3.22**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Table 2.7b Grade Point Average (GPA), 42 Students: Grouped Frequency Distribution Table

GPA	*f*	%
3.80+	3	7%
3.50−3.79	5	12%
3.20−3.49	9	21%
2.90−3.19	10	24%
2.60−2.89	7	17%
2.30−2.59	4	10%
2.00−2.29	3	7%
<2.00	1	2%
Total	42	100%

Table 2.8 Guidelines for Creating Grouped Frequency Distribution Tables

Guideline	Guideline Description
1. *Variables are grouped into approximately 10 intervals.*	The exact number of intervals depends on the data in each sample. For example, data that fall within a small range of values may be accurately represented with a relatively small number of intervals, whereas data extending across a wide range may require a larger number of intervals.
2. *The number of intervals should accurately represent the data.*	The intervals should represent the nature of the data as accurately as possible. For example, there should not be so large a number of intervals that many intervals have frequencies of zero (e.g., 3.00−3.05) or so few intervals that a large majority of the sample falls into only one or two intervals (e.g., 3.00−3.99).
3. *Intervals should be of equal size.*	To make the intervals comparable, they should be of equal size or width. For example, you would not want one interval of 2.01−2.50 (a width of .50) and another of 2.51−2.70 (a width of .20). One exception to this guideline is the lowest or highest interval (e.g., less than 2.00 or greater than 3.70), which is typically wider than the other intervals because it contains a relatively small frequency.
4. *Intervals should not overlap.*	The intervals should be distinct from each other, such that each score may be placed in only one interval. For example, intervals of 2.30− *2.60* and *2.60*− 2.90 are a problem because the GPA 2.60 is in both intervals leading to confusion regarding within which interval to place someone with a GPA of 2.60.

Table 2.9 includes the cumulative percentages for the GPA example. As an example of a cumulative percentage, the cumulative percentage for the 2.30–2.59 interval (19%) has been calculated by combining the percentages for three intervals: < 2.00, 2.00–2.29, and 2.30–2.59: 2% + 7% + 10% = 19%. As a result, we could conclude that 19% of this sample had GPAs less than 2.60. The accumulating of percentages continues until all (100%) of the scores in the sample are combined. As

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

shown in Table 2.9 , cumulative percentages are typically located in the final column of a frequency distribution table and labeled “Cum %.”

Table 2.9 Frequency Distribution Table With Cumulative Percentages for the GPA Variable

GPA	*f*	%	Cum %
3.80+	3	7%	**100%**
3.50−3.79	5	12%	**93%**
3.20−3.49	9	21%	**81%**
2.90−3.19	10	24%	**60%**
2.60−2.89	7	17%	**36%**
2.30−2.59	4	10%	**19%**
2.00−2.29	3	7%	**9%**
< 2.00	1	2%	**2%**
Total	42	100%	

2.5 Examining Data Using Figures

In addition to organizing data into frequency distribution tables, researchers also examine data visually using figures such as charts or graphs. There are four types of figures often used to visually portray data for variables: bar charts, pie charts, histograms, and frequency polygons. Which type of figure a researcher may use depends on the variable’s level of measurement (nominal, ordinal, interval, ratio). Data for variables measured at the nominal or ordinal level of measurement are typically displayed using bar charts or pie charts. The values of interval and ratio variables are graphed using histograms and frequency polygons. Each of these types of visual illustration is discussed in the following sections.

Displaying Nominal and Ordinal Variables: Bar Charts and Pie Charts

The values of variables measured at the *nominal* level of measurement differ in category or type; one example of a nominal variable is the lottery ticket variable discussed earlier, which consisted of four categories: sequence, pattern, nonequilibrated, and random. The values of variables measured at the *ordinal* level of measurement can be placed in an order relative to the other values; age (young, middle age, elderly) is one example of an ordinal variable.

One way to display nominal and ordinal variables is to use a bar chart. A **bar chart** is a figure that uses bars to represent the frequency or percentage of a sample corresponding to each value of a variable, with the different bars not touching each other. Figure 2.1 shows a bar chart for the data from the lottery ticket variable. In this figure, each bar corresponds to one of the four types of lottery tickets; the height of each bar corresponds to the frequency (*f*) of nonproblem gamblers who selected that type of ticket.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

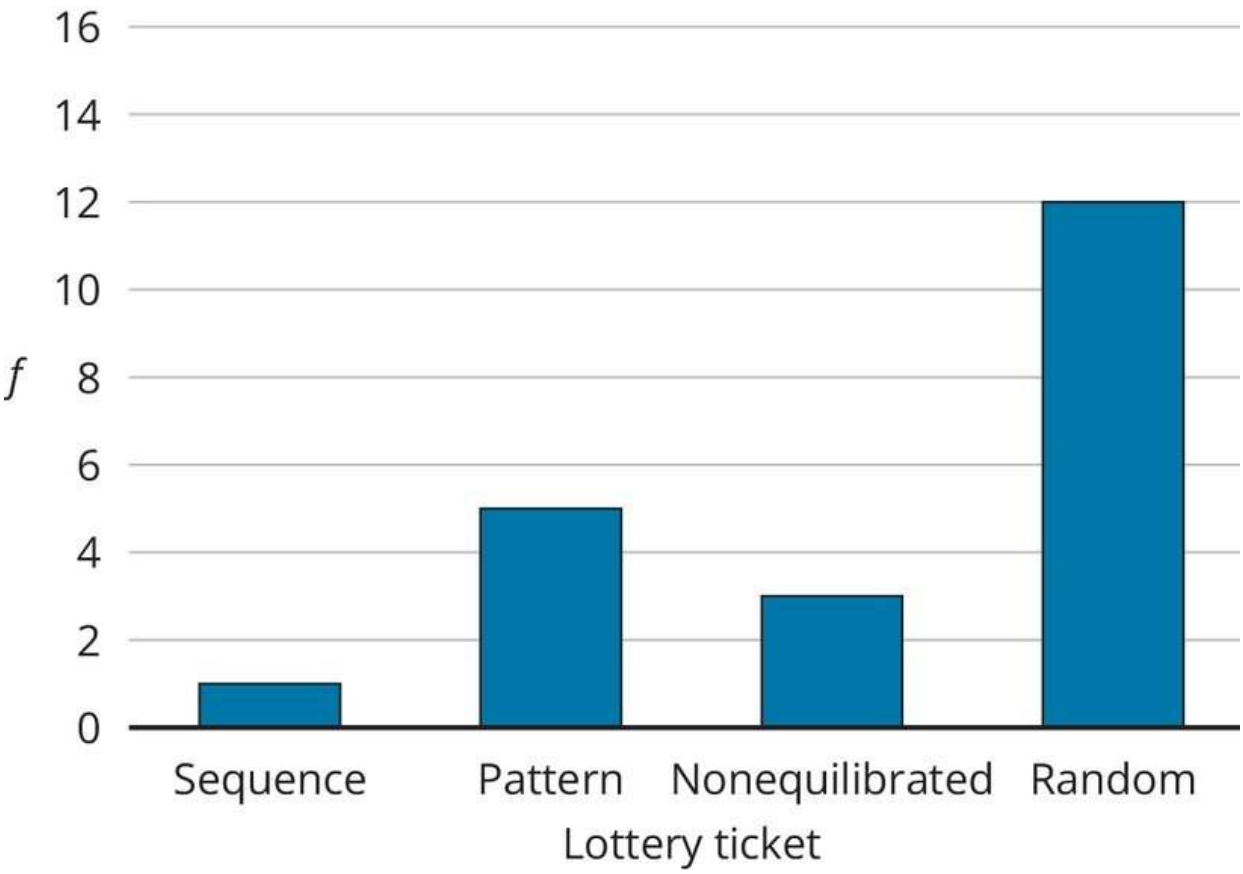


Figure 2.1 Bar Chart Showing the Frequencies of the Lottery Ticket Variable for the Nonproblem Gamblers

When creating a bar chart, the horizontal (X) axis contains the name of the variable and the values of the variable; the vertical (Y) axis contains the frequency (*f*) or percentage (%) for each value. Notice that the bars in the bar chart in Figure 2.1 do not touch each other; the gap between the bars indicates that the values of the variable represent distinct groups or categories that can't be placed along a numeric continuum.

A second type of figure used to display nominal or ordinal variables is a pie chart. A **pie chart** is a figure that uses a circle divided into proportions (slices) that represent the percentage of the sample for each value of a variable. Pie charts are created by dividing the 360 degrees of a circle into proportions corresponding to the percentage of each value.

Figure 2.2 shows a pie chart for the data for the lottery ticket variable. To illustrate how pie charts are constructed, the frequency distribution table in Table 2.4 shows that 5% of nonproblem gamblers chose the sequence ticket. Given that a pie chart consists of 360 degrees, 5% of the 360 degrees are needed to represent the sequence ticket. As 5% * 360 = 18, 18 degrees of the pie is assigned to the sequence ticket. Similarly, 97 degrees of the pie chart (27% of 360) are used to represent the pattern ticket. This process continues until all 360 degrees of the pie chart are accounted for.

The decision whether to use a bar chart or pie chart to illustrate a variable may depend on the number of values for the variable. For example, a pie chart for a variable that has 12 values may

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

result in many of the 12 “slices” being very thin and small, making them hard to see, meaning that a bar chart may be a better choice to illustrate this variable.

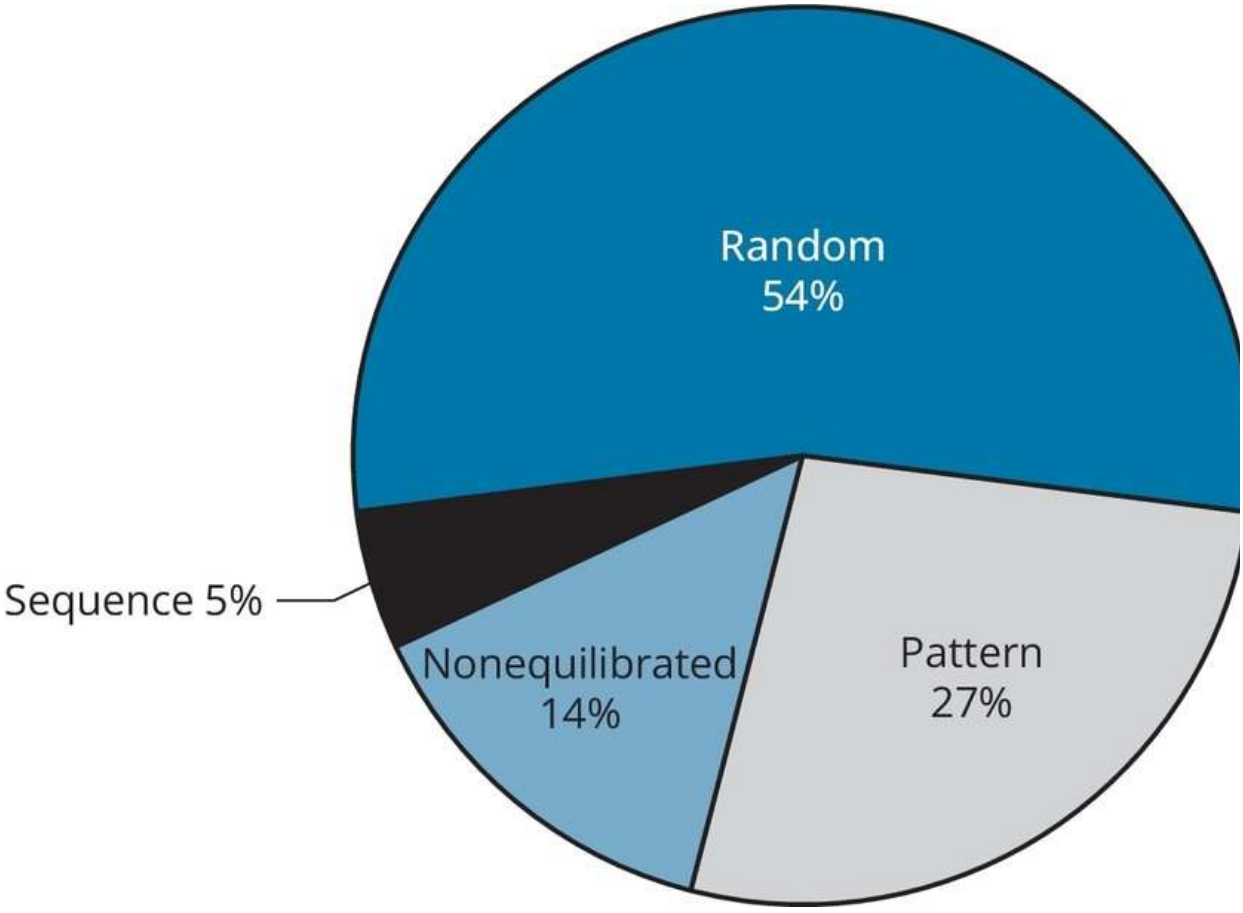


Figure 2.2 Pie Chart Showing the Percentages of the Lottery Ticket Variable for the Non−Problem Gamblers

Displaying Interval and Ratio Variables: Histograms and Frequency Polygons

Interval and ratio variables have values equally spaced along a numeric continuum and are identical to each other with one exception: Ratio scales possess a true zero point, for which the value of zero (0) represents the complete absence of the variable. The values of interval and ratio variables may be illustrated using histograms and frequency polygons, the choice of which depends on the number of values for the variable.

A **histogram** is a figure in which bars are used to represent the frequency of each value of a variable, with the different bars touching each other. The name and values of the variable are located on the X-axis, and the frequency or percentage for each value is on the Y-axis. Histograms are similar to bar charts, with one important difference: The bars in a histogram touch each other, indicating that the values of the variable are connected to each other along a numeric continuum.

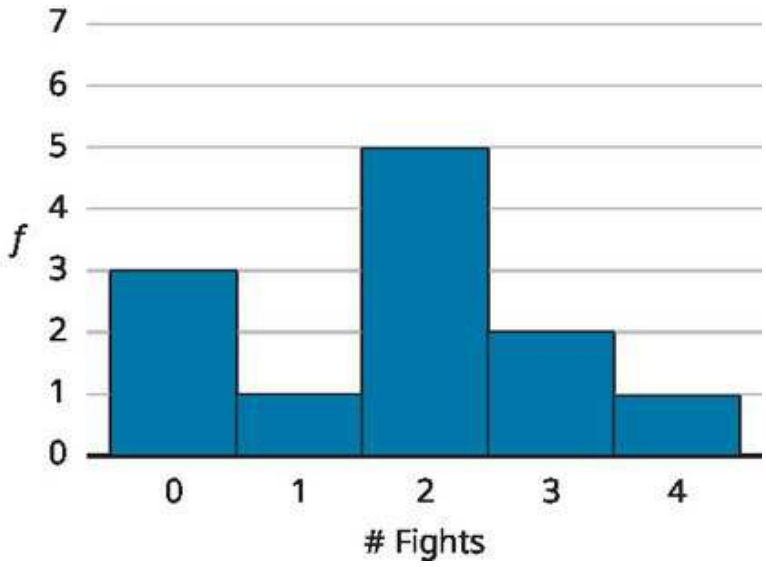
join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

As an example, imagine a teacher is interested in studying aggressive behavior in children and measures this variable by counting the number of fights each of her students is involved in over a 3-day period. “Number of fights” is a ratio variable rather than an interval variable because the value of 0 represents the complete absence of fights. Figure 2.3 provides a frequency distribution table and histogram for this variable.

A histogram is used when an interval or ratio variable consists of a relatively small number of values. But what if, for example, the number of fights could be anywhere from 0 to 20 rather than 0 to 5? In order to have the histogram fit on one page, the 21 bars would need to be very thin and crowded together, making the histogram confusing and difficult to read. In situations such as this, the values of the variable may be more clearly illustrated using a figure known as a **frequency polygon**, which is a line graph that uses data points to represent the frequency of each value of a variable, with straight lines connecting the data points. In a frequency polygon, the X-axis represents the name and values of the variable, while the Y-axis represents the frequency for each value.

# Fights	*f*
4	1
3	2
2	5
1	1
0	3
Total	12



▶ Long Description: Figure 2.3 Frequency Distribution Table and Histogram for the Number of Fights Variable

Figure 2.3 Frequency Distribution Table and Histogram for the Number of Fights Variable

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

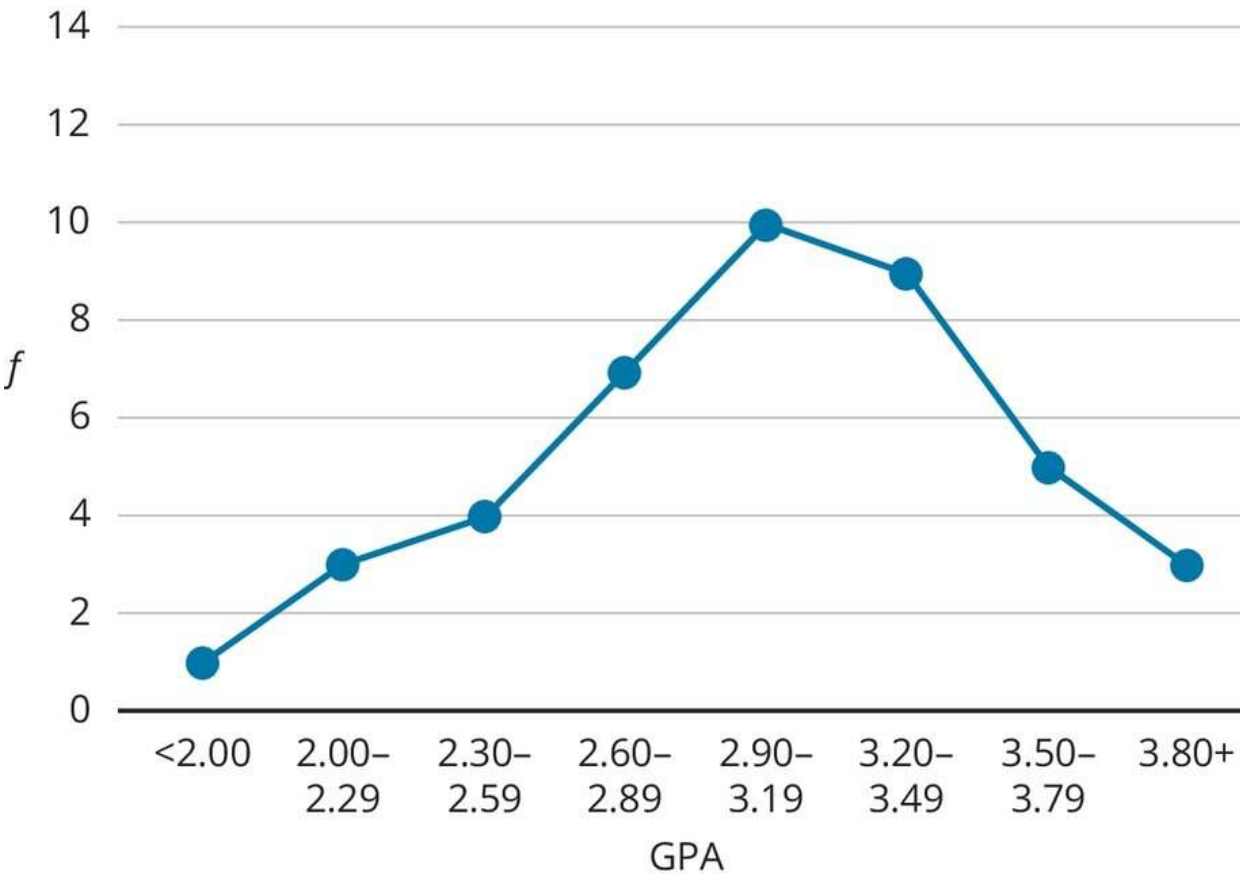


Figure 2.4 Frequency Polygon Showing the Frequency of Students in Each Interval of the Grade Point Average (GPA) Variable

Figure 2.4 shows a frequency polygon based on the grouped frequency distribution table of the earlier example of students' GPA (Table 2.7(b)). In this frequency polygon, the data points representing the frequencies of the eight GPA intervals are connected using straight, sharp lines. These lines indicate that the values of the variable can be connected together along a numeric continuum.

Although a frequency polygon like the one in Figure 2.4 may look jagged as it moves from one data point to the next, it provides a general sense of the shape of the distribution of scores for a variable. For example, looking at Figure 2.4 , you might conclude that the majority of GPAs in this sample are in the middle of the 1.00 to 4.00 range, with relatively low frequencies of students with either low or high GPAs.

Learning Check 2

Reviewing What You've Learned So Far

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



1. Review questions
    a. What is the difference between a frequency distribution table and a grouped frequency distribution table?
    b. For what levels of measurement would you create a pie chart, bar chart, histogram, or a frequency polygon?
    c. What is the difference between a bar chart and a histogram?
    d. Why would it be inappropriate to create a histogram or frequency polygon for a nominal or ordinal variable?
    e. Under what circumstances might you use a frequency polygon rather than a histogram to graph an interval or ratio variable?
2. For each of the following variables, identify whether you could use a bar chart, pie chart, histogram, and/or a frequency polygon to visually display the data.
    a. Miles per gallon (MPG)
    b. Living situation (on-campus, off-campus)
    c. Number of brothers and sisters
    d. Marital status (single, married, divorced)

Drawing Inappropriate Conclusions From Figures

A well-constructed figure can greatly facilitate the understanding of complex information. However, a poorly designed figure may not only be confusing, it can lead to inappropriate or misleading conclusions. Box 2.1 provides guidelines for constructing figures based on recommendations from the American Psychological Association (APA).

A simple example may demonstrate how altering the features of a figure can lead to different interpretations of the same set of data. Imagine a pollster assesses how voters with different political affiliations plan to vote on a ballot measure. She asks a sample of Democrats, Republicans, and Independents the question, “Do you plan on voting in favor of Measure A?” In examining her data, the pollster finds 56% of Democrats, 47% of Republicans, and 52% of Independents intend to vote “yes” on the measure.

Figure 2.5 graphs these three percentages using two bar charts; bar charts are used to display the data because the variable of interest (political affiliation) is a nominal variable (Democrat, Republican, or Independent). The bars don’t touch each other because the three political affiliations are distinct from each other and can’t be placed along a numeric continuum.

Although the two bar charts are similar, there is one important difference in how they are constructed. Following the guidelines in Box 2.1, the bar chart in Figure 2.5 (a) starts the Y-axis with the value 0% and ends with 100%. However, the bar chart in Figure 2.5 (b) starts with the value 45% and ends with 65%. By limiting the range of values of the percentages, the bar chart in Figure 2.5 (b) exaggerates differences in the heights of the three bars. As a result, someone looking at Figure 2.5 (b) may interpret the differences between the three political affiliations as being larger than someone looking at Figure 2.5 (a).

The saying “a picture is worth a thousand words” is particularly relevant for the presentation of statistical data. Because figures can provide a vivid representation of data, it’s the responsibility of researchers to portray their data in an accurate manner and to avoid creating images that may mislead or confuse.

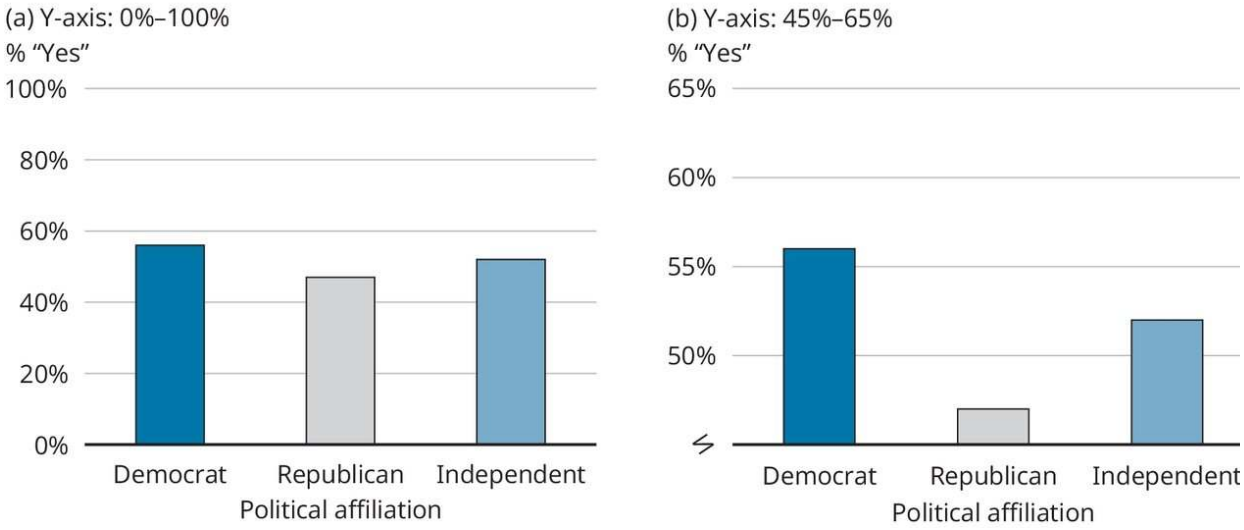
join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Box 2.1 Some APA Guidelines for Creating Figures

1. The figure should be formatted in a portrait rather than landscape page orientation (the reader should not turn the page sideways to read the table).
2. The figure is numbered using Arabic numbers (Figure 1, not Figure *I* or Figure *A*).
3. The figure includes a title that describes the contents of the figure. For example, it is not enough to name a figure simply " Figure 2.1 " or " Figure 2.1 . Bar chart."
4. Figures have a rectangular shape, with the height of the vertical (Y) axis two thirds to three fourths the length of the horizontal (X) axis.
5. The horizontal (X) axis contains the values of the variable; the vertical (Y) axis contains the frequency (*f*) or percentage (%) of the sample having each value of the variable.
6. The *X* - and *Y* -axes include descriptive labels written in both upper- and lowercase letters (as opposed to all capital letters).
7. The labels for the *X* - and *Y* -axes are short enough to fit on one line along the axes.
8. The label of the Y-axis is often placed parallel to the Y-axis (the letters face sideways rather than vertically) but can sometimes be seen on top of the y-axis for accessibility.
9. The Y-axis starts with the value zero (0) and is divided into equally spaced values or intervals.
10. The Y-axis is divided into 6 to 10 values (just enough values to accommodate bar length). The bars should neither be flattened along the bottom of the Y-axis nor bumped up against the top of the Y-axis.

Political affiliation	% "Yes"
Democrat	56%
Republican	47%
Independent	52%



▶ Long Description: Figure 2.5 Different Displays of the Same Data

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Figure 2.5 Different Displays of the Same Data

2.6 Examining Data: Describing Distributions

We've discussed how researchers construct tables and figures to organize their data. Looking at a table or figure enables researchers to describe the distribution of values or scores for a variable. Describing a distribution involves addressing questions such as, "Which values of the variable occur most often?," "How do the frequencies of the values of the variable change in relation to the most common values?," and "What is the amount of differences among the different values?"

Questions such as these pertain to three aspects of distributions: modality, symmetry, and variability. These three aspects of distributions are introduced in the following sections; they'll also be discussed in greater detail in the next two chapters of this book.

Modality

The first feature of a distribution of data for a variable is its modality. The **modality** of a distribution refers to the value or values of the variable that have the highest frequency or occur most often in a set of data. Using the lottery study as an example, identifying the modality of the lottery ticket variable involves answering the question, "Which of the four lottery tickets was chosen the most often?"

The modality of a distribution can be identified by looking at a frequency distribution table; the modality is the value(s) of the variable with the highest frequency (*f*) or percentage (%). The modality may also be determined by examining a figure and locating the tallest bar in a bar chart or histogram, the highest point in a frequency polygon, or the biggest "slice" in a pie chart. For the lottery ticket variable, the modality is the random ticket.

In identifying the modality of a distribution, it may be that one value or a limited number of values adjacent to one another clearly have the highest frequency. In such cases, the distribution is referred to as a unimodal distribution. A **unimodal distribution** is a distribution where one value of a variable occurs with the greatest frequency. Looking at the frequency distribution table (Table 2.4) or figure (Figure 2.1) for the lottery ticket variable, this distribution would be considered unimodal as one value, the random ticket, clearly occurred more often than the other three ticket choices.

In contrast to a unimodal distribution, it's possible that two distinct values may both have relatively high frequencies; if so, the distribution is referred to as a **bimodal distribution** , which is a distribution where two values clearly distinct from each other have the greatest frequency. A bimodal distribution sometimes resembles the two humps on a camel's back. As an example of a bimodal distribution, in April 2024 the Pew Research Center (www.pewresearch.org) asked a sample of 7,000 voters whether they felt Donald Trump had the mental fitness needed to be president, with five responses ranging from "extremely" to "not at all." The majority of voters chose one of the two extreme choices, either "extremely" (23%) or "not at all" (35%), with only 14% choosing the middle response of "somewhat."

And, in addition to unimodal and bimodal distributions, it's possible for a distribution to have more than two values with relatively equally high frequencies; this type of distribution is known as a **multimodal distribution** —a distribution where more than two values have the greatest frequency. In examining people's attitudes toward gun control, in May 2023, the news network CNN asked 1,227 people the question, "Do you think allowing gun owners to carry their guns in public places makes those places safer, less safe, or makes no difference in how safe a public place is?" They found people differed greatly in their responses, with 32% saying "makes public places safer," 36% saying

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

"makes public places less safe," and 32% saying "makes no difference," Figure 2.6 presents an example of a unimodal, bimodal, and multimodal distribution.

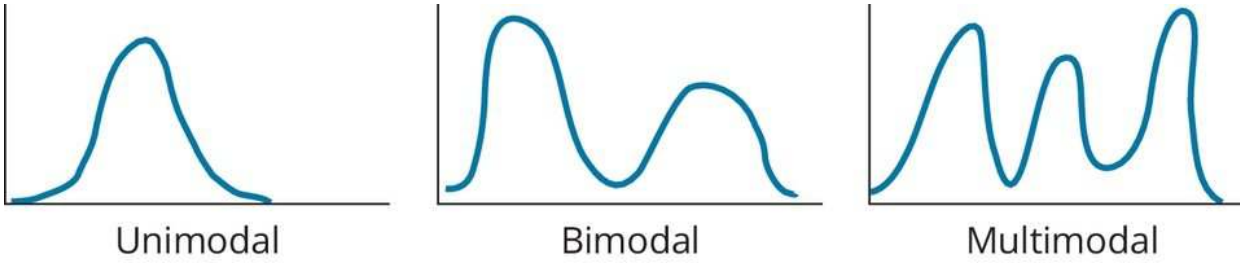


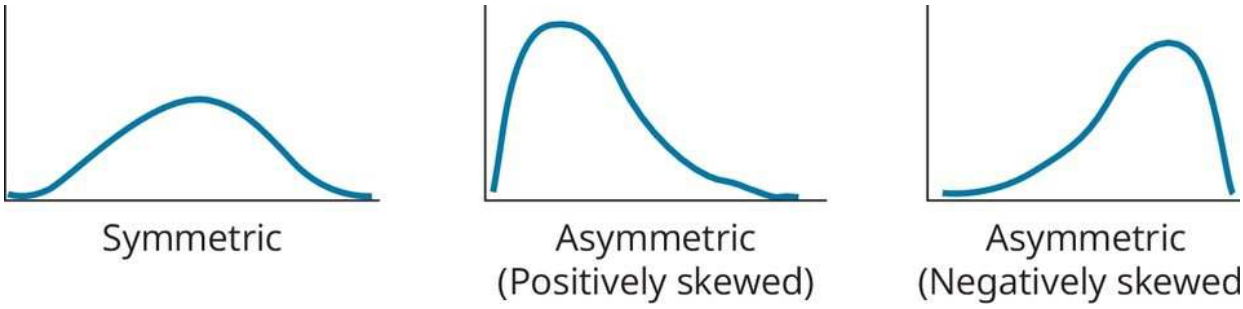
Figure 2.6 Distributions With Different Modalities (Unimodal, Bimodal, and Multimodal)

Symmetry

The second feature of a distribution of data is its symmetry. The **symmetry** of a distribution refers to how the frequencies of values of a variable change in relation to the most common or frequently occurring values. Describing the symmetry of a distribution means answering the question, "What is the shape of the distribution?"

The symmetry of a distribution can be determined by looking at the frequencies in a frequency distribution table or the bars or points in a figure. By doing so, a distribution may be described as being either symmetric or asymmetric. A **symmetric distribution** is one in which the frequencies change in a similar manner moving away in both directions from the most frequently occurring values. An example of a symmetric distribution is illustrated in Figure 2.7 . One way to determine whether a distribution is symmetric is to draw the distribution on a piece of paper and then fold the distribution vertically down the middle so that the two halves are on top of each other. A distribution is symmetric when the two halves are similar in shape such that they resemble mirror images of each other.

However, a distribution may also be asymmetric. An **asymmetric distribution** (also known as a *skewed* distribution) is one in which the frequencies change in a different manner moving away in both directions from the most frequently occurring values. In asymmetric distributions, the highest frequencies are located at one end of the distribution rather than in the middle, such that if you fold an asymmetric distribution down the middle, the shapes of the two halves do not look the same.



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Figure 2.7 Distributions With Different Symmetries (Symmetric, Positively Skewed, and Negatively Skewed)

In an asymmetric distribution, the frequencies take on the shape of a long tail as the values move toward the end of the distribution; scores in the tail of the distribution are often considered to be outliers. For example, in the 2022 Current Population Survey, the U.S. Census Bureau (www.census.gov) estimated that most Americans (63%) had annual incomes less than $50,000. However, a very small percentage (3%) were greater than $200,000; these higher incomes would be considered outliers.

Depending on the location of its tail, asymmetric distributions are referred to either as positively skewed or negatively skewed. A **positively skewed distribution** is one in which the higher frequencies are at the lower end of the distribution, with the tail on the upper (right) end of the distribution; a **negatively skewed distribution** is one in which the higher frequencies are at the upper end of the distribution, with the tail on the lower (left) end of the distribution. Figure 2.7 provides an illustration of a symmetric distribution and two asymmetric distributions (one positively skewed and one negatively skewed). Annual income, illustrated by the U.S. Census Bureau example mentioned earlier, is an example of a positively skewed distribution.

Variability

The third way in which distributions may be described is in terms of the amount or nature of differences among the different values for a variable. The **variability** of a distribution refers to the amount of differences in a distribution of data for a variable. The variability in a distribution is described by answering the question, “To what extent are the scores in the distribution different from each other?”

The amount of variability in a distribution can be estimated by looking at a frequency distribution table or figure. Imagine, for example, an instructor gives an 11-item quiz to her class and finds almost all of the students answer six, seven, or eight of the items correctly. In this situation, the distribution would resemble the shape of a tall mountain: thin and tall, with many of the scores residing in a few values of the variable. A distribution with a small amount of variability such as this one is referred to as a **peaked distribution** , defined as a distribution where much of the data are in a small number of values of a variable.

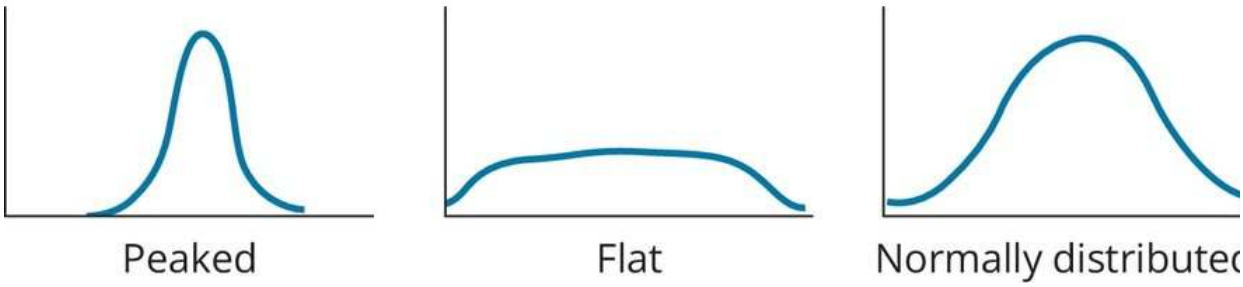


Figure 2.8 Distributions With Different Variability (Peaked, Flat, and Normally Distributed)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

On the other hand, what would the distribution look like if she had found her students' scores were distributed relatively evenly across the entire range of possible scores? In this situation, the distribution would not be tall and thin like a mountain but rather would be low and wide like a plateau. This type of distribution is referred to as a **flat distribution** , which is a distribution where the data are spread evenly across the values of a variable. The CNN gun control example described earlier is an example of a flat distribution.

In discussing modality, symmetry, and variability, distributions may have more than one mode, may be asymmetric (skewed), and may be peaked or flat. But what if a distribution has none of these qualities? A distribution that is unimodal, symmetric, and neither peaked nor flat may be described as "bell shaped." If you look at a bell, you'll notice that it is the highest in the middle, and as you move away from the top of a bell in both the left and right directions, the shape of the bell slopes downward in the same manner. A bell-shaped distribution for a variable is given a special name: a normally distributed variable.

A **normally distributed variable** is a distribution for a variable that is considered unimodal, symmetric, and neither peaked nor flat. Note that the word *normal* is a statistical concept and not a value judgment. In other words, asymmetric, peaked, or flat distributions are not considered "abnormal." The concept of normal distributions will be discussed in greater detail in Chapter 5. Figure 2.8 provides an illustration of three distributions with different amounts of variability: peaked, flat, and normally distributed.

Learning Check 3

Reviewing What You've Learned So Far



1. Review questions
    a. What are the three aspects used to describe distributions of variables?
    b. What is the difference between unimodal, bimodal, and multimodal distributions?
    c. What is the difference between symmetric and asymmetric (skewed) distributions?
    d. What is the difference between positively and negatively skewed distributions?
    e. What is the difference between peaked and flat distributions?
    f. What are the characteristics of a normally distributed variable?
2. For each of the following sets of data, determine whether the distribution of scores is unimodal, bimodal, or multimodal.
    a. 4, 2, 11, 6, 10, 2, 6, 13, 2, 10, 6, 10, 1
    b. 15, 17, 13, 18, 17, 14, 20, 18, 17, 14, 17
    c. 20, 10, 50, 10, 80, 30, 70, 10, 60, 10, 80, 0, 30, 80, 10, 100, 80
3. For each of the following sets of data, determine whether the distribution of scores is symmetric or asymmetric (skewed).
    a. 4, 5, 1, 4, 2, 4, 3, 5, 3, 4, 4, 3
    b. 220, 230, 250, 230, 200, 240, 210, 230, 270, 220, 240, 230, 240, 230
    c. 3, 1, 9, 2, 3, 6, 2, 7, 3, 5, 2, 6, 3, 2, 2
4. For each of the following sets of data, determine whether the distribution of scores is peaked or flat.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

a. 16, 10, 19, 12, 13, 15, 14, 17, 15, 20, 11, 13, 19, 16, 15
b. 12, 15, 12, 11, 13, 11, 12, 12, 14, 10, 12, 11, 16, 12, 13, 12, 11
c. 14, 12, 16, 14, 13, 15, 12, 14, 14, 15, 14, 14, 13, 14

2.7 Looking Ahead

This chapter has discussed why and how researchers examine data. Examining data by creating tables and figures requires a number of different skills, including organizational skills (to prepare a set of data for statistical analysis), communication skills (to present data in tables or figures), and conceptual skills (to describe a distribution). The next two chapters focus on mathematical skills related to the statistical analysis of data, describing how researchers describe distributions numerically rather than visually. Given that this is a statistics textbook, we will focus on calculating statistics to describe and analyze data. However, and we cannot say this strongly enough, it is *extremely* important to examine a set of data before conducting statistical analyses as this will prevent errors both in your calculations of statistics as well as your interpretation of the results of these calculations.

2.8 Summary

Before conducting statistical analyses, researchers typically examine their data to gain an initial sense of the data, detect data coding or data entry errors, identify ***outliers*** (rare, extreme scores that lie outside of the range of the majority of scores in a set of data), evaluate research methodology, and determine whether the data meet statistical criteria and assumptions.

One common way of examining variables involves creating tables known as ***frequency distribution tables*** (tables that summarize the number and percentage of participants that express the different values for a variable) and ***grouped frequency distribution tables*** (tables that group the values of a variable measured at the interval or ratio level of measurement into a small number of intervals and then provide the frequency and percentage within each interval). To create a grouped frequency table, the ***real limits*** of each interval (the values of the variable that fall halfway between the top of one interval and the bottom of the next interval) need to be identified.

Data for a variable may also be examined by creating figures such as bar charts, pie charts, histograms, and frequency polygons. ***Bar charts*** and ***pie charts*** are used for variables measured at the nominal or ordinal level of measurement; a bar chart uses bars to represent the frequency or percentage of a sample corresponding to each value of a variable, with the different bars not touching each other; a pie chart is a circle divided into proportions that represent the percentage of the sample corresponding to each value of a variable. ***Histograms*** and ***frequency polygons*** are used for variables measured at the interval or ratio level of measurement; a histogram is a figure in which bars are used to represent the frequency of each value of a variable, with the different bars touching each other; a frequency polygon is a line graph that uses data points to represent the frequency of each value of a variable, with lines connecting the data points.

From examining their data, researchers can describe the distribution of values or scores for a variable. There are three main aspects of distributions: modality, symmetry, and variability.

The ***modality*** of a distribution is the value(s) of the variable that have the highest frequencies or occur most often in a set of data. ***Unimodal distributions*** are distributions where one value occurs with the greatest frequency, ***bimodal distributions*** have two values with the greatest frequency, and ***multimodal distributions*** have more than two values with the greatest frequencies.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The ***symmetry*** of a distribution refers to how the frequencies of values of a variable change in relation to the most common or frequently occurring values of the variable. In ***symmetric distributions*** , the frequencies change in a similar manner, moving away in both directions from the most common values of the variable. In ***asymmetric (skewed) distributions*** , the frequencies change in a different manner, moving away in both directions from the most common values of the variable.

The ***variability*** of a distribution refers to the amount of differences in a distribution of data. In ***peaked distributions*** , much of the data are in a few values of a variable; in ***flat distributions*** , the data are spread evenly across the values of a variable. A ***normally distributed variable*** has a symmetric distribution that is neither peaked nor flat.

2.9 Important Terms

asymmetric (skewed) distribution

bar chart

bimodal distribution

cumulative percentage

flat distribution

frequency (*f*)

frequency distribution table

frequency polygon

grouped frequency distribution table

histogram

modality

multimodal distribution

negatively skewed distribution

normally distributed variable

outlier

peaked distribution

pie chart

positively skewed distribution

real limits

real lower limit

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

real upper limit

symmetric distribution

symmetry

unimodal distribution

variability

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

2.10 Formula Introduced in This Chapter

Percentage (%)

$$\% = \frac{f}{total\ number\ of\ scores} * 100$$

2.11 Using IBM SPSS Software *

Creating Data Files: Defining Variables and Entering Data

* IBM SPSS software. SPSS is a registered trademark of International Business Machines Corporation.

1. Open a new data file (See Figure 2.9).

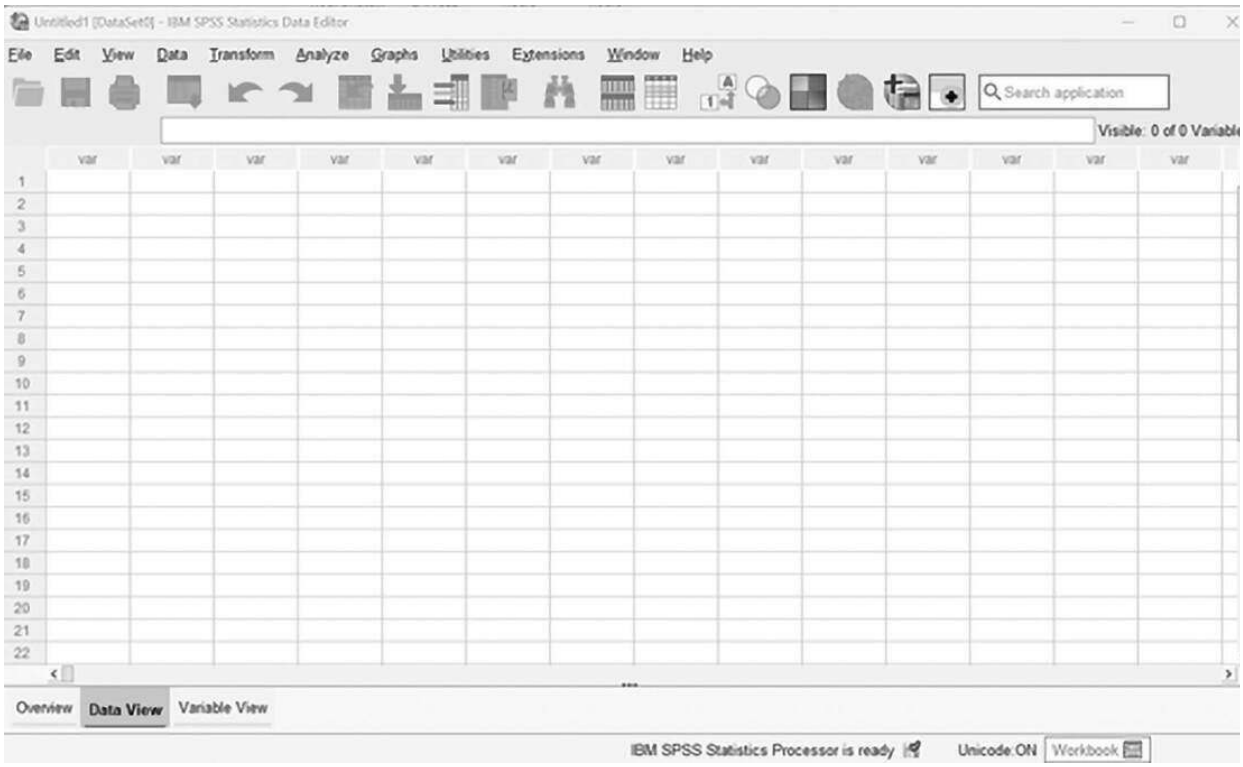


Figure 2.9 Data View

2. Define and name variable(s) (See Figure 2.10).

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

How? (1) Click **Variable View** tab, (2) click below **Name** , and (3) type name of variable in the cell.

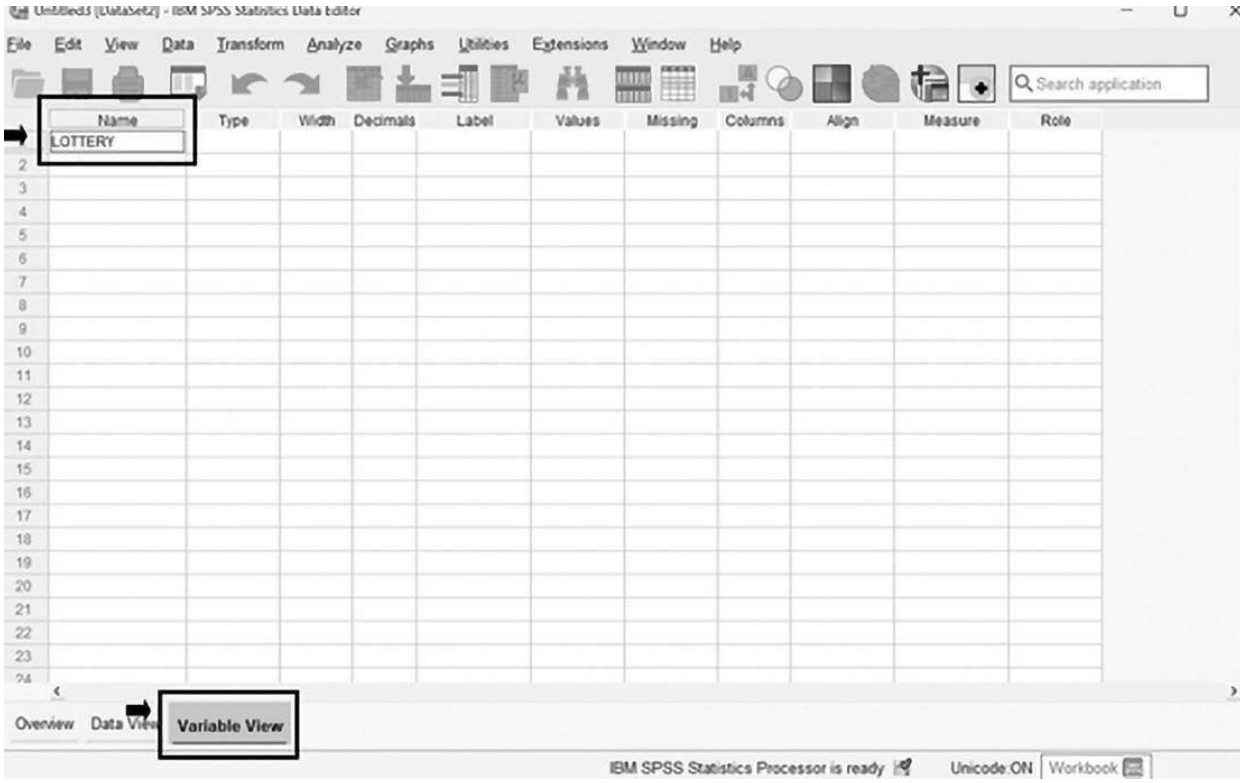


Figure 2.10 Variable View

3. Determine number of decimals for your variable(s) (See Figure 2.11).

How? Click desired number of decimal places in the **Decimals** box.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

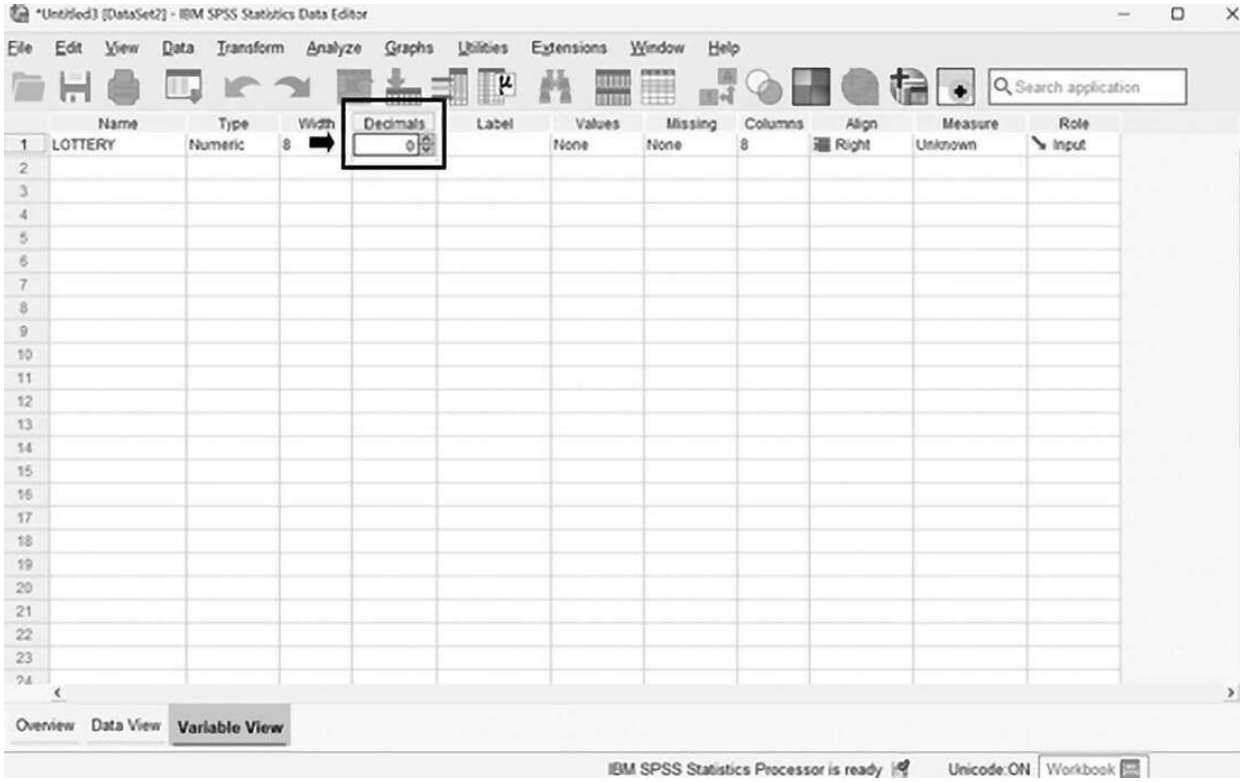


Figure 2.11 Decimals Box

4. Provide labels for your variable(s) (See Figure 2.12).

How? (1) Click below **Label** , and (2) type the label for the variable in the cell.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

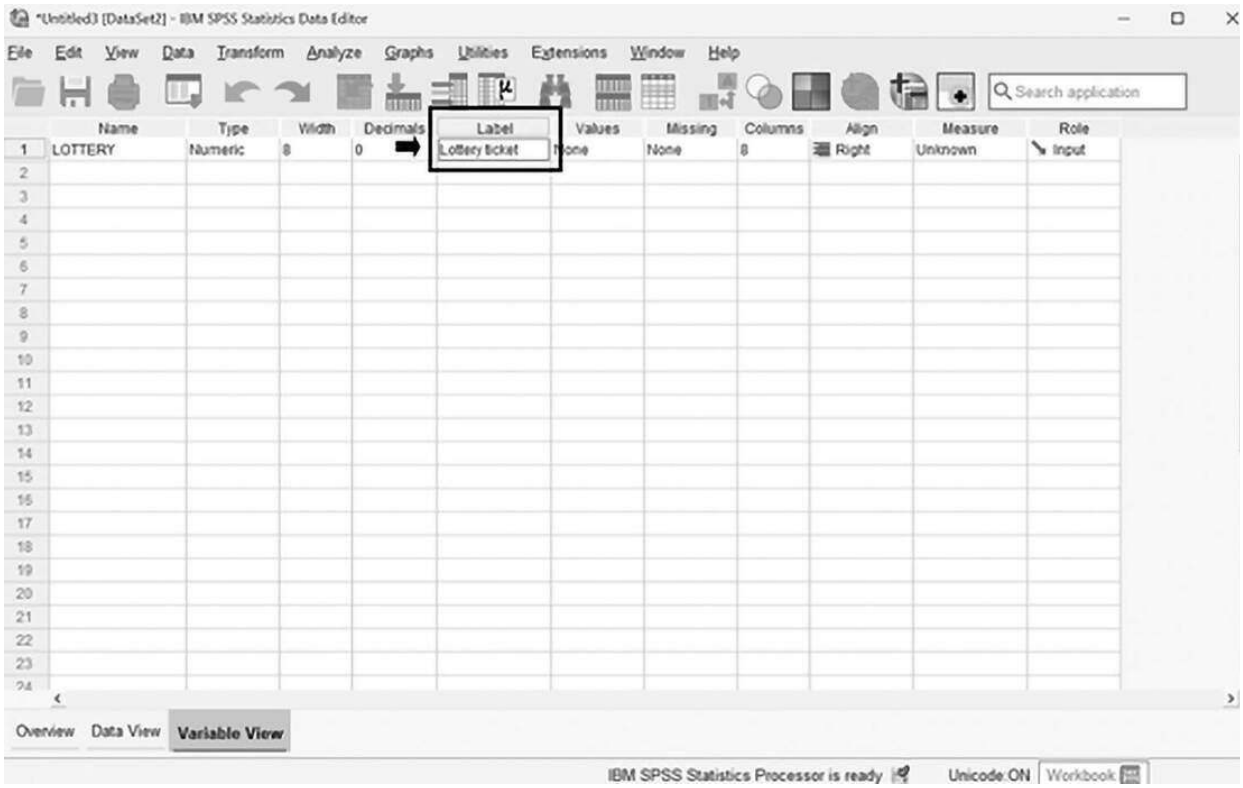


Figure 2.12 Label Box

5. Provide labels for values of nominal (categorical) variable(s) (See Figure 2.13).

How? (1) Click below **Values** , and (2) type labels for coded values of the variable.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

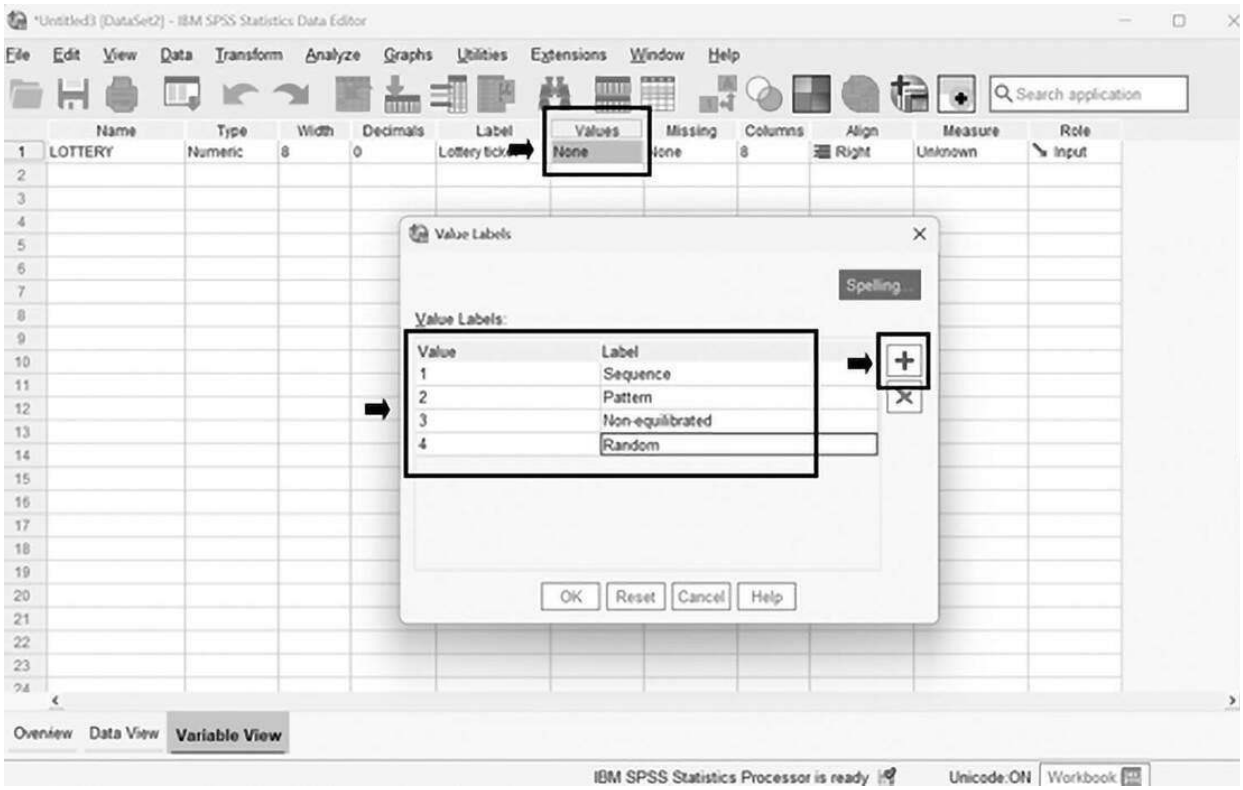


Figure 2.13 Nominal Variables

6. Enter data (See Figure 2.14).

How? (1) Click **Data View** tab and (2) enter data for the variable in the cells.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

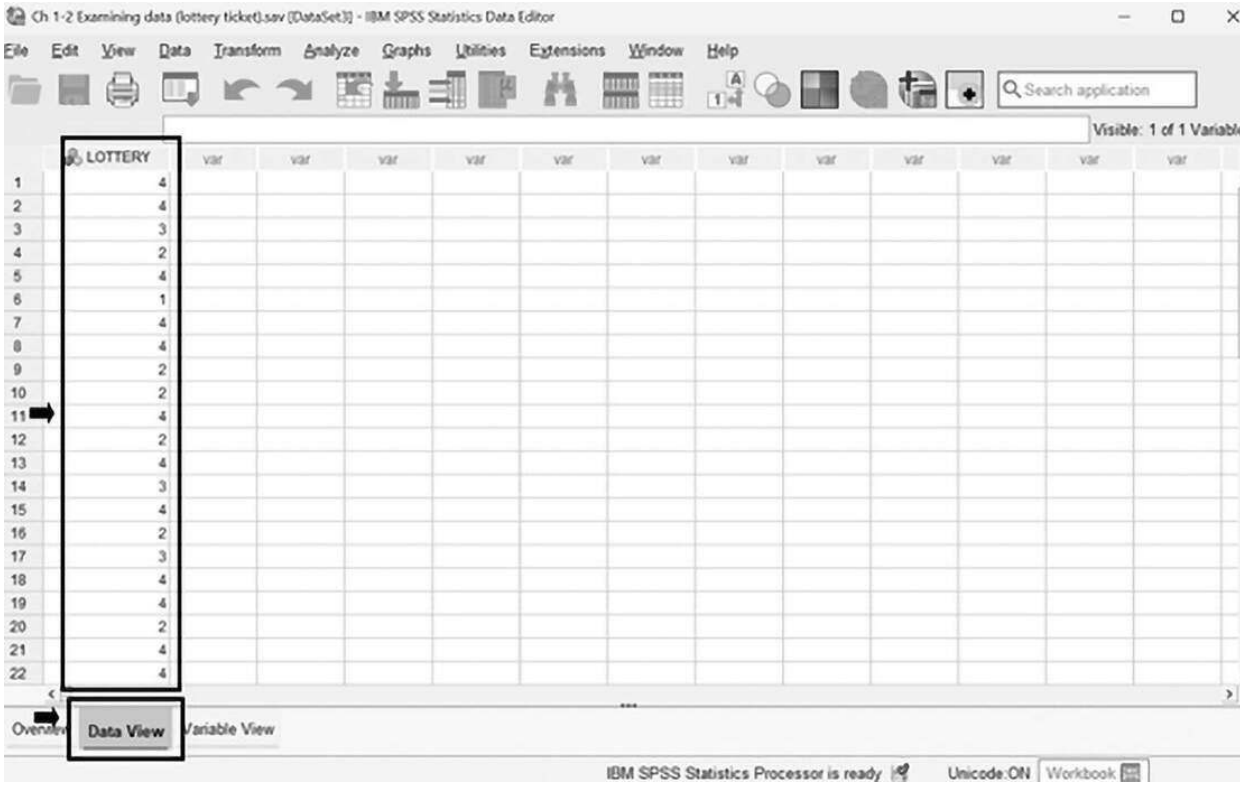


Figure 2.14 Enter Data View

Examining Data: Frequency Distribution Tables and Figures: The Lottery Study (2.1)

1. Select the frequency distribution procedure within SPSS (See Figure 2.15).

   How? (1) Click **Analyze** menu, (2) click **Descriptive Statistics** , and (3) click **Frequencies** .

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

To Jennifer , my kids Meagan and Eika , and my parents, Katsumi and Grayce Tokunaga.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

As a global academic publisher, Sage is driven by the belief that research and education are critical in shaping society. Our mission is building bridges to knowledge—supporting the development of ideas into scholarship that is certified, taught, and applied in the real world.

Sage's founder, Sara Miller McCune, transferred control of the company to an independent trust, which guarantees our independence indefinitely. This enables us to support an equitable academic future over the long term by building lasting relationships, championing diverse perspectives, and co-creating social and behavioral science resources that transform teaching and learning.

▶ Long Description:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

FUNDAMENTAL STATISTICS FOR THE SOCIAL AND BEHAVIORAL SCIENCES

Third Edition

Howard T. Tokunaga

San Jose State University



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

eISBN: 978-1-0718-6160-8

Copyright © 2027 by Sage.

All rights reserved. Except as permitted by U.S. copyright law, no part of this work may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without permission in writing from the publisher.

All third party trademarks referenced or depicted herein are included solely for the purpose of illustration and are the property of their respective owners. Reference to these trademarks in no way indicates any relationship with, or endorsement by, the trademark owner.



FOR INFORMATION:

2455 Teller Road, Thousand Oaks, California 91320, E-mail: orders@sagepub.com

3rd floor, HYLO, 103-105 Bunhill Row, London, EC1Y 8LZ, United Kingdom

10th Floor, Emaar Capital Tower 2, MG Road Sikanderpur, Sector 26, Gurugram, Haryana – 122002, India

18 Cross Street #10-10/11/12, China Square Central, Singapore 048423

Library of Congress Control Number: 2026943277

ISBN: 978-1-0718-6144-8

Printed in the United States of America

This book is printed on acid-free paper.

26 27 28 29 30 10 9 8 7 6 5 4 3 2 1

Acquisitions Editor: Anna Villarruel

Content Development Editor: Kenzie Offley

Editorial Assistant: Daria Terry

Production Editor: Syeda Aina Rahat Ali

Copy Editor: Michelle Ponce

Typesetter: Lumina Datamatics

Cover Designer: Candice Harman

Marketing Manager: Victoria Velasquez

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CONTENTS

Preface
Acknowledgments
About the Author
Chapter 1 Introduction to Statistics
Chapter 2 Examining Data: Tables and Figures
Chapter 3 Measures of Central Tendency
Chapter 4 Measures of Variability
Chapter 5 Normal Distributions
Chapter 6 Probability and Introduction to Hypothesis Testing
Chapter 7 Testing One Sample Mean
Chapter 8 Estimating the Mean of a Population
Chapter 9 Testing the Difference Between Two Means
Chapter 10 Errors in Hypothesis Testing, Statistical Power, and Effect Size
Chapter 11 One-Way Analysis of Variance (ANOVA)
Chapter 12 Two-Way Analysis of Variance (ANOVA)
Chapter 13 Correlation
Chapter 14 Linear Regression and Multiple Correlation
Chapter 15 Chi-Square
Appendix A: Tables
Appendix B: Review of Basic Mathematics
Glossary
References
Index

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

DETAILED CONTENTS

Preface

Acknowledgments

About the Author

Chapter 1 Introduction to Statistics

- 1.1 What Is Statistics?
- 1.2 Why Learn Statistics?
- 1.3 Introduction to the Stages of the Research Process
  - Developing a Research Hypothesis to Be Tested
    - Identifying a Question or Issue to Be Examined
    - Reviewing and Evaluating Relevant Theories and Research
    - Stating a Research Hypothesis: Independent and Dependent Variables
  - Collecting Data
    - Drawing a Sample From a Population
    - Determining How Variables Will Be Measured: Levels of Measurement
    - Selecting a Method to Collect the Data: Experimental and Nonexperimental Research Methods
  - Analyzing the Data
    - Calculating Descriptive Statistics
    - Calculating Inferential Statistics
  - Drawing a Conclusion Regarding the Research Hypothesis
  - Communicating the Findings of the Study
- 1.4 Plan of the Book
- 1.5 Looking Ahead
- 1.6 Summary
- 1.7 Important Terms
- 1.8 Exercises

Answers to Learning Checks

Learning Check 1

Learning Check 2

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Odd-Numbered Exercises

Chapter 2 Examining Data: Tables and Figures

2.1 An Example From the Research: Winning the Lottery

2.2 Why Examine Data?

Gaining an Initial Sense of the Data

Identifying Outliers

Evaluating Research Methodology

Determining Whether Data Meet Statistical Criteria and Assumptions

2.3 Examining Data Using Tables

What Are the Values of the Variable?

How Many Participants in the Sample Have Each Value of the Variable?

What Percentage of the Sample Has Each Value of the Variable?

2.4 Grouped Frequency Distribution Tables

Cumulative Percentages

2.5 Examining Data Using Figures

Displaying Nominal and Ordinal Variables: Bar Charts and Pie Charts

Displaying Interval and Ratio Variables: Histograms and Frequency Polygons

Drawing Inappropriate Conclusions From Figures

2.6 Examining Data: Describing Distributions

Modality

Symmetry

Variability

2.7 Looking Ahead

2.8 Summary

2.9 Important Terms

2.10 Formula Introduced in This Chapter

Percentage (%)

2.11 Using IBM SPSS Software *

Creating Data Files: Defining Variables and Entering Data

Examining Data: Frequency Distribution Tables and Figures: The Lottery Study (2.1)

2.12 Exercises

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 3 Measures of Central Tendency
3.1 An Example From the Research: The 10% Myth
3.2 Understanding Central Tendency
3.3 The Mode
Determining the Mode
3.4 The Median
Calculating the Median: Odd Number of Scores
Calculating the Median: Even Number of Scores
Calculating the Median When Multiple Scores Have the Median Value
3.5 The Mean
Calculating the Mean
Reporting Calculated Numbers
Rounding Calculated Numbers
Calculating the Mean From Frequency Distribution Tables
The Mean as a Balancing Point in a Distribution
The Sample Mean vs. the Population Mean: Statistics vs. Parameters
3.6 Comparison of the Mode, Median, and Mean
Strengths and Weaknesses of the Mode
Strengths and Weaknesses of the Median
Strengths and Weaknesses of the Mean
Choosing a Particular Measure of Central Tendency
3.7 Measures of Central Tendency: Drawing Conclusions
3.8 Looking Ahead
3.9 Summary
3.10 Important Terms
3.11 Formulas Introduced in This Chapter
Median (Odd Number of Scores)
Median (Even Number of Scores)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Mean
Mean (Frequency Distribution Table)
Mean (Balancing Point in a Distribution)
3.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 4 Measures of Variability
4.1 An Example From the Research: How Many Sometimes" in an Always"?
4.2 Understanding Variability
4.3 The Range
Strengths and Weaknesses of the Range
4.4 The Interquartile Range
Strengths and Weaknesses of the Interquartile Range
4.5 The Variance (s^2)
Definitional Formula for the Variance
Computational Formula for the Variance
Why Not Use the Absolute Value of the Deviation in Calculating the Variance?
Why Divide by N − 1 Rather Than N in Calculating the Variance?
4.6 The Standard Deviation (s)
Definitional Formula for the Standard Deviation
Computational Formula for the Standard Deviation
Eyeball Estimating" the Mean and Standard Deviation
4.7 Measures of Variability for Populations
The Population Variance (σ^2)
The Population Standard Deviation (σ)
Measures of Variability for Samples vs. Populations
4.8 Measures of Variability for Nominal Variables
The Index of Qualitative Variation
4.9 Measures of Variability: Drawing Conclusions

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

4.10 Looking Ahead
4.11 Summary
4.12 Important Terms
4.13 Formulas Introduced in This Chapter
Range
Interquartile Range
Variance (s^2) (Definitional Formula)
Variance (s^2) (Computational Formula)
Standard Deviation (s) (Definitional Formula)
Standard Deviation (s) (Computational Formula)
Population Variance (σ^2) (Definitional Formula)
Population Standard Deviation (σ) (Definitional Formula)
Index of Qualitative Variation (IQV)
4.14 Using SPSS
Calculating Measures of Central Tendency and Variability: The Frequency Rating Study (4.1)
4.15 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 5 Normal Distributions
5.1 Normal Distributions
What Are Normal Distributions?
Why Are Normal Distributions Important?
Is There Only One Normal Distribution?
5.2 The Standard Normal Distribution
What Is the Standard Normal Distribution?
Interpreting z-Scores: Distance From the Mean
Interpreting z-Scores: Relative Position Within the Distribution
Finding the Area Between the Mean and a z-Score

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Finding the Area Less or Greater Than a z-Score
Finding the Area Between Two z-Scores
5.3 Applying z-Scores to Normal Distributions
Transforming Normal Distributions Into the Standard Normal Distribution
Understanding the z-Score Formula
Example: Transforming SAT Scores Into z-Scores
Interpreting z-Scores: Distance From the Mean
Interpreting z-Scores: Relative Position Within Distribution
Finding the Area Between the Mean and a Score
Finding the Area Less or Greater Than a Score
Finding the Area Between Two Scores
Transforming Scores Into z-Scores: Drawing Conclusions
5.4 Standardizing Frequency Distributions
Example: Measuring Politicians' Behavior
Standardizing a Frequency Distribution
Interpreting Standardized Scores
Using Standardized Scores to Compare or Combine Variables
Characteristics of Standardized Distributions
When Can You Apply the Normal Curve Table to a Standardized Distribution?
5.5 Looking Ahead
5.6 Summary
5.7 Important Terms
5.8 Formulas Introduced in This Chapter
z-Score
Standardized Score
5.9 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 6 Probability and Introduction to Hypothesis Testing

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

6.1 A Brief Introduction to Probability
- What Is Probability?
- Why Is Probability Important to Researchers?
- Applying Probability to Normal Distributions
- Applying Probability to Binomial Distributions

6.2 Example: Making Heads or Tails of the Super Bowl

6.3 Introduction to Hypothesis Testing
- State the Null and Alternative Hypotheses (H_0 and H_1)
  - The Null Hypothesis (H_0)
  - The Alternative Hypothesis (H_1)
- Make a Decision About the Null Hypothesis
  - Defining a Low" Probability of a Statistic: Setting Alpha (α)
  - Identifying the Values of the Statistic With a Low Probability
  - Stating a Decision Rule
  - Calculating a Value of a Statistic
  - Making the Decision Whether to Reject the Null Hypothesis
- Draw a Conclusion From the Analysis
- Relate the Result of the Analysis to the Research Hypothesis
- Summary of Steps in Hypothesis Testing

6.4 Issues Related to Hypothesis Testing: An Introduction
- The Issue of Proof" in Hypothesis Testing
- Errors in Decision Making
- Factors Influencing the Decision About the Null Hypothesis
  - Sample Size
  - Alpha (α)
  - Directionality of the Alternative Hypothesis

6.5 Looking Ahead

6.6 Summary

6.7 Important Terms

6.8 Formula Introduced in This Chapter

Probability

6.9 Exercises

Answers to Learning Checks

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 7 Testing One Sample Mean
7.1 An Example From the Research: Do You Read Me?
7.2 The Sampling Distribution of the Mean
Characteristics of the Sampling Distribution of the Mean
7.3 Inferential Statistics: Testing One Sample Mean (σ Known)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: z-Test for One Mean
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the z-Test for One Mean
Summary
7.4 A Second Example From the Research: Unique Invulnerability
7.5 Introduction to the t-Distribution
Characteristics of the Student t-Distribution
7.6 Inferential Statistics: Testing One Sample Mean (σ Not Known)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: t-Test for One Mean
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the t-Test for One Mean

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Summary
7.7 Factors Affecting the Decision About the Null Hypothesis
Sample Size
Alpha (α)
Directionality of the Alternative Hypothesis
7.8 Looking Ahead
7.9 Summary
7.10 Important Terms
7.11 Formulas Introduced in This Chapter
z-Test for One Mean
Population Standard Error of the Mean $\left(\mathbf{i}t\sigma_{\mathbf{i}t\overline{X}}\right)$
Degrees of Freedom (df) for One Sample
t-Test for One Mean
Standard Error of the Mean $\left(\mathbf{i}ts_{\mathbf{i}t\overline{X}}\right)$
7.12 Using SPSS
Testing One Sample Mean (σ Not Known): The Unique Invulnerability Study (7.6)
7.13 Exercises
Answers to Learning Checks
Learning Check 2
Learning Check 3
Learning Check 5
Learning Check 6
Answers to Odd-Numbered Exercises
Chapter 8 Estimating the Mean of a Population
8.1 An Example From the Research: Salary Survey
8.2 Introduction to the Confidence Interval for the Mean
Using the Sampling Distribution of the Mean to Create the Confidence Interval for the Mean
8.3 The Confidence Interval for the Mean (σ Not Known)
State the Desired Level of Confidence
Calculate the Confidence Interval and Confidence Limits
Calculate the Standard Error of the Mean $\left(s_{\overline{X}}\right)$

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Calculate the Degrees of Freedom (df) and Identify the Critical Value of t

Calculate the Confidence Interval and Confidence Limits

Draw a Conclusion About the Confidence Interval

Probability of an Interval or the Probability of a Population Mean?

Summary

8.4 The Confidence Interval for the Mean (σ Known)

State the Desired Level of Confidence

Calculate the Confidence Interval and Confidence Limits

Calculate the Population Standard Error of the Mean $\left(\sigma_{\overline{X}}\right)$

Identify the Critical Value of z

Calculate the Confidence Interval and Confidence Limits

Draw a Conclusion About the Confidence Interval

Summary

8.5 Factors Affecting the Width of the Confidence Interval for the Mean

Sample Size

Estimating the Sample Size Needed for a Desired Confidence Interval

Level of Confidence

Choosing Between Different Levels of Confidence: Probability Versus Precision

8.6 Interval Estimation and Hypothesis Testing

Concerns About Hypothesis Testing

The Interpretation of Statistical Significance and Nonsignificance

The Influence of Sample Size on the Decision About the Null Hypothesis

Interval Estimation as an Alternative to Hypothesis Testing

8.7 Looking Ahead

8.8 Summary

8.9 Important Terms

8.10 Formulas Introduced in This Chapter

Confidence Interval for the Mean (σ Not Known)

Confidence Interval for the Mean (σ Known)

Estimated Sample Size for the Confidence Interval for the Mean

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

8.11 Using SPSS
Calculating the confidence interval for the mean (The salary survey study (8.1))
8.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 9 Testing the Difference Between Two Means
9.1 An Example From the Research: You Can Just Wait
9.2 The Sampling Distribution of the Difference
Characteristics of the Sampling Distribution of the Difference
9.3 Inferential Statistics: Testing the Difference Between Two Sample Means
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: t-Test for Independent Means
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the t-Test for Independent Means
Summary
9.4 Inferential Statistics: Testing the Difference Between Two Sample Means (Unequal Sample Sizes)
An Example From the Research: Kids' Motor Skills and Fitness
Inferential Statistics: Testing the Difference Between Two Sample Means (Unequal Sample Sizes)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Relate the Result of the Analysis to the Research Hypothesis

Assumptions of the t-Test for Independent Means (Unequal Sample Sizes)

Summary

9.5 Inferential Statistics: Testing the Difference Between Paired Means

Research Situations Appropriate for Within-Subjects Research Designs

An Example From the Research: Web-Based Family Interventions

Inferential Statistics: Testing the Difference Between Paired Means

Calculate the Difference Between the Paired Scores

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Draw a Conclusion From the Analysis

Relate the Result of the Analysis to the Research Hypothesis

Assumptions of the t-Test for Dependent Means

Summary

9.6 Looking Ahead

9.7 Summary

9.8 Important Terms

9.9 Formulas Introduced in This Chapter

Degrees of Freedom (df), Difference Between Two Sample Means

t-Test for Independent Means

Standard Error of the Difference $\left(s_{\overline{X}_1-\overline{X}_2}\right)$

Standard Error of the Difference (Unequal Sample Sizes)

Degrees of Freedom (df), Difference Between Paired Means

t-Test for Dependent Means

Standard Error of Difference Scores (s D)

9.10 Using SPSS

Testing the Difference Between Two Sample Means: The Parking Lot Study (9.1)

Testing the Difference Between Paired Means: The Web-Based Intervention Study (9.5)

9.11 Exercises

Answers to Learning Checks

Learning Check 1

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Answers to Odd-Numbered Exercises
Chapter 10 Errors in Hypothesis Testing, Statistical Power, and Effect Size
10.1 Hypothesis Testing vs. Criminal Trials
10.2 An Example From the Research: Truth or Consequences
Correct Decisions and Mistakes in Criminal Trials
10.3 Two Errors in Hypothesis Testing: Type I and Type II Error
Type I Error: The Risk in Rejecting the Null Hypothesis
What Is the Probability of Making a Type I Error?
What Is the Probability of Not Making a Type I Error?
Why Is Type I Error a Concern?
Type II Error: The Risk in Not Rejecting the Null Hypothesis
What Is the Probability of Making a Type II Error?
What Is the Probability of Not Making a Type II Error?
Why Is Type II Error a Concern?
Why Does Type II Error Occur?
Type I and Type II Error: Summary
10.4 Controlling Type I and Type II Error
Controlling Type I Error
Concerns About Controlling Type I Error
Controlling Type II Error
Increasing Sample Size
Raising Alpha (α)
Using a Directional Alternative Hypothesis (H1)
Increasing Between-Group Variability
Decreasing Within-Group Variability
Controlling Type I and Type II Error: Summary
10.5 Measures of Effect Size
Interpreting Inferential Statistics: The Influence of Sample Size
The Inappropriateness of “Highly Significant”

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Effect Size as Variance Accounted for

Measure of Effect Size for the Difference Between Two Sample Means: r^2

Interpreting Measures of Effect Size

Presenting Measures of Effect Size

Reasons for Calculating Measures of Effect Size

A Second Measure of Effect Size for the Difference Between Two Sample Means: Cohen's d

10.6 Looking Ahead

10.7 Summary

10.8 Important Terms

10.9 Formulas Introduced in This Chapter

Probability of Type I Error

Probability of Type II error

r Squared (r^2)

Cohen's d

10.10 Exercises

Answers to Learning Checks

Learning Check 1

Learning Check 2

Learning Check 3

Learning Check 4

Answers to Odd-Numbered Exercises

Chapter 11 One-Way Analysis of Variance (ANOVA)

11.1 An Example From the Research: It's Your Move

11.2 Introduction to Analysis of Variance (ANOVA)

Review: Testing the Difference Between the Means of Two Samples

Understanding Between-Group and Within-Group Variability

Relating Between-Group and Within-Group Variability

Introduction to Analysis of Variance (ANOVA) and the F-Ratio

Characteristics of the F-Ratio Distribution

11.3 Inferential Statistics: One-Way Analysis of Variance (ANOVA)

State the Null and Alternative Hypotheses (H_0 and H_1)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Value, and State a Decision Rule
Calculate a Statistic: F-Ratio for the One-Way ANOVA
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (R^2)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Summary
11.4 A Second Example: The Parking Lot Study Revisited
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Value, and State a Decision Rule
Calculate a Statistic: F-Ratio for the One-Way ANOVA
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (R^2)
Draw a Conclusion From the Analysis
The Relationship Between the t-Test and the F-Ratio
11.5 Analytical Comparisons Within the One-Way ANOVA
Planned Versus Unplanned Comparisons
Between-Group and Within-Group Variance in Analytical Comparisons
Conducting Planned Comparisons
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Conducting Unplanned Comparisons
Concerns Regarding Unplanned Comparisons
Methods of Controlling Familywise Error

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Controlling Familywise Error: The Tukey Test

11.6 Looking Ahead

11.7 Summary

11.8 Important Terms

11.9 Formulas Introduced in This Chapter

Degrees of Freedom for Between-Group Variance, One-Way ANOVA (df_{BG})

Degrees of Freedom for Within-Group Variance, One-Way ANOVA (df_{WG})

Between-Group Variance, One-Way ANOVA (MS_{BG})

Total Sample Mean, One-Way ANOVA $\left(\overline{X}_T\right)$

Within-Group Variance, One-Way ANOVA (MS_{WG})

F-Ratio, One-Way ANOVA (F)

Measure of Effect Size, One-Way ANOVA (R^2)

Relationship Between t-Test and F-Ratio for the One-Way ANOVA

F-Ratio, Analytical Comparison (F_{comp})

Between-Group Variance, Analytical Comparison $\left(MS_{\mathrm{BG}_{\mathrm{comp}}}\right)$

Measure of Effect Size, Analytical Comparison $\left(R^2_{\mathrm{comp}}\right)$

Critical Value, Tukey Test for Unplanned Comparisons (F_T)

11.10 Using SPSS

One-Way Analysis of Variance (ANOVA): The Chess Study (11.1)

Analytical Comparisons: Planned Comparisons—Explanation vs. Prediction (11.5)

11.11 Exercises

Answers to Learning Checks

Learning Check 2

Learning Check 3

Learning Check 4

Answers to Odd-Numbered Exercises

Chapter 12 Two-Way Analysis of Variance (ANOVA)

12.1 An Example From the Research: Vote—or Else!

12.2 Introduction to Factorial Research Designs

Testing Interaction Effects

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

An Example of Interaction Effect

An Example of No Interaction Effect

Testing Main Effects

The Relationship Between Main Effects and Interaction Effects

Factorial Designs vs. Single-Factor Designs: Advantages and Disadvantages

12.3 The Two-Factor (A × B) Research Design

Notational System for Components of the Two-Factor Research Design

Analyzing the Two-Factor Research Design: Descriptive Statistics

Notational System for Means in the Two-Factor Research Design

Examining the Descriptive Statistics

12.4 Introduction to Analysis of Variance (ANOVA) for the Two-Factor Research Design

F-Ratios in the Two-Factor Research Design

12.5 Inferential Statistics: Two-Way Analysis of Variance (ANOVA)

State the Null and Alternative Hypotheses (H_0 and H_1)

Main Effect of Factor A

Main Effect of Factor B

A × B Interaction Effect

Make Decisions About the Null Hypotheses

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Values, and State Decision Rules

Calculate Statistics: F-Ratios for the Two-Way ANOVA

Make Decisions Whether to Reject the Null Hypotheses

Determine the Levels of Significance

Calculate Measures of Effect Size (R^2)

Draw Conclusions From the Analyses

Relate the Results of the Analyses to the Research Hypothesis

12.6 Investigating a Significant A × B Interaction Effect: Analysis of Simple Effects

12.7 Looking Ahead

12.8 Summary

12.9 Important Terms

12.10 Formulas Introduced in This Chapter

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Degrees of Freedom for Main Effect of Factor A, Two-Way ANOVA (df_A)
Degrees of Freedom for Main Effect of Factor B, Two-Way ANOVA (df_B)
Degrees of Freedom for A × B Interaction Effect, Two-Way ANOVA ($df_{A\times B}$)
Degrees of Freedom for Within-Group Variance, Two-Way ANOVA (dfWG)
Between-Group Variance for Main Effect of Factor A, Two-Way ANOVA (MS_A)
Between-Group Variance for Main Effect of Factor B, Two-Way ANOVA (MS_B)
Sums of Squares for Between-Group Variance, Two-Way ANOVA (SS_{BG})
Between-Group Variance for A × B Interaction, Two-Way ANOVA ($MS_{A\times B}$)
Within-Group Variance, Two-Way ANOVA (MS_{WG})
F-Ratio for Main Effect of Factor A, Two-Way ANOVA (F_A)
F-Ratio for Main Effect of Factor B, Two-Way ANOVA (F_B)
F-Ratio for A × B Interaction, Two-Way ANOVA ($F_{A\times B}$)
Measure of Effect Size for Main Effect of Factor A, Two-Way ANOVA(R^2_A)
Measure of Effect Size for Main Effect of Factor B, Two-Way ANOVA(R^2_B)
Measure of Effect Size for A × B Interaction, Two-Way ANOVA(R^2_{AB})
12.11 Using SPSS
Two-Way Analysis of Variance (ANOVA): The Voting Message Study (12.1)
12.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 13 Correlation
13.1 An Example From the Research: It's Good for You!
13.2 Introduction to the Concept of Correlation
Describing the Relationship Between Variables
Nature of the Relationship
Direction of the Relationship
Strength of the Relationship

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Measuring the Relationship Between Variables: Correlational Statistics
Introduction to the Pearson Correlation Coefficient (r)
Calculating Correlational Statistics: The Role of Variance and Covariance
13.3 Inferential Statistics: Pearson Correlation Coefficient
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: Pearson Correlation Coefficient (r)
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (r^2)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the Pearson Correlation
A Computational Formula for the Pearson r
Represent the Covariance Between X and Y (SPXY)
Represent the Variance of Variable X (SSX)
Represent the Variance of Variable Y (SSY)
Calculate the Pearson Correlation Coefficient (r)
13.4 Correlating Two Sets of Ranks: The Spearman Rank-Order Correlation
Introduction to Ranked Variables
An Example From the Research: Do Students and Teachers Think Alike?
Inferential Statistics: The Spearman Rank-Order Correlation Coefficient
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
13.5 Correlational Statistics vs. Correlational Research
Does ANOVA = Causality?
“Correlation Is Not Causality”—What Does That Mean?
13.6 Looking Ahead

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

13.7 Summary
13.8 Important Terms
13.9 Formulas Introduced in This Chapter
Degrees of Freedom for Pearson Correlation (df)
Sum of Products (Definitional Formula) (SPXY)
Sum of Squares for the X Variable (Definitional Formula) (SSX)
Sum of Squares for the Y Variable (Definitional Formula) (SSY)
Pearson Correlation (Definitional Formula) (r)
Measure of Effect Size for the Pearson Correlation Coefficient (r^2)
Sum of Products (Computational Formula) (SPXY)
Sum of Squares for the X Variable (Computational Formula) (SSX)
Sum of Squares for the Y Variable (Computational Formula) (SSY)
Pearson Correlation (Computational Formula) (r)
Spearman Rank-Order Correlation (rs)
Measure of Effect Size for the Spearman Rank-Order Correlation $\left(r_s^2\right)$
13.10 Using SPSS
Pearson Correlation: The Picky Eater Study (13.1)
Spearman Rank-Order Correlation: The Business Student Study (13.4)
13.11 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Answers to Odd-Numbered Exercises
Chapter 14 Linear Regression and Multiple Correlation
14.1 Predicting One Variable From Another: Linear Regression
The Linear Regression Equation
Calculating the Linear Regression Equation
Calculate the Slope of the Equation (b)
Calculate the Y-Intercept of the Equation (a)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Report the Linear Regression Equation

Drawing the Linear Regression Equation

A Second Example: Put Down Your Pencils

Calculating the Linear Regression Equation

Drawing the Linear Regression Equation

14.2 Correlation With Two or More Predictors: Introduction to Multiple Correlation and Regression

The Limitations of Using Only One Predictor

Introduction to Multiple Correlation

Multiple Correlation With Two Predictors: An Example

Measuring the Relationship Between Two Predictors and a Criterion: The Multiple Correlation (R)

Testing the Multiple Correlation for Statistical Significance

Predicting a Criterion From Two Predictors: The Multiple Regression Equation

Conclusion

14.3 Looking Ahead

14.4 Summary

14.5 Important Terms

14.6 Formulas Introduced in This Chapter

Linear Regression Equation

Estimate of Slope (b)

Estimate of Y-Intercept (a)

Multiple Correlation Coefficient (R)

F-Ratio for the Multiple Correlation Coefficient (F)

Multiple Regression Equation

Regression Coefficient

Estimate of Y-Intercept (a) (Multiple Regression)

14.7 Using SPSS

Linear Regression: The Picky Eater Study (14.1)

Multiple Regression: The College Admission Example (14.2)

14.8 Exercises

Answers to Learning Checks

Learning Check 1

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 2

Answers to Odd-Numbered Exercises

Chapter 15 Chi-Square

15.1 An Example From the Research (One Categorical Variable): Are You My Type?

15.2 Introduction to the Chi-Square Statistic

Observed and Expected Frequencies: The Basis of the Chi-Square Statistic

Characteristics of the Distribution of Chi-Square Statistics

15.3 Inferential Statistic: Chi-Square Goodness-of-Fit Test

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Value, and State a Decision Rule

Calculate a Statistic: Chi-Square (χ^2)

Make a Decision Whether to Reject the Null Hypothesis

Determine the Level of Significance

Calculate a Measure of Effect Size (Cramér's φ)

Draw a Conclusion From the Analysis

Relate the Result of the Analysis to the Research Hypothesis

Assumptions Underlying the Chi-Square Statistic

Conducting the Chi-Square Goodness-of-Fit Test With Unequal Hypothesized Proportions

Calculate Expected Frequencies (f_e)

Calculate the Chi-Square Statistic (χ^2)

15.4 An Example From the Research (Two Categorical Variables): Seeing Red

15.5 Inferential Statistic: Chi-Square Test of Independence

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Value, and State a Decision Rule

Calculate a Statistic: Chi-square (χ^2)

Make a Decision Whether to Reject the Null Hypothesis

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Determine the Level of Significance
Calculate a Measure of Effect Size (Cramér's φ)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
15.6 Parametric and Nonparametric Statistical Tests
Parametric vs. Nonparametric Statistical Tests
Reasons for Using Nonparametric Statistical Tests
Examples of Nonparametric Statistical Tests
15.7 Looking Ahead
15.8 Summary
15.9 Important Terms
15.10 Formulas Introduced in This Chapter
Degrees of Freedom for the Chi-Square Goodness-of-Fit Test
Expected Frequency for the Chi-Square Goodness-of-Fit Test
Chi-Square Statistic (χ^2)
Cramér's φ for the Chi-Square Goodness-of-Fit Test
Degrees of Freedom for the Chi-Square Test of Independence
Expected Frequency for the Chi-Square Test of Independence
Cramér's φ for the Chi-Square Test of Independence
15.11 Using SPSS
Chi-Square Goodness-of-Fit Test—Equal Hypothesized Proportions: The Personality Study (15.1)
Chi-Square Goodness-of-Fit Test—Unequal Hypothesized Proportions: The Personality Study (15.3)
Chi-Square Test of Independence: The Color Preference Study (15.4)
Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Learning Check 6

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Odd-Numbered Exercises

Appendix A: Tables

Appendix B: Review of Basic Mathematics

I. Symbols

II. Terms

III. Addition, Subtraction, Multiplication, Division

A. All Positive Numbers

B. Adding All Negative Numbers

C. Adding a Combination of Positive and Negative Numbers

D. Subtracting Negative Numbers

E. Multiplying Two Negative Numbers

F. Multiplying a Combination of Positive and Negative Numbers

G. Dividing Two Negative Numbers

H. Dividing One Positive Number and One Negative Number

IV. Order of Operations

V. Fractions, Decimals, Percentages

A. Fraction

B. Decimal

C. Percentage

VI. Solving Simple Equations With One Unknown (X)

A. X Has a Value Subtracted From It (Solve for X Using Addition)

B. X Has a Value Added to It (Solve for X Using Subtraction)

C. X Is Multiplied by a Value (Solve for X Using Division)

D. X Is Divided by a Value (Solve for X Using Multiplication)

VII. Solving More Complex Equations With One Unknown (X)

A. X Is Multiplied by a Value and Then Has a Value Either Added to It or Subtracted From It (Solve for X Using Either Subtraction or Addition and Then Division)

B. X Is Divided by a Value and Then Has a Value Either Added to It or Subtracted From It (Solve for X Using Either Addition or Subtraction and Then Multiplication)

C. X Has a Value Added to It or Subtracted From It and Is Then Multiplied by a Value (Solve for X Using Division and Then Either Subtraction or Addition)

D. X Has a Value Either Added to It or Subtracted From It and Is Then Divided by a Value (Solve for X Using Multiplication and Then Either Subtraction or

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Addition)
Glossary
References
Index

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

PREFACE

A Few (?) Words About Me

When I entered college, I thought my path would take me to law school; however, taking an Intro to Psych class made me realize I enjoy the challenge of trying to understand the human mind. The college I attended, U.C. Santa Cruz, was a fairly unconventional university, but somehow in the midst of a sea of humanistic psychologists I became attracted to the empirical and methodological aspects of psychology by taking classes from instructors such as David Harrington and Dane Archer, who showed me that statistics could appeal to students if taught with a gentle, guiding approach that addresses questions relevant to students' lives. After graduating from college, I was able to get a job as a research assistant for a human resource consulting firm. Despite my lack of work experience, I was hired primarily as a result of having taken statistics and research methods courses, which taught me that learning statistics has benefits both inside and outside of the classroom.

Several years later, I started grad school at U.C. Berkeley, where two events critical to this book took place. First, serving as a teaching assistant, I found I really enjoyed helping students, particularly in statistics and research methods classes often approached with fear and suspicion. Second, I took graduate classes from Geoff Keppel, who had developed his own unique system for analyzing experimental research designs. His lectures and books showed me that statistics can be taught in a systematic way that highlights similarities rather than differences between different research situations. Geoff managed to transform something as daunting-sounding as a "3 x 2 x 4 research design" into the mathematical equivalent of playing with wooden toy alphabet blocks labeled "A," "B," and "C."

After leaving Cal, I was hired by San Jose State, where I teach an introductory statistics course aimed at students with a wide range of backgrounds, ability, and motivation. When I went to select a textbook for the course, for the first time I looked carefully at the wide range of offerings. What I found striking (and still find striking) was that the majority of books focused on providing formulas and very small sets of data designed to demonstrate how to correctly calculate the correct numbers from these formulas. Little emphasis, however, was given to what these numbers meant or implied. Given my own experiences learning and using statistics, I thought a book was needed that focusses on how statistics may be used to answer questions and test ideas within the larger research process.

Goal of this Book

This textbook, aimed primarily at students majoring in the behavioral and social sciences, includes many of the same topics and chapters as the majority of existing introductory statistics books. However, what I believe distinguishes this book is its thematically-based approach that places statistics within the research process. The primary purpose of this

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

book is to not just teach students how to calculate statistics but how to interpret the results of statistical analyses in light of a study's research hypothesis, and to communicate one's results and interpretations to a broader audience. Hopefully, this book will not only help students understand the purpose and use of statistics, but it will also give them a greater appreciation of how research studies are conceived, conducted, and communicated.

The book is intended for students with a limited knowledge of mathematics, statistics, and research design and methodology. Some common obstacles faced by students are their math anxiety, fear of conducting arithmetic calculations, and an inability to see the relevance of the material to their daily lives. To address these concerns, this books attempts to judiciously use mathematical terminology, write in a conversational manner, and utilize actual research situations rather than contrived sets of numbers.

Organization of the Book

The 15 chapters of this book may be placed into three general categories. Chapters 1 to 4 are designed to introduce students to the research process and how data that have been collected may be organized, presented, and summarized:

Chapter 1 (Introduction to Statistics) defines the concept of statistics, explains why statistics are important to learn, and describes the steps in the research process that are the framework for the rest of the book.

Chapter 2 (Examining Data: Tables and Figures) discusses the importance of examining data that have been collected and describes how tables and figures may be created to represent data.

Chapter 3 (Measures of Central Tendency) presents measures of central tendency (the mode, median, and the mean) and compares them in terms of their relative strengths and weaknesses.

Chapter 4 (Measures of Variability) presents measures of variability (range, interquartile range, variance, standard deviation, index of qualitative variation) and the concept of parameters (population variance and standard deviation).

Chapters 5 through 10 discuss the process of conducting statistical analyses to test research questions and hypotheses, as well as issues and controversies regarding this process:

Chapter 5 (Normal Distributions) introduces theoretical distributions, starting with normal distributions, before moving on to the standard normal distribution (normal curve).

Chapter 6 (Probability and Introduction to Hypothesis Testing) presents the structure for later chapters by introducing, defining, and illustrating steps in hypothesis testing.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Chapter 7 (Testing One Sample Mean) tests one sample mean using the *z* -test and *t* -test and introduces sampling distributions, the central limit theorem, and the Student *t* distribution.

Chapter 8 (Estimating the Mean of a Population) introduces interval estimation by developing the confidence interval for the mean and discusses interval estimation as an alternative to null hypothesis significance testing (NHST).

Chapter 9 (Testing the Difference Between Two Means) covers the *t* -test for independent means for both equal and unequal sample size and introduces within-subjects research designs using the *t* -test for dependent means.

Chapter 10 (Errors in Hypothesis Testing, Statistical Power, and Effect Size) discusses Type I and Type II error and introduces measures of effect size as a way of drawing appropriate conclusions about an analysis.

Chapters 11 to 15 discuss different statistical procedures used in research situations that vary in the number of independent variables as well as how the variables have been measured:

Chapter 11 (One-way Analysis of Variance (ANOVA)) discusses the one-way ANOVA, the relationship between the *t* -statistic and the F-ratio, and planned and post-hoc comparisons.

Chapter 12 (Two-way Analysis of Variance (ANOVA)) introduces factorial research designs, the calculation of F-ratios for main effects and interaction effects, and illustrates the use of simple effects analyses to test research hypotheses.

Chapter 13 (Correlation) defines and illustrates the concept of correlation, discusses the Pearson and Spearman correlation coefficients, and makes the distinction between correlational statistics and correlational research methods.

Chapter 14 (Linear Regression and Multiple Correlation) introduces the concept of regression, discusses linear regression, and works through an example of multiple regression with two predictor variables.

Chapter 15 (Chi-Square) describes the analysis of categorical variables using the chi-square goodness-of-fit test and the chi-square test of independence before ending the book with a brief introduction to nonparametric statistics.

In most cases, the chapters in this book share a uniform format and structure, centered on the research process, based on examples of actual published research studies addressing topics relevant to students' lives. For each example, the research hypothesis is clearly identified; after a brief description of how the data in the study were collected, the calculation and interpretation of both descriptive and inferential statistics are presented,

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ending with the extent to which the results of the statistical analyses support the research hypothesis and how these findings may be communicated to a broader audience.

Pedagogical Features

The chapters in this textbook contain several reoccurring features designed to enhance students' comprehension of the material. These include the following:

Periodic learning checks within chapters. Short learning checks, in the form of conceptual review questions and computational exercises, are provided throughout each chapter.

Explicit presentation of formulas and calculations. This book provides both definitional formulas to highlight their relation to the concept being measured and computational formulas to facilitate calculations. In applying these formulas, all of the steps in the calculation of statistics such as the *t* -test and F-ratio are included to help students in completing exercise and homework assignments.

Illustration of the presentation and communication of analyses. Every chapter includes a discussion and illustration of how statistical analyses may be reported as well how they may be communicated to a larger audience in both written and visual form.

Highlighted definition of terms. Every term and concept introduced in this book is boldfaced, immediately followed by an explicit definition.

End of chapter summaries and exercises. The end of each chapter contains a summary of critical points, a list of terms and formulas introduced in the chapter, and a large number of exercises, many of which refer to published research studies.

Calculation of statistics using SPSS. Many chapters include end-of-chapter content that provides an annotated presentation of how statistics may be calculated using the SPSS software. Screen shots of data files, menu selections, and output files are included and annotated to locate and identify relevant aspects of the analysis.

What's New to the 3rd Edition

I'd like to start by expressing my most heartfelt appreciation to the instructors and reviewers of the first two editions of this book as their evaluation, insights, criticisms, and suggestions were absolutely vital to the changes made in this third edition. I'd like to particularly thank the instructors who've adopted the book who were kind enough to spontaneously send me their thoughts and suggestions. It is my sincere hope that all of you see the incorporation of your input in this edition.

There were two main goals in this edition. The first goal was to make the writing as brief and concise as possible, manifested in the goal of reducing the length of each chapter by 5% to

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

10%. This goal was achieved by simplifying the language and shortening or eliminating what I, the book's adopters, and the book's reviewers felt was unnecessary detail or discussion. The second goal was to increase and update the exercises at the end of many of the chapters, including the use of recent published research as the basis of these exercises. Another goal was to update the demonstration of the use of SPSS to analyze data using the most recent version (Version 31 at the time of this book's publication).

More specifically, in terms of content, the following changes were made to these chapters:

Chapter 1 (Introduction to Statistics): The published research studies used to illustrate concepts in the text, learning checks, and exercises were added to and updated using topics relevant to students (e.g., fear of missing out (FOMO), cryptocurrency, legalization of cannabis, Gen Z, COVID, fake news).

Chapter 6 (Probability and Introduction to Hypothesis Testing): The Super Bowl example used to introduce and illustrate the process of hypothesis testing was updated.

Chapter 8 (Estimating the Mean of a Population): The main research example, which involves data from an actual national salary survey, has been updated throughout this chapter with more current salary data.

As with the first two editions, any and all questions, concerns, suggestions, and complaints you have about the book are most welcome and can be sent to me at howard.tokunaga@sjsu.edu .

A Few Tips for Students

To you, the college student about to read this book as part of taking a statistics course: "Welcome!" and "Great job!" We welcome you because you're about to embark on a semester-long journey that we hope enhances your skills and widens your perspective; we congratulate you because it's a journey not everyone is willing to take. We say to our students that by taking just one statistics course they are ahead of 95% of the population in their knowledge of statistics.

We also tell our students that our primary goal for them is not to become a statistics genius but just to become comfortable with the process of analyzing and interpreting data. In that regard, we agree with a 2012 *Forbes* magazine article entitled *The Six Classes That Will Make Any College Grad Employable* that noted "a good statistics class will help the student become comfortable pushing numbers around and avoiding false interpretations."

At the present moment, we know our encouragement may be of little comfort because you might be anxious about having to learn statistics. Some of you might be anxious about having to learn statistical concepts and formulas; others might be anxious because the research process seems pretty complicated. Talking with students who have taken our courses over the years has helped us assemble the following advice:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Master the Material Presented in the Early Chapters

Chapters 1 through 6 discuss how research is typically conducted, how data are summarized and described, and the process by which researchers conduct statistical analyses to test their ideas. It is absolutely critical for you to have a firm understanding of these chapters as they lay the groundwork for later chapters that discuss a variety of statistical procedures used by researchers to test hypotheses.

What does it mean to master this material? First, *read the chapters both before and after they're discussed in class* . Reading the chapters beforehand helps you identify anything that's unclear and have your questions addressed by your instructor; re-reading the chapters afterward helps confirm your understanding of the material. Next, *be able to define and explain key conceptspresented in these chapters* . These concepts are highlighted in bold-faced letters, and it's important to learn them when they're first introduced because they'll appear throughout the remainder of the book. Next, *do the learning checks within the chapters and the exercises at the end of the chapters* , because they include exercises and review questions you can ask yourself to assess your understanding of the material. Most importantly, *do not miss class during the early part of the semester* ! It's been our experience that students who miss critical lectures at the beginning of the term often have difficulty keeping up as the semester continues.

Review High School Algebra

If you happen to move beyond the introductory statistics course for which you're reading this book, you'll find that statistics are often calculated using computers and statistical software rather than hand calculations. Consequently, some statistics textbooks no longer include mathematical formulas but instead have students use statistical software or Microsoft Excel. However, this book takes a different approach for two main reasons. First, we believe that in order to understand the results of statistical analyses, one must understand their underlying foundation. Learning statistics via computer software can lead to a "brain-dead" approach in which students are at the mercy (rather than control) of their computers. We refer to this as the "I'm only as smart as my printout" method of learning statistics, also known as "Because this is what my computer told me."

Second, this book contains mathematical formulas and calculations for one simple reason: The statistics in this book are not hard to calculate. To assess whether you're prepared to read this book, take the following test:

1. Do you know how to add, subtract, multiply, and divide?
2. Do you know how to calculate the square or square root of a number?
3. Do you understand the order of operations and parentheses within mathematical formulas?

If your answers to these questions are "Yes," congratulations! You possess the ability to conduct *every* mathematical calculation in this book because *none* of the formulas involve geometry, trigonometry, or calculus. However, if you're unsure of your mathematical ability, we suggest you review your high school algebra. Also, there's an appendix at the end of this

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

book that includes the mathematical concepts and operations needed to conduct the statistical analyses in this book. We highly recommend you read this appendix to review and assess your mathematical skills.

Learn to Use a Statistical Calculator

Although learning the statistics in this book doesn't require the use of computers or software, you will need a hand calculator. Calculators with statistical capabilities are typically identified as "statistical" or "scientific" calculators. Although there's a wide selection of brands and models, know that the calculators most appropriate for this book cost only $10 to $15. We do *not* recommend you use or purchase a more expensive calculator! We prefer these simpler calculators because students are able to do their calculations more quickly and with fewer errors. We do suggest you purchase a calculator similar to the one your instructor uses as they will be more familiar with its features and idiosyncrasies.

In selecting a calculator, look closely at the keyboard; you'll need a calculator whose keys have labels such as $\bar{X}$, σ, and ΣX (Chapters 3 and 4 discuss what these symbols represent). When you go calculator shopping, you'll find calculators that have keys with labels such as ΣY and XY. Although calculators with these Y symbols are only slightly more expensive than calculators with just the X symbols, we do *not* recommend their purchase because these features are not needed to perform the calculations in this book. In fact, we've found students with Y calculators have a more difficult time performing simpler calculations.

Ask Questions

Conducting research is the process of asking and answering questions. Accordingly, throughout this book issues are often framed in the form of questions, which we hope encourages you to ask questions as well. The first person you should direct any questions to is yourself. As you work your way through the chapters, get into the habit of asking yourself, "What does this mean?" A good test of your level of comprehension is whether you can explain the material to yourself in a meaningful way.

You should also be sure to work with your instructor to confirm your level of understanding and clarify unanswered questions. Instructors often say to students, "The only bad question is the one you don't ask." Don't be hesitant about asking questions in class. Instructors will tell you they use one student's question to assess an entire class's level of understanding. This is what we call the "pencil test"—when one student asks a question, we look to see how many other pencils get raised, which reflects the number of students who had the same question. Asking questions enhances the learning experience for yourself, your classmates, and your instructor.

Form a Study Group

No matter how much we encourage students to ask questions in class, it seems they hesitate for fear of drawing attention to themselves. As a result, we turn to another well-

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

traveled saying: “Misery loves company.” Forming a study group with other students will help you keep up with reading assignments, confirm your comprehension of the material, clarify any confusion you have, prepare for exams, and provide social support. We do suggest that your group contain at least one person who will raise the group’s questions and concerns to the instructor during class or office hours.

Practice, Practice, Practice

The best way to learn the topics covered by this book is to obtain as much practice as possible. Throughout the chapters are examples in which the calculations have been worked out for you; in the middle and end of each chapter are a number of exercises. The answers to some of the exercises are provided at the back of this book, and your instructor has access to the answers to the other exercises. Do the end-of-chapter exercises even if your instructor does not assign them as homework in order to identify any reoccurring mistakes you make. We find that the vast majority of errors students make are simple computational errors, committed when students do the calculations too quickly.

Also note that the publisher of this book, Sage Publications, has developed a webpage for this book that contains a huge amount of supplemental tools such as videos and quizzes for each chapter. Find it by going to the Sage Publishing website **collegepublishing.sagepub.com** , searching for my name (Tokunaga), and looking for the **Resources** tab.

Enough cheerleading and preaching – let’s get started!

WELCOME!

GREAT JOB!

Additional Teaching Resources

Visit **collegepublishing.sagepub.com** , and navigate to the Resources tab on your book’s page to find the teaching materials designed to accompany this textbook. On this site you will find an array of materials that will save you time and to help you keep students engaged, including the following:

Learning management system cartridges that easily integrate with your course management system so student test results and graded assignments seamlessly flow into your gradebook

Test banks , aligned to Bloom’s Taxonomy, that provide a diverse range of test items, including multiple choice, true/false, and essay questions

Lecture notes that provide an outline and the key concepts in each chapter to aid in lecture preparation

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

PowerPoint ® slides that offer a flexible, accessible, and customizable solution for creating multimedia lectures

Tables and Figures from the book to support lecture preparation and class discussions

Sample course syllabi , including suggested models for structuring your course

Answers to even-numbered questions from the text help facilitate grading

SPSS datasets for use with exercises from the text.

Multimedia content , including original SAGE videos featuring tutorials with author Howard T. Tokunaga that bring concepts to life and appeal to diverse learners

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ACKNOWLEDGMENTS

This book has traveled a long journey, with many people playing both transient and persevering roles in its development. First and foremost, I'd like to thank my kids, Meagan and Will, for providing me perspective and inspiration amid all of the twists and turns we faced while this book was being written. I'm so very blessed to have you as my "burden." Away from home, I'm grateful to my graduate students at San Jose State—working on this book has given me much empathy for their thesis-related travails. However, they should be aware that the completion of this book ends the unspoken agreement that I not ask, "So how's your thesis?" if they don't ask, "So how's your book?" A number of friends have provided most welcome relief after long days staring at a computer screen, with particular thanks going to Ellie, Eileen, Susan, Bonnie, Ann, Lisa, and Terry. I would like to thank Charles Linsmeier and David Shirley for their many helpful comments and advice on earlier versions of this book, Julie West for her help in developing chapter exercises, and Youn-Jeong Choi and Naomi Ran for their careful examination of the chapters and exercises.

I would like to express my appreciation to everyone at SAGE who has played a role in shepherding this book to completion. First and foremost, I must express my gratitude to my editor, Vicki Knight. It has been a joy and an honor to work with someone with such a truly exceptional combination of experience, knowledge, perspective, and humor. In production I was fortunate enough to work with others on the SAGE team who made the process a pleasure: Yvonne McDuffee, editorial assistant; Katie Guarino, assistant editor; Nicole Elliott, executive marketing manager; Candice Harman, cover designer; and the copyeditor, Michelle Ponce. Furthermore, I am grateful to the reviewers who took the time to read the chapters, raise questions, and provide constructive comments and suggestions to assist me in clarifying and strengthening the manuscript. The following are the reviewers for third edition:

Erin M. Fekete, University of Indianapolis

Brooke Magnus, Boston College

Loretta Neal McGregor; Arkansas State University

Michael T. Moore, Adelphi University

Megan M. Ringel, St. Edward's University

Bettina Viereck, University of Hartford

Janice Weaver, Community College of Allegheny County

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ABOUT THE AUTHOR

Howard T. Tokunaga is professor of Psychology at San Jose State University, where he serves as coordinator of the MS Program in Industrial/Organizational (I/O) Psychology and teaches undergraduate and graduate courses in statistics, research methods, and I/O psychology. He received his bachelor's degree in psychology at U.C. Santa Cruz and his PhD in psychology at U.C. Berkeley. In addition to his teaching, he has consulted with a number of public sector and private sector organizations on a wide variety of management and human resource issues. He is author of *Moving from IBM SPSS to R and RStudio: A Statistics Companion* , and coauthor (with G. Keppel) of *Introduction to Design and Analysis* : *A Student's Handbook* . In his spare time, he enjoys the outdoors, museums, live theater, and being as physically, socially, and politically active as possible.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CHAPTER 1 INTRODUCTION TO STATISTICS

Learning Objectives

1.1 Define the concept of statistics.

1.2 Describe different reasons for learning statistics.

1.3 Describe the main steps in the research process and the tasks and concepts within each step.

In introducing this book to you, we assume you're a college student taking what is perhaps your first course in statistics to fulfill a requirement for your major or a general education requirement. If so, you may be asking yourself two questions:

What is statistics?

Why learn statistics?

The ultimate goal of this book is to help you begin to answer these two questions.

1.1 What Is Statistics?

You may not know it, but you encounter a variety of "statistics" in your day-to-day activities: the typical cost of going to college, the yearly income of the average college graduate, the average price of a car, and so on. So what exactly is "statistics"?

The Merriam-Webster dictionary defines **statistics** as *a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data* . When most people think about statistics, they focus on the "analysis" aspect of this definition—that is to say, they focus on numbers that result from analyzing data. However, statistics not only is concerned about how data are *analyzed* but also how data are *collected* and how the results of statistical analyses are *interpreted* and *communicated* . The main purpose of this book is to introduce, describe, and illustrate the role of statistics within the larger research process.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1.2 Why Learn Statistics?

There are many reasons why you should learn statistics. First, not only do you currently encounter statistics in your daily activities, but *throughout your life you are affected by the results of research and statistical analyses* . Which college you attend is based in part on test scores developed by researchers. You may have to take a personality or intelligence test to get a job. The choices of drugs and medicines available to you are based on medical research. Learning about statistics will help you become a more informed and aware consumer of research and statistical analyses that affect many aspects of your life.

A second reason for learning statistics is that *you may be required to interpret the results of statistical analyses* . Many college courses require students to read academic research journal articles. Evaluating published research is complicated by the fact that different people studying the same topic may come up with diverse or even opposing conclusions. Understanding statistics and their role in the research process will help you decide whether conclusions drawn in research articles are appropriate and justified.

Another reason for learning statistics is that *you might be required to do your own research* . College courses sometimes have students design and conduct mini-research studies; undergraduate majors might encourage students to do senior honors theses; graduate programs often require masters' theses and doctoral dissertations. Learning to collect and analyze data will help you address your own questions in an objective, systematic manner.

A final reason for learning statistics is that it may *help you in your future career* . The website Careercast.com conducts an annual survey in which they evaluate jobs on four dimensions: environmental factors, income, employment outlook, and stress. Recently, their #1 rated job was "data scientist," and the #2 rated job was "statistician." They noted, "A statistician's skill set can be used to break down and analyze large quantities of data. The demand for these skills spans a variety of industries, including marketing, banking, government, sports, retail, and even healthcare."

It's generally a good idea for students to maintain a healthy level of curiosity or even skepticism in regard to their education. However, we find that when it comes to learning statistics, students' feelings may be characterized as fear or anxiety. We understand these feelings, and we hope the benefits of learning statistics will become clear to you and help overcome any concerns you have.

1.3 Introduction to the Stages of the Research Process

Much of scientific research involves asking questions. Throughout this book, we'll examine a broad range of questions regarding human attitudes and behavior that contemporary researchers have asked and attempted to answer. The following are research questions we'll address in this chapter to introduce the stages of the research process:

> Does appealing to one's fear of missing out (FOMO) affect the intention to invest in cryptocurrencies?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Do pro-cannabis messages on social media lead to cannabis use?

Do parents differ in their awareness of their children's online activities?

Do nonbinary college students have the same experiences as cis men or cis women?

How might you try to answer questions such as these? You could base your answers on your personal beliefs, or you could adopt the answers given to you by others. But rather than relying on subjective beliefs and feelings, researchers in many academic disciplines test their ideas using the scientific method. The **scientific method** is a method of investigation that uses the objective and systematic collection and analysis of empirical data to test theories and hypotheses.

At its simplest, this book will portray the scientific method as consisting of five main steps or phases:

developing a research hypothesis to be tested,

collecting data,

analyzing the data,

drawing a conclusion regarding the research hypothesis, and

communicating the findings of the study.

Accomplishing each of these five steps requires completing a number of tasks, as shown in Figure 1.1 . Because this sequence of steps serves as the model for the research process we'll discuss throughout this book, each step is briefly introduced in the following sections. It's important to understand that the research process depicted in Figure 1.1 represents an ideal way of doing research. The "real" way, as you may discover in your own efforts, is often anything but a smooth ride but rather is filled with false starts, dead ends, and wrong turns.

Developing a Research Hypothesis to Be Tested

The initial stage—and the first step—of the research process is to develop a research hypothesis to be tested. A **research hypothesis** may be defined as a statement regarding an expected or predicted relationship between variables. A **variable** is a property or characteristic of an object, event, or person that can take on different values. One example of a variable is "U.S. state," a variable with 50 possible values (Alabama, Arkansas, etc.).

Research hypotheses are usually developed through the completion of several tasks:

identifying a question or issue to be examined,

reviewing and evaluating relevant theories and research, and

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

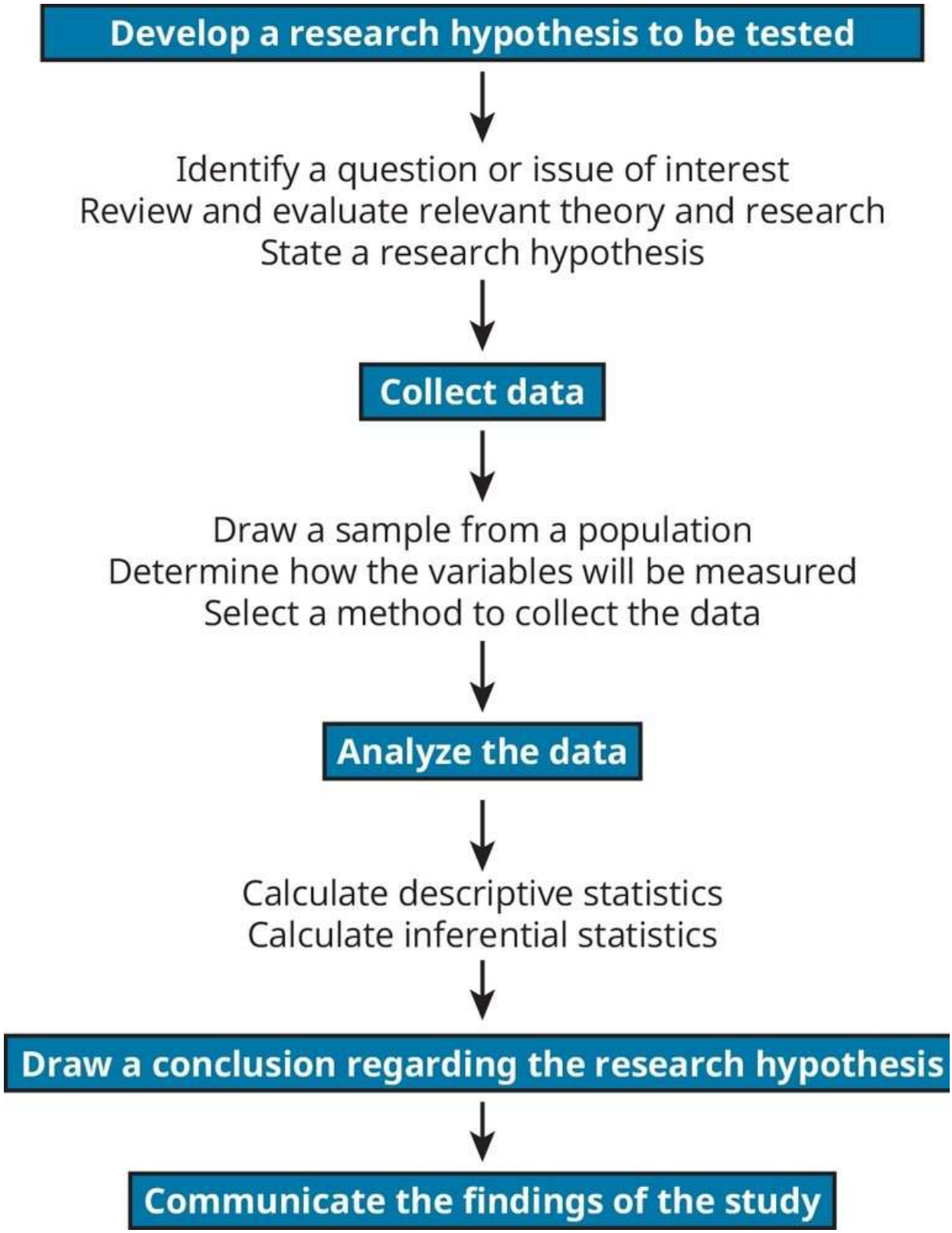
stating a research hypothesis.

Identifying a Question or Issue to Be Examined

Most research starts with a question posed by the researcher that often comes from the researcher's own ideas and daily observations. Although this may not seem terribly scientific, there's an advantage in using one's own experience as a starting point: People are generally much more motivated to explore issues that concern them personally. In teaching statistics, we advise students developing their own research projects to study something of interest to them. Conducting research can be tedious, difficult, and frustrating. At various points, you may ask yourself, "Why am I doing this?" Being able to provide a satisfactory answer to this question will help you overcome obstacles you encounter along the way.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

▶ Long Description: Figure 1.1 Steps in the Research Process Within the Scientific Method

Figure 1.1 Steps in the Research Process Within the Scientific Method

Reviewing and Evaluating Relevant Theories and Research

Beyond your own curiosity, research questions often arise from studying the theories, ideas, and research of others. A **theory** is a set of propositions used to describe or explain a phenomenon. The purpose of a theory is to summarize and explain specific facts using a logically consistent framework. Placing a question within a theoretical framework provides guidance and structure to research.

Reviewing and evaluating existing theories and research helps you decide whether it's worth the time and energy required to conduct a study. By seeing what others have done, you may conclude that the current way of thinking on a topic is incomplete or mistaken. By doing this review, you're able to ensure that the research you conduct adds to and improves upon an existing body of knowledge.

Stating a Research Hypothesis: Independent and Dependent Variables

Understanding and evaluating an existing literature not only helps develop a question of interest, but it may also lead to a predicted answer to that question. Within the scientific method, this answer is stated as a research hypothesis, defined earlier as a statement regarding an expected or predicted relationship between variables. Table 1.1 lists the research questions and research hypotheses for the studies listed at the beginning of this section. For example, the first research hypothesis states that "consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos."

One characteristic of research hypotheses such as those listed in Table 1.1 is that they identify the variables that are the focus of their research. As mentioned earlier, a variable is a property or characteristic with different values. Variables can be classified in several ways. In specifying a research hypothesis, researchers often speak in terms of "independent" and "dependent" variables. An **independent variable** may be defined as a variable manipulated or controlled by the researcher. A **dependent variable** , on the other hand, is a variable measured by the researcher that is expected to change or vary as a function of the independent variable. The independent variable is seen as the "cause" or "treatment," whereas the dependent variable is the "effect" or "outcome."

Researchers are interested in examining the effect of the independent variable on the dependent variable. Consider the first research hypothesis provided in Table 1.1 : "Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos." Here, the independent variable is the type of appeal consumers are exposed to: FOMO or non-FOMO. The dependent variable is the intention to invest in cryptos. In this study, the effect of the independent variable on the dependent variable is that

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

differences in the intention to invest in cryptos (the dependent variable) may “depend” upon which type of appeal (FOMO or non-FOMO) a consumer receives. Table 1.2 lists the independent and dependent variables for each of the research hypotheses in Table 1.1 .

In addition to identifying variables, a second characteristic of research hypotheses is that they specify the nature and direction of the relationship between variables. For example, the first research hypothesis in Table 1.1 states that “Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *higher* intention to invest in cryptos.” The word *higher* indicates the nature and direction of the relationship between the independent variable (type of appeal) and the dependent variable (intention to invest in cryptos). The direction of the relationship would not have been stated if the hypothesis had included the less specific phrase “will have a *different* intention,” which simply indicates that the two groups of consumers are not expected to be the same. Table 1.3 provides directional and nondirectional research hypotheses for the research studies from Table 1.1 .

Table 1.1 Examples of Research Questions and Hypotheses

Research Question	Research Hypothesis
Does appealing to one’s fear of missing out (FOMO) affect the intention to invest in cryptocurrencies? (Friederich et at., 2024)	Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.
Do pro-cannabis messages on social media lead to cannabis use? (Willoughby et al., 2024)	Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.
Do parents differ in their awareness of their children’s online activities? (Symons et al., 2017)	Mothers have more accurate knowledge of their children’s online activities than do fathers.
Do nonbinary college students have the same experiences as cis men or cis women? (Marx et al., 2024)	Nonbinary college students experience higher levels of victimization than do cis men or cis women.

The ability to form a directional research hypothesis is dependent on the state of the literature on the question of interest. If little or perhaps conflicting research has been conducted, researchers may not be able to form a directional hypothesis. For example, one study examining the relationship between exercise deprivation (not allowing people to exercise) and depression. reported, “We did not have a directional hypothesis, and when participants asked what we expected to find in this study, we replied: ‘We really are not sure since the results of earlier work on exercise deprivation are mixed’” (Mondin et al., 1996, p. 1200).

Table 1.2 Research Hypotheses and Their Independent and Dependent Variables

Research Hypothesis	Independent Variable	Dependent Variable
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher	Type of appeal	Intention to invest in cryptos

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Independent Variable	Dependent Variable
intention to invest in cryptos.		
Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.	Exposure to pro-cannabis messages	Intention to use cannabis
Mothers have more accurate knowledge of their children's online activities than do fathers.	Parent	Knowledge of children's online activities
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	Gender identity	Amount of victimization

Collecting Data

Once a research hypothesis has been formulated, researchers are ready to proceed to the second stage in the research process: collecting data. This step is seen as being composed of three tasks:

drawing a sample from a population,

determining how the variables will be measured, and

selecting a method by which to collect the data.

Table 1.3 Directional and Nondirectional Research Hypotheses

Directional Research Hypothesis	Nondirectional Research Hypothesis
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *higher* intention to invest in cryptos.	Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *different* intention to invest in cryptos.
Exposure to pro-cannabis messages on social media is related to *increased* intentions to use cannabis.	Exposure to pro-cannabis messages on social media is *related to* intentions to use cannabis.
Mothers have *more* accurate knowledge of their children's online activities than do fathers.	Mothers have *different* knowledge of their children's online activities than do fathers.
Nonbinary college students experience *higher* levels of victimization than do cis men or cis women.	Nonbinary college students experience *different* levels of victimization than do cis men or cis women.

Drawing a Sample From a Population

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The first step in collecting data is to identify the group of participants to which the research hypothesis applies. The group to which the results of a study may be applied or generalized is called a population. A **population** is the total number of possible units or elements that could potentially be included in a study. Although people often think the word "population" refers to everyone in the world, researchers typically define their populations more narrowly. For example, the population for the first research hypothesis in Table 1.1 could be "consumers," "U.S. consumers," "Georgia consumers," or "University of Georgia consumers." Researchers typically try to define their populations as broadly as possibly (e.g., "consumers" rather than "University of Georgia consumers") to maximize the applications or implications of their research.

Learning Check 1

Reviewing What You've Learned So Far



1. Review questions
    a. What are the five steps involved in the research process?
    b. Why is it useful to review and evaluate theories and research before conducting a study?
    c. What are the two main characteristics of research hypotheses?
    d. What is the difference between an independent variable and a dependent variable?
2. The following are several research hypotheses from published studies. For each research hypothesis, identify the independent and dependent variable.
    a. "College students will rate instructors who dress formally (i.e., business suit and tie) as having more expertise than instructors who dress casually (i.e., slacks and shirt)" (Sebastian & Bristow, 2008, p. 197).
    b. "It was expected tha. . . . greater amounts of television viewin. . . . would predict greate. . . . posttraumatic stress symptoms" (McLeish & Del Ben, 2008, p. 417).
    c. "Do borrowers who did not complete college have greater financial distress compared with . . . non-borrowers who did not complete college [or] borrowers who earned a degree?" (Jabbari et al., 2023, p. 6).
    d. "We expected that drivers would believe that texting while driving is more detrimental to their driving performance than browsing social media" (Hashash et al., 2019, p. 70).

It's extremely difficult to collect data from all members of a population. Imagine, for example, the time and money needed to collect information from every consumer in the United States! For this reason, researchers typically draw conclusions about populations based on information collected from a sample of the population. A **sample** is a subset or portion of a

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

population. Table 1.4 describes the samples used in the four studies introduced in Table 1.1 . As you can see from Table 1.4 , samples vary greatly in terms of their targeted population and size (the number of participants). The extent to which the results of a study can be generalized to the target population depends on the extent to which the sample is representative of the population.

Determining How Variables Will Be Measured: Levels of Measurement

The research hypotheses described in Table 1.1 involve variables such as type of appeal, intentions to use cannabis, and gender identity. To conduct a research study, the researcher must decide how to measure the variables stated in the research hypothesis. **Measurement** is the assignment of numbers or categories to objects or events according to rules.

For example, to measure the variable “height,” a researcher might use “number of inches from the ground in bare feet” as a form of measurement. To measure “success in college,” a student’s grade point average (GPA) may be obtained from school transcripts. “Self-esteem” might be measured by having people complete a questionnaire and be categorized as having either “low” or “high” self-esteem. As these examples demonstrate, different types of variables have different forms of measurement. Consequently, researchers have identified four distinct levels for measuring variables: nominal, ordinal, interval, and ratio.

The values of variables measured at the **nominal level of measurement** differ in category or type. The word *nominal* implies having to do with “names,” such that we use first and last names to distinguish between people. Gender identity is an example of a nominal variable, in that it consists of categories or types (such as cis male and cis female) rather than numeric values. In the first research hypothesis in Table 1.1 , the independent variable of “type of appeal” is measured at the nominal level with two categories: FOMO and non-FOMO.

Variables measured at the **ordinal level of measurement** have values that can be placed in an order relative to the other values. Rankings (such as finishing first, second, or third in a race), size (small, medium, large, or extra-large), and ratings (below average, average, or above average) are examples of ordinal variables. The values of ordinal variables indicate that one value represents relatively more or less of the variable than other values, but they don’t specify the precise difference between values. For example, although you can say that a “medium” drink is smaller than a “large” drink which is smaller than an “extra-large” drink, you can’t specify the exact difference in size between the three drinks.

Table 1.4 Research Hypotheses and Study Samples

Research Hypothesis	Study Sample
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.	“We recruited 150 participants from Amazon MTurk” (Friederich et at., 2024, p. 105).
Exposure to pro-cannabis messages on social media is	“Teens (N = 350) . . . needed to be between 13 and 17 years old and from Washington state” (Willoughby et

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Study Sample
related to increased intentions to use cannabis.	al., 2024, p. 185).
Mothers have more accurate knowledge of their children's online activities than do fathers.	"triadic data including reports from children aged 13 to 18, mothers and fathers . . . recruited in Flanders, the northern, Dutch speaking region of Belgium . . . a total of 357 valid triads was achieved" (Symons et al., 2017, p. 405).
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	"undergraduate college students aged 18–30 who were enrolled at 4-year institutions. This yielded a sample of 46,510 students" (Marx et al., 2024, p. 334).

The values of variables measured at the **interval level of measurement** are numbers equally spaced along a continuum. One example of an interval variable is the Fahrenheit scale of temperature. Here, a difference of 5 degrees has the same meaning anywhere along the scale; for example, the difference between 45°F and 50°F is the same as the difference between 65°F and 70°F. Many variables studied in the behavioral sciences (e.g., personality characteristics or attitudes) are considered to be measured at the interval level of measurement.

Variables at the **ratio level of measurement** are identical to interval variables, with one exception: Ratio scales possess what is known as a true zero point, for which the value of zero (0) represents the complete absence of the variable. Variables that describe a physical dimension (such as height, weight, distance, and time duration) typically have a true zero point. In the last research hypothesis in Table 1.1 , the variable "number of times victimized in last 12 months" is a ratio variable because it has a true zero point, where zero would indicate the complete absence of victimizations.

One advantage of ratio measurement versus interval is that ratio variables allow for a greater number of comparisons among values. Consider, for example, the value 6 for the ratio variable "inches." Not only is the difference between 4 and 6 inches the same as the difference between 6 and 8 inches (thereby involving addition and subtraction, which also can be used with interval variables), 6 inches is twice as long as 3 inches and half as long as 12 inches (involving multiplication and division). Multiplication and division comparisons cannot be made with interval variables. You cannot, for example, say that a temperature of 90°F is three times as much temperature as 30°F.

The boldfaced and italicized text in Table 1.5 illustrates how the four studies described in the earlier tables measured their independent and dependent variables. As you can see, the variables in these studies have been measured in a variety of ways at different levels of measurement.

Table 1.5 Research Hypotheses and Measurement of Variables

Research Hypothesis	Independent Variable	Dependent Variable

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Independent Variable	Dependent Variable
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.	Type of appeal **FOMO, non-FOMO (nominal)**	Intention to invest in cryptos **Likelihood of investing** **1-7 scale (1 = Highly unlikely to 7 = Highly likely) (ordinal)**
Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.	Exposure to pro-cannabis messages on social media **How often saw pro-cannabis messages (ratio)**	Intentions to use cannabis **Likelihood of using cannabis in the future (ratio)**
Mothers have more accurate knowledge of their children's online activities than do fathers.	Type of parent **Mother, Father (nominal)**	Accuracy of online activities **Accurate, somewhat accurate, inaccurate (nominal)**
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	Gender identity **Nonbinary, cis man, cis woman** **(nominal)**	Number of times victimized in last 12 months **(ratio)**

What difference does the level of measurement make?

A variable's level of measurement has important implications for researchers in that it influences how research hypotheses are stated and how data are analyzed. For example, imagine you study the variable "success in college." You could measure this variable at the *nominal* level of measurement by classifying students into one of two groups: successful or unsuccessful. Researchers typically use nominal variables when they are interested in studying differences between groups, such as, "Do successful and unsuccessful students differ in their study habits?"

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

You could instead measure success in college as an *ordinal* variable by ranking students from top to bottom, which enables a researcher to study relative differences with less concern for the precise size of these differences. For example, using ranked data, you could ask, “How similar are younger and older faculty members’ rankings of their students?”

To study success in college at the *interval* level, you could rate each student on a scale from 1 (low) to 100 (high). You could then use these ratings to address a research question such as, “Is there a relationship between faculty ratings of their students and students’ ratings of themselves?”

Finally, to measure success in college using the *ratio* level of measurement, you could use students’ salary after graduation (in dollars) as an indicator of success. Salary is a ratio variable because it contains a true zero point. In practice, ratio variables can be used to ask many of the same questions as those involving interval variables. For example, you could ask, “Is there a relationship between faculty members’ ratings of students and students’ salaries after graduation?”

As you see, how researchers measure their variables influences how they state their research questions and hypotheses. The level of measurement has another important implication: It helps determine the statistics researchers use to analyze their data. Certain statistics can be applied to variables measured at the interval or ratio levels of measurement, whereas others are appropriate for nominal or ordinal variables. As we move from chapter to chapter of this book, the level of measurement of the variables being analyzed will be noted.

Selecting a Method to Collect the Data: Experimental and Nonexperimental Research Methods

In addition to drawing a sample from a population and determining how the variables will be measured, the third step in collecting data is to determine the type of research method to use to collect the data on these variables. Research methods may be classified into two main types:

experimental research methods, and

nonexperimental research methods.

Experimental research methods.

Experimental research methods are methods designed to test causal relationships between variables—more specifically, whether changes in independent variables produce or cause changes in dependent variables. To make inferences about cause-effect relationships, researchers conducting an experiment must first eliminate other possible causes or explanations for changes in the dependent variable besides the independent variable. If it can be claimed that a variable other than the independent variable created the changes in the dependent variable, the research results are said to be confounded. A **confounding variable** is a variable related to an independent variable that provides an alternative explanation for the relationship between the independent and dependent variables.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

To illustrate the concept of confounding variables, consider the following question: Is there a relationship between a mother's ethnicity and the birth weight of her baby? Although research has shown differences in the birth weights of babies of different ethnicities, one study noted that some of this research did not eliminate the possibility that these differences may be partly due to differences between ethnic groups on factors associated with low birth weights, such as the mother's average age at the time of pregnancy and behaviors such as smoking and drinking (Ma, 2008). These other factors are considered confounding variables in that they provide alternative explanations for the relationship between ethnicity and babies' birth weights. Confounding variables make it difficult for researchers to draw firm conclusions regarding the relationship between the independent and dependent variable.

Learning Check 2

Reviewing What You've Learned So Far



1. Review questions
    a. What is the difference between a population and a sample?
    b. Within the research process, what is the relationship between populations and samples?
    c. What are the four levels of measurement? How do they differ?
2. For each of the following variables, name the scale of measurement (nominal, ordinal, interval, or ratio).
    a. Type of school attended (public, private)
    b. Probability of graduating college in 4 years (0% to 100%)
    c. Rating of television (good, better, best)
    d. Number of computers in home

Researchers minimize the influence of confounding variables in two main ways. First, researchers exercise experimental control by making the research setting (i.e., characteristics of the research participants, the instruments or measures administered, the methods used to collect data) the same for all participants. In the birth weight study, for example, the research design might control for the effect of parental smoking on birth weight by only including nonsmokers in the study sample, thereby excluding those who smoke.

Another way researchers exert control over the research setting is by including a condition known as a **control group** , a group of participants in an experiment not exposed to the independent variable. For example, if you conducted an experiment examining the effects of caffeine on one's health, one group (the experimental group) might be instructed to drink coffee while a second group (the control group) drinks decaffeinated coffee. By comparing the results of the two groups, researchers can then assess the impact of caffeinated coffee on the health of those who drink it.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Because researchers cannot possibly identify and control for all potential confounding variables, a second strategy used to minimize the influence of confounding variables is **random assignment** , which is assigning participants to each category of an independent variable in such a way that each participant has an equal chance of being assigned to each category. For example, a researcher could randomly assign a participant to the experimental or control condition by flipping a coin: heads for experimental, tails for control. The purpose of random assignment is to equalize or neutralize the effects of confounding variables by distributing them equally over the levels of the independent variable.

Nonexperimental (correlational) research methods.

Experimental research designs are one of the best tools researchers have for making causal inferences. However, the ability to make causal inferences requires a great deal of control that may not always be possible or desirable. For this reason, researchers often employ **nonexperimental research methods** , which are research methods designed to measure relationships between variables without having the ability to infer cause-effect relationships. These research methods are sometimes referred to as *correlational* research methods. It is important to understand that correlational research methods are different from correlational statistics we'll introduce in Chapter 13. Correlational research methods pertain to how data are *collected* ; correlational statistics pertain to how data are *analyzed* . Some of the most common types of nonexperimental research designs include quasi-experiments, survey research, observational research, and archival research.

Quasi-experimental research compares naturally formed or preexisting groups rather than randomly assigning participants to conditions. One example of this type of research was a study examining the effects of the COVID-19 pandemic on three groups of students: those living on campus, those living off campus but within commuting distance to school, and those living far away in another city, state, or country (Bisconer & McGill, 2023). This study is a quasi-experiment because the students were not randomly assigned to one of the three living conditions before collecting data.

Survey research methods obtain information directly from a group of people regarding their opinions, beliefs, or behavior. The goal of survey research, which can include using questionnaires and interviews, is to obtain information from a sample to represent or estimate the views of a larger population. One example of survey research examined college students' mental health by having them fill out a survey in which they reported their spirituality (the desire for truth, meaning, and significance of one's life) and their self-acceptance (knowledge and acceptance of one's self and limitations) (Nonis et al., 2024). Survey research is nonexperimental in that the variables being measured are not manipulated by the researcher.

Observational research is the systematic and objective observation of naturally occurring behavior or events. Observational research is often used to study phenomena that the researcher cannot or should not deliberately manipulate. For example, researchers studying interpersonal conflict among children would never force children to fight each other. One study resolved this issue by videotaping groups of adolescents discussing topics such as competition and peer pressure and then counting the number of times the adolescents teased or mocked each other (Connolly et al., 2015).

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Rather than observing or measuring behavior directly, **archival research** is the use of archives (records or documents of the activities of individuals, groups, or organizations) to examine research questions or hypotheses. One example of an archival research study studied the question of whether men are perceived to be more creative than women (Proudfoot et al., 2015). In this study, the 100 most viewed talks on TED.com were identified, and the researchers compared viewers' posted ratings of the talks given by men versus women.

Labeling variables in nonexperimental research.

One issue that often confuses students when learning about nonexperimental research is the practice of referring to the variables as the "independent variable" and the "dependent variable." For example, in the study described previously that involved biological sex and TED.com ratings, biological sex may be called the "independent variable." However, because biological sex isn't manipulated by the researcher, it would be inappropriate to say that biological sex "causes" differences in TED.com ratings. To avoid confusion, researchers sometimes refer to the independent variable and the dependent variable in a nonexperimental study as the **predictor variable** and **criterion variable**, respectively.

Analyzing the Data

Once data collection has been completed, the next step in the research process involves data analysis. This part of the research process addresses the primary topic of this book: statistics. Because this is the first chapter of this book, we won't describe analyzing data in detail. Instead, we introduce the notion that there are two main purposes of analyzing data: (1) to organize, summarize, and describe data that has been collected and (2) to test and draw conclusions about ideas and hypotheses. These two purposes are met by calculating two main types of statistics: descriptive statistics and inferential statistics.

Calculating Descriptive Statistics

Descriptive statistics are statistics used to summarize and describe a set of data for a variable. For example, you've probably heard of statistics used to describe or summarize aspects of our society, such as crime statistics and unemployment statistics. Using a research-related example, Caitlin Abar, a researcher at Brown University, conducted a study examining increases in students' alcohol-related behaviors after entering college (Abar, 2012). To measure students' level of alcohol use, she created a variable called "typical weekend drinking," which "was measured as the sum of drinks consumed on a typical Friday and Saturday within the past 30 days" (p. 22). One way to summarize students' responses to the typical weekend drinking variable (0, 1, 2, etc.) would be to calculate the mean, which is the mathematical average of a set of scores. The mean, described in Chapter 3, is an example of a descriptive statistic. The first part of this book describes a variety of descriptive statistics used by researchers to summarize data they've collected.

Calculating Inferential Statistics

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Besides summarizing and describing data, a second purpose of statistics is to test hypotheses. **Inferential statistics** are statistical procedures used to test hypotheses and draw conclusions from data collected during research studies. Using inferential statistics, researchers make inferences about what's believed to exist in populations based on data gathered from samples of the population.

As an example of an inferential statistic, let's return to the example of students' alcohol-related behaviors (Abar, 2012). This study was interested in seeing whether these behaviors were related to aspects of these students' relationships with their parents. One aspect was "alcohol communications," defined as "how often they discussed alcohol related topics with their parents" (p. 22). To test a hypothesis regarding the relationship between students' alcohol use and their alcohol communications, we could use an inferential statistic known as the Pearson correlation coefficient, which will be discussed in Chapter 13. The chapters in the last half of this book discuss a broad variety of inferential statistics, along with detailed instructions for analyzing and drawing conclusions from statistical data and presenting research findings.

Drawing a Conclusion Regarding the Research Hypothesis

Once statistical analyses have been completed, the next step is to interpret the results of the analyses as they relate to the study's research hypothesis. More specifically, do the findings of the analyses support or not support the research hypothesis? The word *support* in the previous sentence is very important. Students and researchers are sometimes tempted to say their findings "prove" their hypotheses are either "true" or "false." However, as we'll discuss in Chapter 5, because researchers typically don't collect data from the entire population, they can't know with 100% certainty whether their hypothesis has been proven to be true or false. A *huge* part of this book is learning how to draw appropriate conclusions from statistical analyses.

Communicating the Findings of the Study

Conducting research requires a variety of different skills: conceptual skills to develop research hypotheses, methodological skills to collect data, and mathematical skills to analyze these data. Another integral part of the research process is the communication skills needed to inform others about a study. Researchers must be not only skilled scientists but also effective writers.

Learning Check 3

Reviewing What You've Learned So Far



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1. Review questions
    a. What are the main differences between experimental and nonexperimental research methods?
    b. What are confounding variables? What do researchers do to minimize the effects of confounding variables?
    c. What are the main types of nonexperimental research methods?
    d. What are the relative strengths and weaknesses of experimental and nonexperimental research?
    e. What are the main purposes of descriptive and inferential statistics?

For many years, the American Psychological Association (APA), the professional association for psychologists, has provided researchers guidance on how to communicate the results of their research. In 1929, the APA published a seven-page article entitled, “Instructions in Regard to Preparation of Manuscript.” In contrast, the seventh edition of the *Publication Manual of the American Psychological Association* , published in 2020, consists of 428 pages! This increase in length highlights the challenge of communicating research effectively and appropriately.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1.4 Plan of the Book

The primary purpose of this chapter was to introduce statistics and place it within the larger research process. The remainder of this book discusses different statistics used by researchers to examine various questions of interest. Although it's critical for you to learn to correctly calculate statistics, it is equally important to understand the role of statistics within the research process and appreciate issues related to the use (and sometimes misuse) of statistics.

The first half of the book (Chapters 2 through 6) introduces conceptual and mathematical issues that are the foundation of statistical analyses. Chapters 2, 3, and 4 discuss how data may be examined, described, summarized, and presented in numeric, visual, and graphic form. Descriptive statistics are introduced and discussed in these chapters. Chapters 5 and 6 introduce characteristics of the inferential statistics used to test research hypotheses. The chapters in the second half of the book cover a number of different inferential statistics used with different types of variables.

In most cases, the chapters in this book share a uniform format and structure and are centered on the research process, using examples of actual published research studies that address questions with which you may be familiar. For each example, we'll clearly state the research hypothesis, briefly describe how the data in the study were collected, illustrate the calculation and interpretation of descriptive and inferential statistics, discuss the extent to which the statistical analyses support the research hypothesis, and show how to communicate one's research to a broader audience.

1.5 Looking Ahead

We began this chapter by defining statistics and providing reasons why learning about them may be of value to you. Next, we placed statistics within the stages of the larger research process, a process that will be the foundation of this textbook. As we've explained, the research process is centered on a research hypothesis, which is a predicted relationship between variables. Once a hypothesis has been stated, the next step is to collect data to be analyzed. However, before conducting statistical analyses on a set of data, there is preliminary work that must be completed. The next chapter will focus on methods used by researchers to examine data.

1.6 Summary

Statistics may be defined as a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data. As such, statistics not only is concerned about how data are analyzed but also recognizes the importance of understanding how data are collected and how the results of analyses are interpreted and communicated. There are a variety of reasons why students should learn statistics: Statistics are encountered in a wide variety of daily activities, students may be asked or required to read and interpret the results of statistical analyses in their courses, students may use statistics in conducting their own research, and statistics may help one's career.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Researchers conduct research using the **scientific method** of inquiry, a method of investigation that uses the objective and systematic collection and analysis of empirical data to test theories and hypotheses. The research process used within the scientific method of inquiry consists of five main steps: developing a research hypothesis to be tested, collecting data, analyzing the data, drawing a conclusion regarding the research hypothesis, and communicating the findings to the study.

A **research hypothesis** is a statement regarding an expected or predicted relationship between variables. Developing a research hypothesis involves identifying a question or issue to be examined, reviewing and evaluating relevant theories and research, and stating the research hypothesis to be tested in the study. A **theory** is a set of propositions used to describe or explain a phenomenon. A research hypothesis contains **variables** , which are properties or characteristics of some object, event, or person that can take on different values. More specifically, a research hypothesis states the nature and direction of a proposed relationship between an **independent variable** (a variable manipulated or controlled by the researcher) and a **dependent variable** (a variable measured by the researcher that is expected to change or vary as a function of the independent variable).

Collecting data involves drawing a sample from a population, determining how the variables will be measured, and selecting a method to collect the data. A **population** is the total number of possible units or elements that could potentially be included in a study; a **sample** is a subset or portion of a population. Variables can be measured at one of four levels of measurement: **nominal** (values differing in category or type), **ordinal** (values placed in an order relative to the other values), **interval** (values equally spaced along a numeric continuum), or **ratio** (values equally spaced along a numeric continuum with a true zero point).

There are two main types of methods used to collect data: **experimental research methods** , designed to test whether changes in independent variables produce or cause changes in dependent variables, and **nonexperimental (correlational) research methods** , designed to examine the relationship between variables without the ability to infer cause-effect relationships.

In experimental research, researchers are concerned about possible **confounding variables** , which are variables related to independent variables that provide an alternative explanation for the relationship between independent and dependent variables. To minimize the impact of confounding variables, a research study may exercise experimental control over various aspects of the situation, including the use of a **control group** , which is a group of participants not exposed to the independent variable the research is designed to test, or **random assignment** , which involves assigning participants to each category of an independent variable in such a way that each participant has an equal chance of being assigned to each category.

Examples of nonexperimental research methods are **survey research** , in which information is directly obtained from a group of people regarding their opinions, beliefs, or behavior; **observational research** , which involves the systematic and objective observation of naturally occurring events; and **archival research** , which uses archives (records or documents of the activities of individuals, groups, or organizations) to examine research questions or hypotheses. In nonexperimental research, the independent variable and

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

dependent variable may be referred to as the **predictor variable** and **criterion variable**, respectively.

Once data have been collected, the next step is to analyze them using two main types of statistics: **descriptive statistics**, which summarize and describe a set of data, and **inferential statistics**, which are statistical techniques used to test hypotheses and draw conclusions from data collected during research studies.

Once the statistical analyses have been completed, the results of the analyses are interpreted regarding whether they support or not support the study's research hypothesis. The final step in the research process is to communicate the study to a broader audience.

1.7 Important Terms

archival research

confounding variable

control group

criterion variable

dependent variable

descriptive statistics

experimental research methods

independent variable

inferential statistics

interval level of measurement

level of measurement (nominal, ordinal, interval, ratio)

measurement

nonexperimental research methods

observational research

population

predictor variable

quasi-experimental research

random assignment

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

research hypothesis

sample

scientific method

statistics

survey research

theory

variable

1.8 Exercises

1. Identify the independent and dependent variable for each of the following research questions and hypotheses from published articles.
    a. “We hypothesize that finance majors will embody narcissistic traits more than other business majors” (Wood et al., 2021, p. 924).
    b. “Men are expected to employ physical characteristics (particularly those that are directly related to sex) more often than women in selecting dating candidates” (Hetsroni, 2000, p. 91).
    c. “It is hypothesized that perceptions of emotional pain and support needs will be higher for the community of the Black victim than those of the officers’” (Kahn et al., 2024).
    d. “Gen Z students are more likely to believe that ‘Group projects bring out the worst in people’ than Millennials” (Schlee et al., 2020, p. 145).
    e. “We hypothesized to find . . . higher stress and anxiety post COVID in the Portuguese group in contrast with the British ” (Penalba-Sanchez et al., 2022, p. 322).
    f. “We hypothesized that consumers would think that a ‘sale price’ presentation would generate a greater monetary savings than an ‘everyday low price’ presentation” (Tom & Ruiz, 1997, p. 403).
2. The following are a number of research questions and hypotheses from actual published articles. For each hypothesis, identify the independent and dependent variable.
    a. “The use of color in a Yellow Pages advertisement will increase the perception of quality of the products for a particular business when compared with noncolor advertisements” (Lohse & Rosen, 2001, p. 75).
    b. “We hypothesized that parents who use more frequent corporal and verbal punishmen. . . . will report more problem behaviors in their children” (Brenner & Fox, 1998, p. 252).
    c. “It was hypothesized that adolescents with anorexia nervosa woul. . . . be more respectful when compared with peers with bulimia nervosa” (Pryor & Wiederman, 1998, p. 292).
    d. “The purpose of the present research was to assess brand name recognition as a function of humor in advertisements. . . . It was predicted that participants would recognize product brand names that had been presented with humorous

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

advertisements more often than brand names presented with nonhumorous advertisements" (Berg & Lippman, 2001, p. 197).

e. "The purpose of this study was to explore the relation between time spent in daycare and the quality of exploratory behaviors in 9-month-old infant. . . . it was hypothesized tha. . . . infants who spent greater amounts of time in center-based care would demonstrate more advanced exploratory behaviors than infants who did not spend as much time in center-based care" (Schuetze et al., 1999, p. 269).

f. "It was hypothesized that students would score higher on test items in which the narrative contains topics and elements that resonate with their daily experiences versus test items comprised of material that is unfamiliar" (Erdodi, 2012, p. 172).

3. Identify the independent and dependent variable for each of the following research questions and hypotheses from published articles.
    a. "It was hypothesized that the number of selfies posted . . . would be positively correlated with dimensions of narcissism" (Barry et al., 2017, p. 51).
    b. "We hypothesized that preschoolers' self-reported preferences for aggressive media . . . would be positively related to parent reports of children's aggressive behaviors" (Jamnik & DiLalla, 2018, p. 159).
    c. "We hypothesized that helicopter parenting would lead to higher levels of depression" (Darlow et al., 2017, p. 3).
    d. "We predicted that weight-related shame would be associated with greater binge eating" (Craven & Fekete, 2019, p. 45).
    e. "We hypothesize that men with strong gender identities will be more sensitive to manipulated threats to their masculinity compared with men with gender identities that are less central to their identity" (Harrison & Michelson, 2019, p. 65).
    f. "After exposure to fake news story with user comments identifying the news as fake, people are less likely to share the fake news than after exposure to the fake news story with user comments supportive of the content" (Colliander, 2019, p. 204).
4. Name the scale of measurement (nominal, ordinal, interval, ratio) for each of the following variables:
    a. The amount of time needed to react to a sound
    b. Gender
    c. Score on the Scholastic Aptitude Test (SAT)
    d. Political orientation (not at all conservative, conservative, very conservative)
    e. Political affiliation (Democrat, Republican, Independent)
5. Name the scale of measurement (nominal, ordinal, interval, ratio) for each of the following variables:
    a. One's age (in years)
    b. Size of soft drink (small, medium, large, extra large)
    c. Voting behavior (in favor vs. against)
    d. IQ score
    e. Parent (mother vs. father)
6. A faculty member wishes to assess the relationship between students' scores on the SAT and their performance in college.
    a. What is a possible research hypothesis in this situation?
    b. What are the independent and dependent variables?
    c. How could you measure the variable "performance in college" at each of the four levels of measurement?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

7. A researcher hypothesizes that drivers who use cellular phones will get into a greater number of traffic accidents than drivers who do not use these phones. For this example,
    a. What are the independent and dependent variables?
    b. How could you measure the dependent variable?
    c. How could you conduct this study using an experimental research method? How could you conduct this study using a nonexperimental method?

Answers to Learning Checks

Learning Check 1

1.

    a. IV: Instructor's dress; DV: Expertise

    b. IV: Amount of television viewing; DV: Posttraumatic stress symptoms

    c. IV: Student group; DV: Financial distress

    d. IV: Activity while driving; DV: Driving performance

Learning Check 2

2.

    a. Nominal

    b. Ratio

    c. Ordinal

    d. Ratio

Answers to Odd-Numbered Exercises

1.

    a. IV: Type of business major; DV: Narcissistic traits

    b. IV: Biological sex; DV: Employment of physical characteristics

    c. IV: Community; DV: Perceptions of emotional pain and support needs

    d. IV: Generation; DV: Beliefs about group projects

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

e. IV: Country; DV: Stress and anxiety

f. IV: Price presentation; DV: Perception of savings

3.

a. IV: Number of selfies posted; DV: Narcissism

b. IV: Preference for aggressive media; DV: Parent reports of children's aggressive behaviors

c. IV: Helicopter parenting; DV: Depression

d. IV: Weight-related shame; DV: Binge eating

e. IV: Strength of gender identity; DV: Sensitivity to threat to masculinity

f. IV: Type of user comments to news story; DV: Sharing of news story

5.

a. Ratio

b. Ordinal

c. Nominal

d. Interval

e. Nominal

7.

a. IV: Cellular phone use; DV: Number of traffic accidents

b. Example: Determine the number of traffic accidents each person has experienced in the past year.

c. Example of experimental: Use a driving simulation with the "driver" talking on a phone and measure the number of potential driving mistakes or accidents.

d. Example of nonexperimental: Take a survey of the number of accidents people have been in and whether or not they were using a phone during the accident.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CHAPTER 2 EXAMINING DATA: TABLES AND FIGURES

Learning Objectives

2.1 Describe the main purpose and the research hypothesis of the “lottery” study.

2.2 Describe different reasons for examining data.

2.3 Describe the steps used to create a frequency distribution table.

2.4 Describe the steps used to create a grouped frequency distribution table.

2.5 Explain the differences between bar charts, pie charts, histograms, and frequency polygons.

2.6 Describe the three aspects and types of distributions of data.

Chapter 1 described a process by which research may be conducted, which begins by stating a research hypothesis. The next step involves collecting data that will ultimately be used in statistical analyses designed to test the research hypothesis. However, researchers don’t conduct analyses until they’ve first examined the data to make sure appropriate conclusions can be drawn from the analyses. This chapter describes *why* researchers examine data and *how* data may be examined using tables, charts, and graphs. As with many of the chapters in this book, this discussion will be guided by the findings from a published research study.

2.1 An Example From the Research: Winning the Lottery

A friend of yours flips a coin and asks you to guess whether it’s “heads” or “tails.” You guess “tails,” but the answer is “heads.” Your friend flips the coin a second time, and again you predict “tails.” But again the answer is “heads.” On the third coin flip, you guess “tails” once more—but, once again, you’re wrong. As your friend prepares to flip the coin a fourth time, you tell yourself there’s little chance the coin could be “heads” again. But the reality is that the fourth coin flip is equally likely to be “heads” or “tails” because the outcome of the first three coin flips has absolutely no impact on the fourth flip. In other words, the outcome of a coin flip is a random event that can’t be controlled or predetermined. But do you really believe this?

The impact of random events can take on greater significance than the outcome of a coin flip. For example, government-run lotteries provide the opportunity to win huge amounts of money through the random selection of numbers. In 2001, Dr. Karen Hardoon, a researcher at McGill University in Canada, studied the thought processes people use when purchasing lottery tickets (Hardoon et al., 2001). As previous research had found gamblers believe they can control random events such as the rolling of dice or the spinning of a roulette wheel, Dr. Hardoon and her colleagues were interested in

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

determining “whether individuals with gambling problems perceive the purchase of lottery tickets in a similar manner as non-problem gamblers” (p. 752). On the basis of earlier findings, Dr. Hardoon hypothesized that problem gamblers are less likely to believe lottery outcomes are random than are nonproblem gamblers.

To test their research hypothesis, Dr. Hardoon and her associates located a group of gamblers and asked whether they had ever experienced problems as a result of their gambling. Each participant was classified into one of two groups: problem gamblers and nonproblem gamblers.

The participants in the study, which we’ll refer to as the lottery study, were shown four lottery tickets and asked, “If you were to buy a ticket to play in the lottery, which one would you select?” Each ticket contained six numbers selected from among the numbers 1 through 49; the four tickets were given a particular label. The *sequence* ticket had numbers in consecutive order (30-31-32-33-34-35); numbers on the *pattern* ticket increased by 5s (5-10-15-20-25-30); the *nonequilibrated* numbers (35-37-40-43-44-49) were clustered at the upper end of the 49 numbers, and the *random* ticket (7-8-23-34-36-42) shared none of the characteristics of the other three tickets. We’ll call the variable in this study “lottery ticket,” a variable with four categories: sequence, pattern, nonequilibrated, and random.

The researchers recorded which of the four tickets each participant chose. Table 2.1 lists the data for the lottery ticket variable for the nonproblem gamblers (the data for the problem gamblers are provided in the exercises at the end of this chapter). Looking at this table, we see that the nonproblem gamblers differed in their lottery ticket choices; however, it’s difficult to describe these differences given the unorganized nature of the collected data. Consequently, before conducting statistical analyses on their data, researchers typically organize and examine them. In the following sections, we will discuss why and how data are organized and examined.

2.2 Why Examine Data?

There are a variety of reasons why researchers examine data they’ve collected before analyzing the data. These reasons include the following:

To gain an initial sense of the data

To identify outliers

To evaluate research methodology

To determine whether data meet statistical criteria and assumptions

Each of these reasons is discussed in the following sections.

Table 2.1 The Lottery Ticket Choices of 22 Nonproblem Gamblers

Participant	Lottery Ticket
1	**Random**
2	**Random**
3	**Nonequilibrated**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Participant	Lottery Ticket
4	**Pattern**
5	**Random**
6	**Sequence**
7	**Random**
8	**Random**
9	**Pattern**
10	**Pattern**
11	**Random**
12	**Pattern**
13	**Random**
14	**Nonequilibrated**
15	**Random**
16	**Pattern**
17	**Nonequilibrated**
18	**Random**
19	**Random**
20	**Pattern**
21	**Random**
22	**Random**

Gaining an Initial Sense of the Data

Because researchers spend a great amount of time and energy designing their research studies, they're eager to analyze the data they've collected. Students taking courses in statistics, on the other hand, face a different challenge: They're given sets of data with which they have no prior experience. Although both researchers and students may be motivated to analyze their data as quickly as possible, it's important to first examine the data in order to gain an initial sense of them.

As an example, imagine you conduct a study regarding the fuel efficiency of automobiles and collect data on the miles per gallon (MPG) of different types of cars. You might expect the scores in your data to range from 10 to 50 MPG, with the majority between 25 and 35 MPG. Examining data helps researchers gain an initial sense of whether the data are in line with their expectations. Examining data also provides a way of confirming the results of later statistical analyses. We've found that students' errors in their calculations could have been avoided had they first looked at the set of data. For example, examining data will help you avoid making statements such as, "The average MPG in my sample was 127 MPG."

Identifying Outliers

Researchers also examine their data to identify **outliers** , defined as rare, extreme scores that lie outside of the range of the majority of scores in a set of data. Imagine, for example, that a group of 15 students is asked to take a 20-item test of mathematical calculations; each student's score is the number of test items answered correctly. In scoring the tests, a researcher finds that 14 students

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

correctly answered between 8 and 17 questions. However, one student provided the wrong answer to every question for a score of 0. In this case, the score of 0 would be considered an outlier.

If not identified, outliers may distort conclusions researchers draw about their sample and their data, particularly when the sample is relatively small. In the previous example, the single score of 0 might lead an observer to conclude the class as a whole performed more poorly on the test than they actually did.

Evaluating Research Methodology

Examining data is also useful in assessing the effectiveness of the methods used to collect the data. For the 20-item math test, for example, it might be reasonable to expect the majority of scores to fall in the middle of the 0 to 20 range. But what if every student answered at least 17 of the 20 questions correctly? You might conclude the test was too easy and that more difficult questions are needed to accurately assess students' math skills. By comparing the collected data for a variable with the range of possible values, researchers are able to evaluate and modify their measurement tools.

Determining Whether Data Meet Statistical Criteria and Assumptions

Researchers also examine their data to see whether they meet certain statistical criteria and assumptions. Many of the statistics in this book are based on assumptions regarding the shape and nature of a set of data that has been collected. For example, these procedures assume that in most data sets, the majority of scores will fall in the middle of the range of possible values, with a small amount of scores at the highest and lowest possible values (you may have heard this referred to as "the bell curve"). In the example of the 20-item math test, it might be assumed that most scores will be between 7 and 13, with smaller number of scores less than 7 or greater than 13. When assumptions such as this are not met, it's possible to draw inappropriate conclusions from statistical analyses that are conducted.

2.3 Examining Data Using Tables

One part of examining data that have been collected for a variable involves answering a series of questions:

> What are the values of the variable?
>
> How many participants in the sample have each value of the variable?
>
> What percentage of the sample has each value of the variable?

As it may be difficult to answer questions such as these just by looking at the data in Table 2.1 , researchers typically organize their data by creating a table known as a **frequency distribution table** , a table that summarizes the number and percentage of participants for the different values of a variable. This section discusses the steps involved in creating frequency distribution tables using these questions.

What Are the Values of the Variable?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The first step in organizing the data for a variable is to identify all of the possible values for the variable. Table 2.2 begins the construction of a frequency distribution table by identifying the four values for the lottery ticket variable: sequence, pattern, nonequilibrated, and random.

How Many Participants in the Sample Have Each Value of the Variable?

Once the values of the variable have been identified, the next step in organizing a set of data is to determine the **frequency** of each value, which is the number of participants in a sample corresponding to a value of a variable. These frequencies help indicate the nature and shape of the distribution of scores for a variable in a sample.

Table 2.2 Identifying the Values of the Lottery Ticket Variable

Number	Lottery Ticket
1	Sequence
2	Pattern
3	Nonequilibrated
4	Random

Table 2.3 Using the Tick Mark Method to Determine the Frequencies of the Lottery Ticket Variable

Lottery Ticket	Tick Marks	*f*
Sequence	|	1
Pattern	||||— |	6
Nonequilibrated	|||	3
Random	||||— ||||— ||	12
Total	||||— ||||— ||||—||||—||	22

One way to determine the frequencies for a variable is to sort the data into the different values and then count the number of participants having each value using a “tick mark method,” where a tick mark is placed next to a value of the variable each time that value appears in the sample. As shown in Table 2.3 , the tick marks are then counted to determine the frequency for each value. For example, the number of tick marks and therefore the frequency (f) for the sequence ticket is 1 because it was picked by one participant (Participant 6 in Table 2.1). Similarly, the tick marks for each of the other three types of tickets can be recorded and counted until all of the data in the sample are accounted for.

What Percentage of the Sample Has Each Value of the Variable?

Although it’s important to determine the frequency for each value of a variable, these frequencies by themselves are not always particularly informative. For example, how you interpret “There are 35 Democrats in my sample” is different when the sample consists of 40 voters rather than 400 voters.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Because the interpretation of frequencies depends on the size of the sample, it's useful to calculate the *percentage* of the sample having each value of a variable. These percentages can be calculated using the following formula:

$$\% = \frac{f}{total\,number\,of\,scores} * 100$$

where *f* is the frequency for a value of a variable.

Looking at Table 2.3 , 1 of the 22 nonproblem gamblers chose the sequence ticket. Therefore, the percentage of the sample choosing this ticket is

$$\begin{aligned} \% &= \frac{f}{total\,number\,of\,scores} * 100 \\ &= \frac{1}{22} * 100 = .05 * 100 \\ &= 5\% \end{aligned}$$

After calculating the percentage for each of the four tickets, Table 2.4 provides the frequency distribution table for the lottery ticket variable.

Once a frequency distribution table has been constructed for a variable, researchers can begin to draw conclusions about their sample. For the lottery ticket variable, the random ticket had the highest frequency and percentage (*f* = 12, 54%) compared to the pattern (*f* = 6, 27%), nonequilibrated (*f* = 3, 14%), and sequence (*f* = 1, 5%) tickets. As a result, the researchers in the lottery study made the following observation:

"The results of the present study indicated that, for the entire sample . . . the greatest percentage of tickets selected by the entire sampl. . . . were random tickets." (Hardoon et al., 2001, p. 760)

Table 2.4 Frequency Distribution Table for the Lottery Ticket Variable

Lottery Ticket	***f***	**%**
Sequence	1	**5%**
Pattern	6	**27%**
Nonequilibrated	3	**14%**
Random	12	**54%**
Total	22	**100%**

Learning Check 1

Reviewing What You've Learned So Far

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



1. Review questions
    a. What are some reasons for examining data before conducting statistical analyses?
    b. Why are outliers of concern to researchers?
    c. What questions do you address about a set of data for a variable in constructing frequency distribution tables?
    d. Why is it useful to calculate percentages in addition to frequencies for each value of a variable?
2. For each of the following situations, create a frequency distribution table.
    a. A college counselor asks a group of 116 seniors what their plans are after graduating from college; 71 say they are going to “work,” 34 are planning on going to “graduate school,” and 11 say they are “not sure” what their plans are.
    b. A high school instructor teaching a course finds that of the students she teaches, 69 are Freshmen, 18 are Sophomores, 12 are Juniors, and 9 are Seniors.
    c. A pollster stops 12 people and asks if they are either “in favor” or “against” a local proposition in Table 2.5 :

Table 2.5 Example of People’s Position

Person	Position
1	**In favor**
2	**In favor**
3	**Against**
4	**In favor**
5	**Against**
6	**Against**
7	**In favor**
8	**Against**
9	**In favor**
10	**Against**
11	**In favor**
12	**In favor**

    d. The company that makes M&Ms candy (www.mms.com) conducted a survey asking people which new color they would like to have added: purple, aqua, or pink. Table 2.6 are the votes of a hypothetical sample:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Table 2.6
Hypothetical Sample of New M&M Color

Person	Color
1	**Pink**
2	**Purple**
3	**Pink**
4	**Aqua**
5	**Purple**
6	**Purple**
7	**Purple**
8	**Aqua**
9	**Purple**
10	**Pink**
11	**Aqua**
12	**Pink**
13	**Pink**
14	**Purple**
15	**Aqua**
16	**Pink**
17	**Purple**
18	**Purple**
19	**Aqua**
20	**Purple**
21	**Purple**
22	**Pink**
23	**Pink**
24	**Purple**

2.4 Grouped Frequency Distribution Tables

The frequency distribution table in Table 2.4 was sufficient for presenting the frequencies for the lottery ticket variable. However, other situations may involve variables with a large number of possible values. Consider, for example, the variable grade point average (GPA). The values for GPA typically range from 0.00 to 4.00, with hundreds of possible values in between. Creating a frequency distribution table for GPA could result in an extremely large table, with many values having frequencies of zero.

To examine a variable with a large number of values, it is useful to construct a **grouped frequency distribution table** , a table that groups the values of a variable measured at the interval or ratio level

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

of measurement into a small number of intervals and then provides the frequency and percentage for each interval. To illustrate how grouped frequency distribution tables are constructed, Table 2.7 (a) provides a sample of 42 hypothetical GPAs. In Table 2.7 (b), each of these 42 GPAs has been placed into one of eight intervals that comprise a grouped frequency distribution table.

To illustrate how to read and interpret a grouped frequency distribution table, the lowest interval in Table 2.7 (b) consists of students with GPAs less than (<) 2.00. The frequency in this interval (*f* = 1) consists of one student: Student 29 (GPA = 1.96). The next interval consists of students with GPAs ranging from 2.00 to 2.29 and contains three students (4, 22, and 38). This grouping continues until all of the GPAs in the sample have been accounted for.

To create a grouped frequency distribution table, the real limits of each interval should be identified. The **real limits** of an interval are the values of the variable that are halfway between the top of one interval and the bottom of the next interval. For example, because of how values of GPA are rounded, the lowest value of GPA that would lead someone to be placed in the interval 2.30–2.59 is 2.295, which is halfway between the top of the 2.00–2.29 interval and the bottom of the 2.30–2.59 interval. The **real lower limit** is the smallest value of a variable that would be grouped into a particular interval. In this case, 2.295 represents the real lower limit for the 2.30–2.59 interval. On the other hand, the **real upper limit** is the largest value of a variable that would be grouped into a particular interval. For the 2.30–2.59 interval, the real upper limit is 2.594.

Once the grouped frequency table for the GPA example has been created (Table 2.7(b)), what preliminary conclusions might be drawn about these students? First, because the 2.90–3.19 interval has the greatest frequency (*f* = 10), you might conclude that the typical student in this sample has a GPA close to 3.00. Second, moving in both directions away from the 2.90–3.19 interval, the frequencies of the different intervals grow progressively smaller, with fewer and fewer students having either very high or very low GPAs. Table 2.8 provides a list of guidelines for creating grouped frequency distribution tables.

Cumulative Percentages

In addition to determining the percentage of the sample that has each individual value or interval for a variable, it is sometimes useful to combine these percentages. For the GPA example, you may be interested in knowing, “What percentage of the sample had GPAs less than 2.60?” To answer this question, the percentages in the “%” column of the frequency distribution table must be combined. This type of percentage is referred to as a **cumulative percentage** , defined as the percentage of a sample at or below a particular value of a variable.

Table 2.7a Grade Point Average (GPA), 42 Students: Individual GPAs

Student	GPA
1	**3.05**
2	**2.83**
3	**3.26**
4	**2.19**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Student	GPA
5	**3.52**
6	**3.34**
7	**3.02**
8	**2.97**
9	**3.71**
10	**2.50**
11	**2.77**
12	**3.10**
13	**3.06**
14	**3.83**
15	**2.39**
16	**3.16**
17	**3.92**
18	**2.37**
19	**3.65**
20	**3.70**
21	**3.00**
22	**2.26**
23	**2.65**
24	**3.40**
25	**2.87**
26	**2.92**
27	**2.89**
28	**3.41**
29	**1.96**
30	**3.77**
31	**3.20**
32	**2.86**
33	**2.94**
34	**2.71**
35	**3.43**
36	**3.95**
37	**3.32**
38	**2.28**
39	**3.08**
40	**3.25**
41	**2.40**
42	**3.22**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Table 2.7b Grade Point Average (GPA), 42 Students: Grouped Frequency Distribution Table

GPA	*f*	%
3.80+	3	7%
3.50−3.79	5	12%
3.20−3.49	9	21%
2.90−3.19	10	24%
2.60−2.89	7	17%
2.30−2.59	4	10%
2.00−2.29	3	7%
<2.00	1	2%
Total	42	100%

Table 2.8 Guidelines for Creating Grouped Frequency Distribution Tables

Guideline	Guideline Description
1. *Variables are grouped into approximately 10 intervals.*	The exact number of intervals depends on the data in each sample. For example, data that fall within a small range of values may be accurately represented with a relatively small number of intervals, whereas data extending across a wide range may require a larger number of intervals.
2. *The number of intervals should accurately represent the data.*	The intervals should represent the nature of the data as accurately as possible. For example, there should not be so large a number of intervals that many intervals have frequencies of zero (e.g., 3.00−3.05) or so few intervals that a large majority of the sample falls into only one or two intervals (e.g., 3.00−3.99).
3. *Intervals should be of equal size.*	To make the intervals comparable, they should be of equal size or width. For example, you would not want one interval of 2.01−2.50 (a width of .50) and another of 2.51−2.70 (a width of .20). One exception to this guideline is the lowest or highest interval (e.g., less than 2.00 or greater than 3.70), which is typically wider than the other intervals because it contains a relatively small frequency.
4. *Intervals should not overlap.*	The intervals should be distinct from each other, such that each score may be placed in only one interval. For example, intervals of 2.30− *2.60* and *2.60*− 2.90 are a problem because the GPA 2.60 is in both intervals leading to confusion regarding within which interval to place someone with a GPA of 2.60.

Table 2.9 includes the cumulative percentages for the GPA example. As an example of a cumulative percentage, the cumulative percentage for the 2.30–2.59 interval (19%) has been calculated by combining the percentages for three intervals: < 2.00, 2.00–2.29, and 2.30–2.59: 2% + 7% + 10% = 19%. As a result, we could conclude that 19% of this sample had GPAs less than 2.60. The accumulating of percentages continues until all (100%) of the scores in the sample are combined. As

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

shown in Table 2.9 , cumulative percentages are typically located in the final column of a frequency distribution table and labeled “Cum %.”

Table 2.9 Frequency Distribution Table With Cumulative Percentages for the GPA Variable

GPA	*f*	%	Cum %
3.80+	3	7%	**100%**
3.50−3.79	5	12%	**93%**
3.20−3.49	9	21%	**81%**
2.90−3.19	10	24%	**60%**
2.60−2.89	7	17%	**36%**
2.30−2.59	4	10%	**19%**
2.00−2.29	3	7%	**9%**
< 2.00	1	2%	**2%**
Total	42	100%	

2.5 Examining Data Using Figures

In addition to organizing data into frequency distribution tables, researchers also examine data visually using figures such as charts or graphs. There are four types of figures often used to visually portray data for variables: bar charts, pie charts, histograms, and frequency polygons. Which type of figure a researcher may use depends on the variable’s level of measurement (nominal, ordinal, interval, ratio). Data for variables measured at the nominal or ordinal level of measurement are typically displayed using bar charts or pie charts. The values of interval and ratio variables are graphed using histograms and frequency polygons. Each of these types of visual illustration is discussed in the following sections.

Displaying Nominal and Ordinal Variables: Bar Charts and Pie Charts

The values of variables measured at the *nominal* level of measurement differ in category or type; one example of a nominal variable is the lottery ticket variable discussed earlier, which consisted of four categories: sequence, pattern, nonequilibrated, and random. The values of variables measured at the *ordinal* level of measurement can be placed in an order relative to the other values; age (young, middle age, elderly) is one example of an ordinal variable.

One way to display nominal and ordinal variables is to use a bar chart. A **bar chart** is a figure that uses bars to represent the frequency or percentage of a sample corresponding to each value of a variable, with the different bars not touching each other. Figure 2.1 shows a bar chart for the data from the lottery ticket variable. In this figure, each bar corresponds to one of the four types of lottery tickets; the height of each bar corresponds to the frequency (*f*) of nonproblem gamblers who selected that type of ticket.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

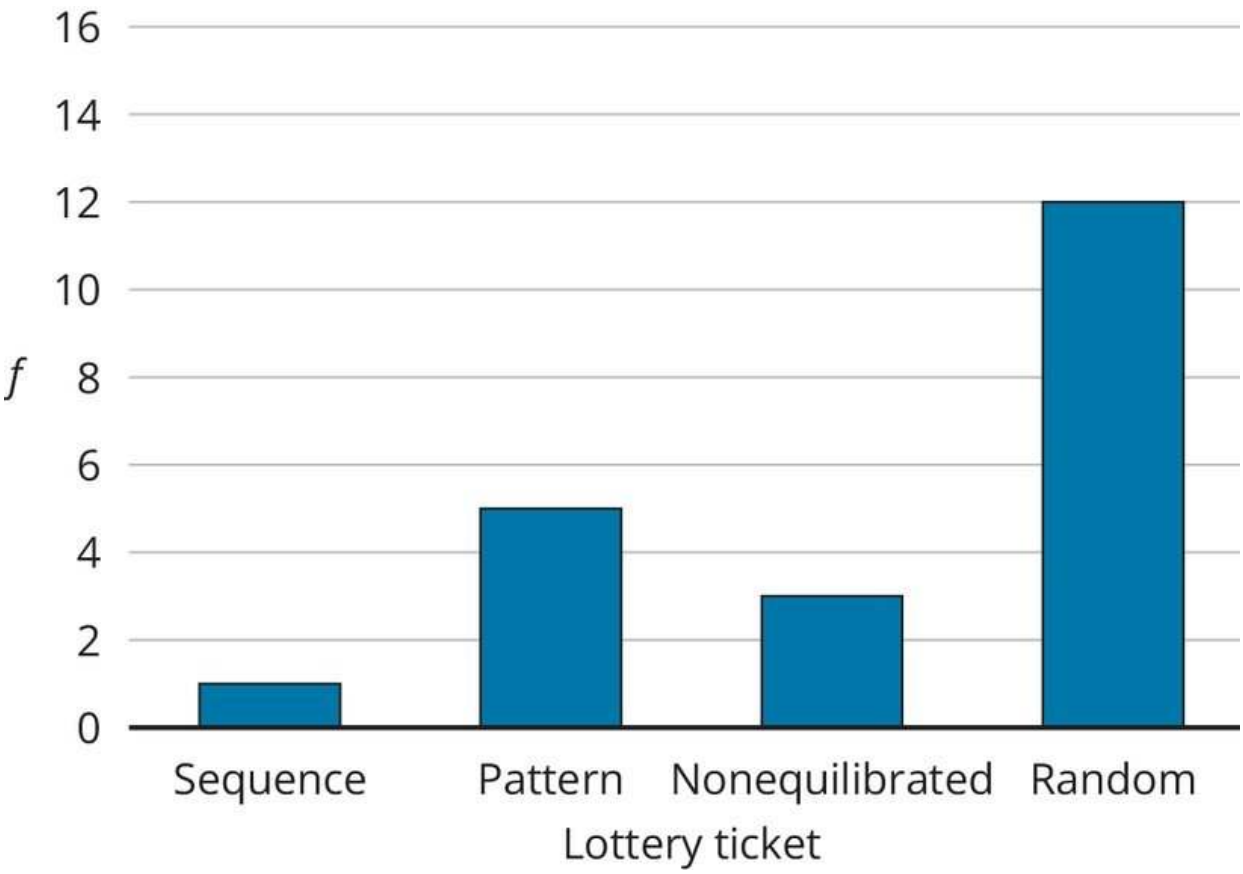


Figure 2.1 Bar Chart Showing the Frequencies of the Lottery Ticket Variable for the Nonproblem Gamblers

When creating a bar chart, the horizontal (X) axis contains the name of the variable and the values of the variable; the vertical (Y) axis contains the frequency (*f*) or percentage (%) for each value. Notice that the bars in the bar chart in Figure 2.1 do not touch each other; the gap between the bars indicates that the values of the variable represent distinct groups or categories that can't be placed along a numeric continuum.

A second type of figure used to display nominal or ordinal variables is a pie chart. A **pie chart** is a figure that uses a circle divided into proportions (slices) that represent the percentage of the sample for each value of a variable. Pie charts are created by dividing the 360 degrees of a circle into proportions corresponding to the percentage of each value.

Figure 2.2 shows a pie chart for the data for the lottery ticket variable. To illustrate how pie charts are constructed, the frequency distribution table in Table 2.4 shows that 5% of nonproblem gamblers chose the sequence ticket. Given that a pie chart consists of 360 degrees, 5% of the 360 degrees are needed to represent the sequence ticket. As 5% * 360 = 18, 18 degrees of the pie is assigned to the sequence ticket. Similarly, 97 degrees of the pie chart (27% of 360) are used to represent the pattern ticket. This process continues until all 360 degrees of the pie chart are accounted for.

The decision whether to use a bar chart or pie chart to illustrate a variable may depend on the number of values for the variable. For example, a pie chart for a variable that has 12 values may

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

result in many of the 12 “slices” being very thin and small, making them hard to see, meaning that a bar chart may be a better choice to illustrate this variable.

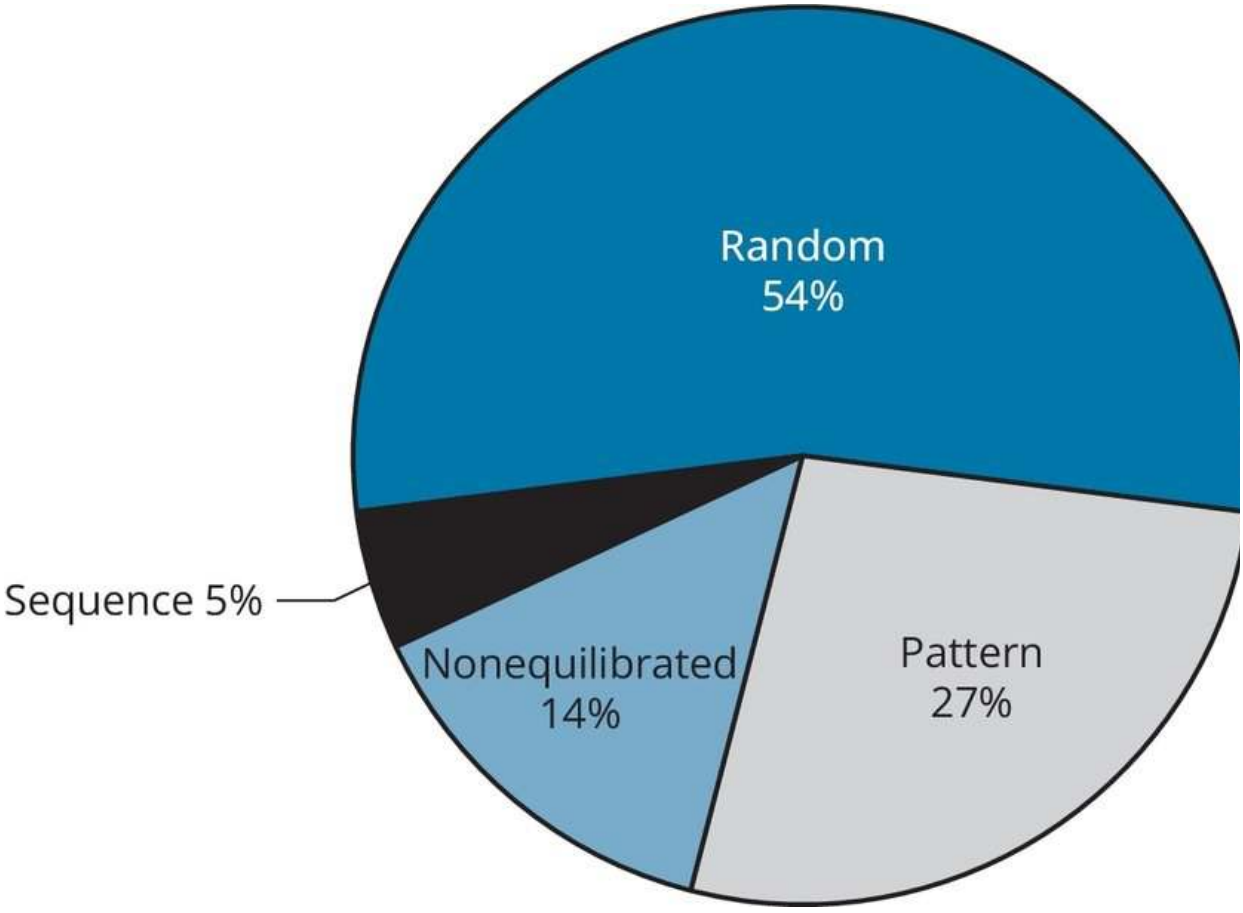


Figure 2.2 Pie Chart Showing the Percentages of the Lottery Ticket Variable for the Non−Problem Gamblers

Displaying Interval and Ratio Variables: Histograms and Frequency Polygons

Interval and ratio variables have values equally spaced along a numeric continuum and are identical to each other with one exception: Ratio scales possess a true zero point, for which the value of zero (0) represents the complete absence of the variable. The values of interval and ratio variables may be illustrated using histograms and frequency polygons, the choice of which depends on the number of values for the variable.

A **histogram** is a figure in which bars are used to represent the frequency of each value of a variable, with the different bars touching each other. The name and values of the variable are located on the X-axis, and the frequency or percentage for each value is on the Y-axis. Histograms are similar to bar charts, with one important difference: The bars in a histogram touch each other, indicating that the values of the variable are connected to each other along a numeric continuum.

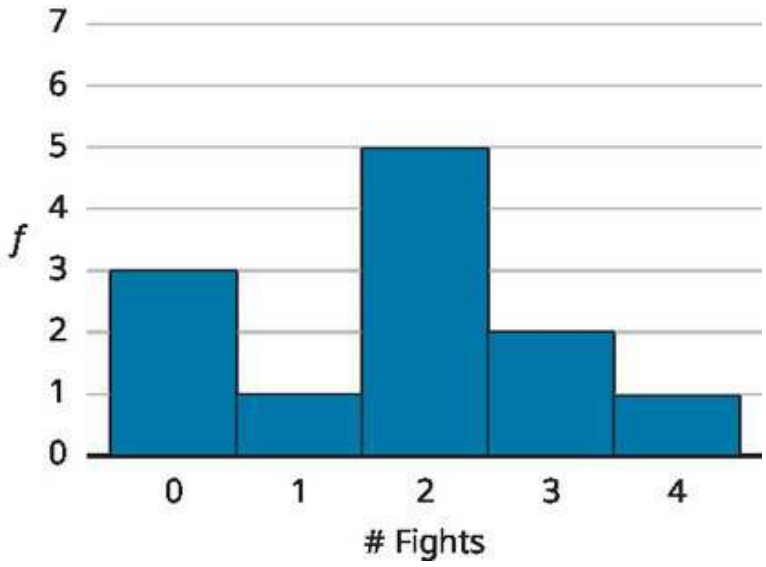
join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

As an example, imagine a teacher is interested in studying aggressive behavior in children and measures this variable by counting the number of fights each of her students is involved in over a 3-day period. “Number of fights” is a ratio variable rather than an interval variable because the value of 0 represents the complete absence of fights. Figure 2.3 provides a frequency distribution table and histogram for this variable.

A histogram is used when an interval or ratio variable consists of a relatively small number of values. But what if, for example, the number of fights could be anywhere from 0 to 20 rather than 0 to 5? In order to have the histogram fit on one page, the 21 bars would need to be very thin and crowded together, making the histogram confusing and difficult to read. In situations such as this, the values of the variable may be more clearly illustrated using a figure known as a **frequency polygon**, which is a line graph that uses data points to represent the frequency of each value of a variable, with straight lines connecting the data points. In a frequency polygon, the X-axis represents the name and values of the variable, while the Y-axis represents the frequency for each value.

# Fights	*f*
4	1
3	2
2	5
1	1
0	3
Total	12



▶ Long Description: Figure 2.3 Frequency Distribution Table and Histogram for the Number of Fights Variable

Figure 2.3 Frequency Distribution Table and Histogram for the Number of Fights Variable

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

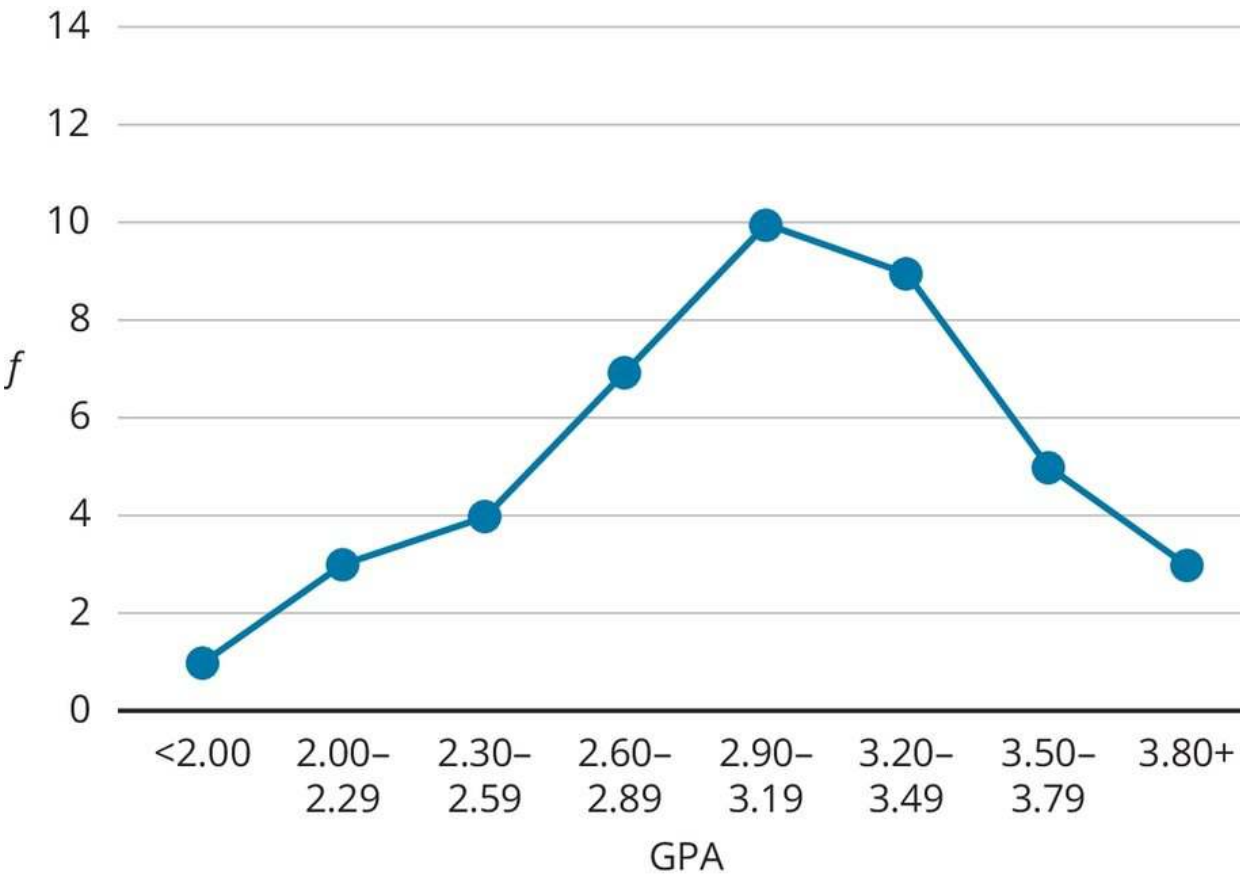


Figure 2.4 Frequency Polygon Showing the Frequency of Students in Each Interval of the Grade Point Average (GPA) Variable

Figure 2.4 shows a frequency polygon based on the grouped frequency distribution table of the earlier example of students' GPA (Table 2.7(b)). In this frequency polygon, the data points representing the frequencies of the eight GPA intervals are connected using straight, sharp lines. These lines indicate that the values of the variable can be connected together along a numeric continuum.

Although a frequency polygon like the one in Figure 2.4 may look jagged as it moves from one data point to the next, it provides a general sense of the shape of the distribution of scores for a variable. For example, looking at Figure 2.4 , you might conclude that the majority of GPAs in this sample are in the middle of the 1.00 to 4.00 range, with relatively low frequencies of students with either low or high GPAs.

Learning Check 2

Reviewing What You've Learned So Far

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



1. Review questions
    a. What is the difference between a frequency distribution table and a grouped frequency distribution table?
    b. For what levels of measurement would you create a pie chart, bar chart, histogram, or a frequency polygon?
    c. What is the difference between a bar chart and a histogram?
    d. Why would it be inappropriate to create a histogram or frequency polygon for a nominal or ordinal variable?
    e. Under what circumstances might you use a frequency polygon rather than a histogram to graph an interval or ratio variable?
2. For each of the following variables, identify whether you could use a bar chart, pie chart, histogram, and/or a frequency polygon to visually display the data.
    a. Miles per gallon (MPG)
    b. Living situation (on-campus, off-campus)
    c. Number of brothers and sisters
    d. Marital status (single, married, divorced)

Drawing Inappropriate Conclusions From Figures

A well-constructed figure can greatly facilitate the understanding of complex information. However, a poorly designed figure may not only be confusing, it can lead to inappropriate or misleading conclusions. Box 2.1 provides guidelines for constructing figures based on recommendations from the American Psychological Association (APA).

A simple example may demonstrate how altering the features of a figure can lead to different interpretations of the same set of data. Imagine a pollster assesses how voters with different political affiliations plan to vote on a ballot measure. She asks a sample of Democrats, Republicans, and Independents the question, “Do you plan on voting in favor of Measure A?” In examining her data, the pollster finds 56% of Democrats, 47% of Republicans, and 52% of Independents intend to vote “yes” on the measure.

Figure 2.5 graphs these three percentages using two bar charts; bar charts are used to display the data because the variable of interest (political affiliation) is a nominal variable (Democrat, Republican, or Independent). The bars don’t touch each other because the three political affiliations are distinct from each other and can’t be placed along a numeric continuum.

Although the two bar charts are similar, there is one important difference in how they are constructed. Following the guidelines in Box 2.1, the bar chart in Figure 2.5 (a) starts the Y-axis with the value 0% and ends with 100%. However, the bar chart in Figure 2.5 (b) starts with the value 45% and ends with 65%. By limiting the range of values of the percentages, the bar chart in Figure 2.5 (b) exaggerates differences in the heights of the three bars. As a result, someone looking at Figure 2.5 (b) may interpret the differences between the three political affiliations as being larger than someone looking at Figure 2.5 (a).

The saying “a picture is worth a thousand words” is particularly relevant for the presentation of statistical data. Because figures can provide a vivid representation of data, it’s the responsibility of researchers to portray their data in an accurate manner and to avoid creating images that may mislead or confuse.

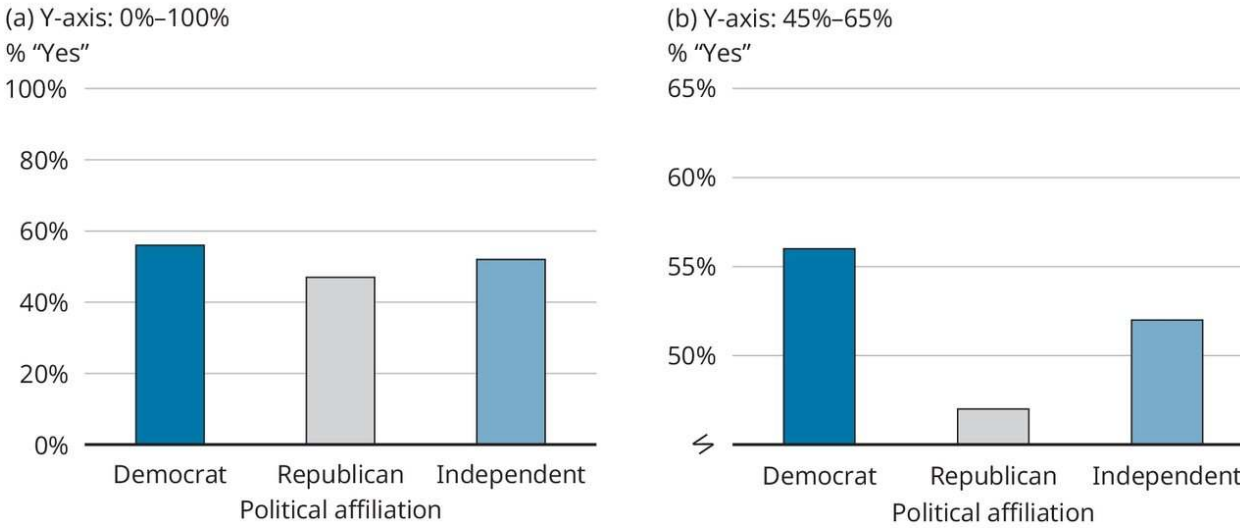
join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Box 2.1 Some APA Guidelines for Creating Figures

1. The figure should be formatted in a portrait rather than landscape page orientation (the reader should not turn the page sideways to read the table).
2. The figure is numbered using Arabic numbers (Figure 1, not Figure *I* or Figure *A*).
3. The figure includes a title that describes the contents of the figure. For example, it is not enough to name a figure simply " Figure 2.1 " or " Figure 2.1 . Bar chart."
4. Figures have a rectangular shape, with the height of the vertical (Y) axis two thirds to three fourths the length of the horizontal (X) axis.
5. The horizontal (X) axis contains the values of the variable; the vertical (Y) axis contains the frequency (*f*) or percentage (%) of the sample having each value of the variable.
6. The *X* - and *Y* -axes include descriptive labels written in both upper- and lowercase letters (as opposed to all capital letters).
7. The labels for the *X* - and *Y* -axes are short enough to fit on one line along the axes.
8. The label of the Y-axis is often placed parallel to the Y-axis (the letters face sideways rather than vertically) but can sometimes be seen on top of the y-axis for accessibility.
9. The Y-axis starts with the value zero (0) and is divided into equally spaced values or intervals.
10. The Y-axis is divided into 6 to 10 values (just enough values to accommodate bar length). The bars should neither be flattened along the bottom of the Y-axis nor bumped up against the top of the Y-axis.

Political affiliation	% "Yes"
Democrat	56%
Republican	47%
Independent	52%



▶ Long Description: Figure 2.5 Different Displays of the Same Data

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Figure 2.5 Different Displays of the Same Data

2.6 Examining Data: Describing Distributions

We've discussed how researchers construct tables and figures to organize their data. Looking at a table or figure enables researchers to describe the distribution of values or scores for a variable. Describing a distribution involves addressing questions such as, "Which values of the variable occur most often?," "How do the frequencies of the values of the variable change in relation to the most common values?," and "What is the amount of differences among the different values?"

Questions such as these pertain to three aspects of distributions: modality, symmetry, and variability. These three aspects of distributions are introduced in the following sections; they'll also be discussed in greater detail in the next two chapters of this book.

Modality

The first feature of a distribution of data for a variable is its modality. The **modality** of a distribution refers to the value or values of the variable that have the highest frequency or occur most often in a set of data. Using the lottery study as an example, identifying the modality of the lottery ticket variable involves answering the question, "Which of the four lottery tickets was chosen the most often?"

The modality of a distribution can be identified by looking at a frequency distribution table; the modality is the value(s) of the variable with the highest frequency (*f*) or percentage (%). The modality may also be determined by examining a figure and locating the tallest bar in a bar chart or histogram, the highest point in a frequency polygon, or the biggest "slice" in a pie chart. For the lottery ticket variable, the modality is the random ticket.

In identifying the modality of a distribution, it may be that one value or a limited number of values adjacent to one another clearly have the highest frequency. In such cases, the distribution is referred to as a unimodal distribution. A **unimodal distribution** is a distribution where one value of a variable occurs with the greatest frequency. Looking at the frequency distribution table (Table 2.4) or figure (Figure 2.1) for the lottery ticket variable, this distribution would be considered unimodal as one value, the random ticket, clearly occurred more often than the other three ticket choices.

In contrast to a unimodal distribution, it's possible that two distinct values may both have relatively high frequencies; if so, the distribution is referred to as a **bimodal distribution** , which is a distribution where two values clearly distinct from each other have the greatest frequency. A bimodal distribution sometimes resembles the two humps on a camel's back. As an example of a bimodal distribution, in April 2024 the Pew Research Center (www.pewresearch.org) asked a sample of 7,000 voters whether they felt Donald Trump had the mental fitness needed to be president, with five responses ranging from "extremely" to "not at all." The majority of voters chose one of the two extreme choices, either "extremely" (23%) or "not at all" (35%), with only 14% choosing the middle response of "somewhat."

And, in addition to unimodal and bimodal distributions, it's possible for a distribution to have more than two values with relatively equally high frequencies; this type of distribution is known as a **multimodal distribution** —a distribution where more than two values have the greatest frequency. In examining people's attitudes toward gun control, in May 2023, the news network CNN asked 1,227 people the question, "Do you think allowing gun owners to carry their guns in public places makes those places safer, less safe, or makes no difference in how safe a public place is?" They found people differed greatly in their responses, with 32% saying "makes public places safer," 36% saying

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

"makes public places less safe," and 32% saying "makes no difference," Figure 2.6 presents an example of a unimodal, bimodal, and multimodal distribution.

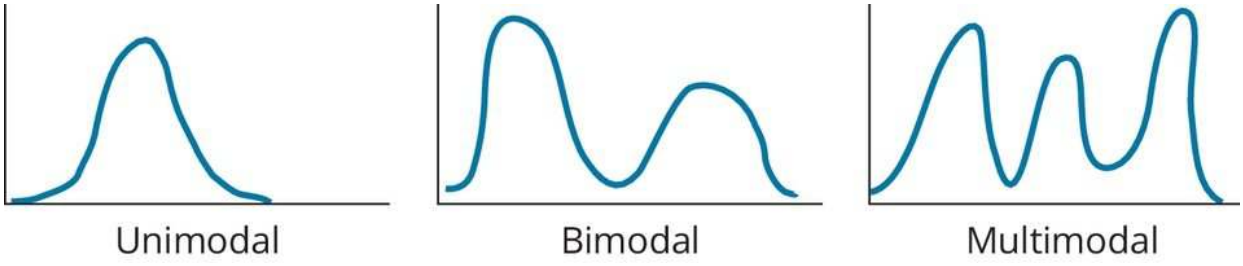


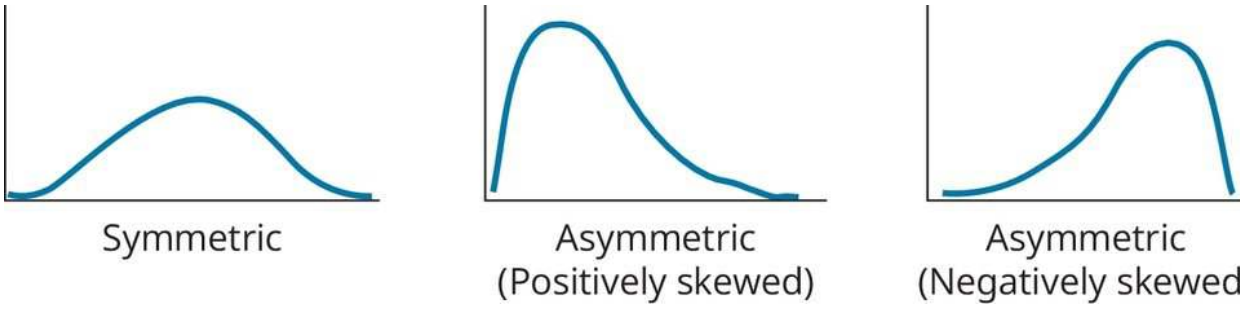
Figure 2.6 Distributions With Different Modalities (Unimodal, Bimodal, and Multimodal)

Symmetry

The second feature of a distribution of data is its symmetry. The **symmetry** of a distribution refers to how the frequencies of values of a variable change in relation to the most common or frequently occurring values. Describing the symmetry of a distribution means answering the question, "What is the shape of the distribution?"

The symmetry of a distribution can be determined by looking at the frequencies in a frequency distribution table or the bars or points in a figure. By doing so, a distribution may be described as being either symmetric or asymmetric. A **symmetric distribution** is one in which the frequencies change in a similar manner moving away in both directions from the most frequently occurring values. An example of a symmetric distribution is illustrated in Figure 2.7 . One way to determine whether a distribution is symmetric is to draw the distribution on a piece of paper and then fold the distribution vertically down the middle so that the two halves are on top of each other. A distribution is symmetric when the two halves are similar in shape such that they resemble mirror images of each other.

However, a distribution may also be asymmetric. An **asymmetric distribution** (also known as a *skewed* distribution) is one in which the frequencies change in a different manner moving away in both directions from the most frequently occurring values. In asymmetric distributions, the highest frequencies are located at one end of the distribution rather than in the middle, such that if you fold an asymmetric distribution down the middle, the shapes of the two halves do not look the same.



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Figure 2.7 Distributions With Different Symmetries (Symmetric, Positively Skewed, and Negatively Skewed)

In an asymmetric distribution, the frequencies take on the shape of a long tail as the values move toward the end of the distribution; scores in the tail of the distribution are often considered to be outliers. For example, in the 2022 Current Population Survey, the U.S. Census Bureau (www.census.gov) estimated that most Americans (63%) had annual incomes less than $50,000. However, a very small percentage (3%) were greater than $200,000; these higher incomes would be considered outliers.

Depending on the location of its tail, asymmetric distributions are referred to either as positively skewed or negatively skewed. A **positively skewed distribution** is one in which the higher frequencies are at the lower end of the distribution, with the tail on the upper (right) end of the distribution; a **negatively skewed distribution** is one in which the higher frequencies are at the upper end of the distribution, with the tail on the lower (left) end of the distribution. Figure 2.7 provides an illustration of a symmetric distribution and two asymmetric distributions (one positively skewed and one negatively skewed). Annual income, illustrated by the U.S. Census Bureau example mentioned earlier, is an example of a positively skewed distribution.

Variability

The third way in which distributions may be described is in terms of the amount or nature of differences among the different values for a variable. The **variability** of a distribution refers to the amount of differences in a distribution of data for a variable. The variability in a distribution is described by answering the question, “To what extent are the scores in the distribution different from each other?”

The amount of variability in a distribution can be estimated by looking at a frequency distribution table or figure. Imagine, for example, an instructor gives an 11-item quiz to her class and finds almost all of the students answer six, seven, or eight of the items correctly. In this situation, the distribution would resemble the shape of a tall mountain: thin and tall, with many of the scores residing in a few values of the variable. A distribution with a small amount of variability such as this one is referred to as a **peaked distribution** , defined as a distribution where much of the data are in a small number of values of a variable.

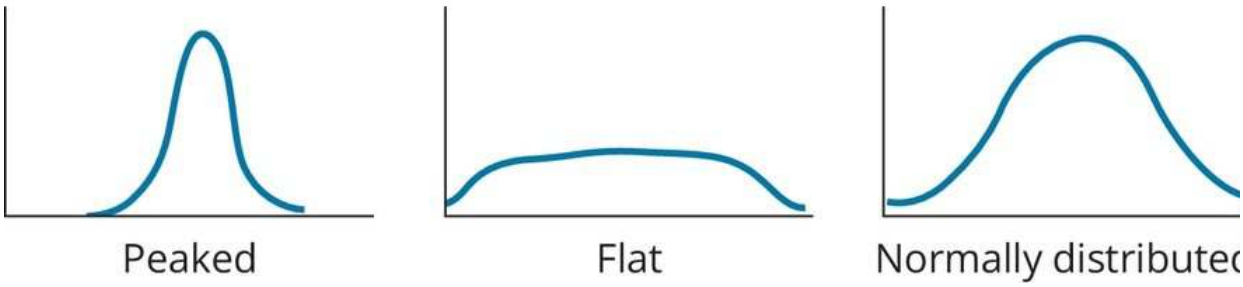


Figure 2.8 Distributions With Different Variability (Peaked, Flat, and Normally Distributed)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

On the other hand, what would the distribution look like if she had found her students' scores were distributed relatively evenly across the entire range of possible scores? In this situation, the distribution would not be tall and thin like a mountain but rather would be low and wide like a plateau. This type of distribution is referred to as a **flat distribution** , which is a distribution where the data are spread evenly across the values of a variable. The CNN gun control example described earlier is an example of a flat distribution.

In discussing modality, symmetry, and variability, distributions may have more than one mode, may be asymmetric (skewed), and may be peaked or flat. But what if a distribution has none of these qualities? A distribution that is unimodal, symmetric, and neither peaked nor flat may be described as "bell shaped." If you look at a bell, you'll notice that it is the highest in the middle, and as you move away from the top of a bell in both the left and right directions, the shape of the bell slopes downward in the same manner. A bell-shaped distribution for a variable is given a special name: a normally distributed variable.

A **normally distributed variable** is a distribution for a variable that is considered unimodal, symmetric, and neither peaked nor flat. Note that the word *normal* is a statistical concept and not a value judgment. In other words, asymmetric, peaked, or flat distributions are not considered "abnormal." The concept of normal distributions will be discussed in greater detail in Chapter 5. Figure 2.8 provides an illustration of three distributions with different amounts of variability: peaked, flat, and normally distributed.

Learning Check 3

Reviewing What You've Learned So Far



1. Review questions
    a. What are the three aspects used to describe distributions of variables?
    b. What is the difference between unimodal, bimodal, and multimodal distributions?
    c. What is the difference between symmetric and asymmetric (skewed) distributions?
    d. What is the difference between positively and negatively skewed distributions?
    e. What is the difference between peaked and flat distributions?
    f. What are the characteristics of a normally distributed variable?
2. For each of the following sets of data, determine whether the distribution of scores is unimodal, bimodal, or multimodal.
    a. 4, 2, 11, 6, 10, 2, 6, 13, 2, 10, 6, 10, 1
    b. 15, 17, 13, 18, 17, 14, 20, 18, 17, 14, 17
    c. 20, 10, 50, 10, 80, 30, 70, 10, 60, 10, 80, 0, 30, 80, 10, 100, 80
3. For each of the following sets of data, determine whether the distribution of scores is symmetric or asymmetric (skewed).
    a. 4, 5, 1, 4, 2, 4, 3, 5, 3, 4, 4, 3
    b. 220, 230, 250, 230, 200, 240, 210, 230, 270, 220, 240, 230, 240, 230
    c. 3, 1, 9, 2, 3, 6, 2, 7, 3, 5, 2, 6, 3, 2, 2
4. For each of the following sets of data, determine whether the distribution of scores is peaked or flat.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

a. 16, 10, 19, 12, 13, 15, 14, 17, 15, 20, 11, 13, 19, 16, 15
b. 12, 15, 12, 11, 13, 11, 12, 12, 14, 10, 12, 11, 16, 12, 13, 12, 11
c. 14, 12, 16, 14, 13, 15, 12, 14, 14, 15, 14, 14, 13, 14

2.7 Looking Ahead

This chapter has discussed why and how researchers examine data. Examining data by creating tables and figures requires a number of different skills, including organizational skills (to prepare a set of data for statistical analysis), communication skills (to present data in tables or figures), and conceptual skills (to describe a distribution). The next two chapters focus on mathematical skills related to the statistical analysis of data, describing how researchers describe distributions numerically rather than visually. Given that this is a statistics textbook, we will focus on calculating statistics to describe and analyze data. However, and we cannot say this strongly enough, it is *extremely* important to examine a set of data before conducting statistical analyses as this will prevent errors both in your calculations of statistics as well as your interpretation of the results of these calculations.

2.8 Summary

Before conducting statistical analyses, researchers typically examine their data to gain an initial sense of the data, detect data coding or data entry errors, identify ***outliers*** (rare, extreme scores that lie outside of the range of the majority of scores in a set of data), evaluate research methodology, and determine whether the data meet statistical criteria and assumptions.

One common way of examining variables involves creating tables known as ***frequency distribution tables*** (tables that summarize the number and percentage of participants that express the different values for a variable) and ***grouped frequency distribution tables*** (tables that group the values of a variable measured at the interval or ratio level of measurement into a small number of intervals and then provide the frequency and percentage within each interval). To create a grouped frequency table, the ***real limits*** of each interval (the values of the variable that fall halfway between the top of one interval and the bottom of the next interval) need to be identified.

Data for a variable may also be examined by creating figures such as bar charts, pie charts, histograms, and frequency polygons. ***Bar charts*** and ***pie charts*** are used for variables measured at the nominal or ordinal level of measurement; a bar chart uses bars to represent the frequency or percentage of a sample corresponding to each value of a variable, with the different bars not touching each other; a pie chart is a circle divided into proportions that represent the percentage of the sample corresponding to each value of a variable. ***Histograms*** and ***frequency polygons*** are used for variables measured at the interval or ratio level of measurement; a histogram is a figure in which bars are used to represent the frequency of each value of a variable, with the different bars touching each other; a frequency polygon is a line graph that uses data points to represent the frequency of each value of a variable, with lines connecting the data points.

From examining their data, researchers can describe the distribution of values or scores for a variable. There are three main aspects of distributions: modality, symmetry, and variability.

The ***modality*** of a distribution is the value(s) of the variable that have the highest frequencies or occur most often in a set of data. ***Unimodal distributions*** are distributions where one value occurs with the greatest frequency, ***bimodal distributions*** have two values with the greatest frequency, and ***multimodal distributions*** have more than two values with the greatest frequencies.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The ***symmetry*** of a distribution refers to how the frequencies of values of a variable change in relation to the most common or frequently occurring values of the variable. In ***symmetric distributions*** , the frequencies change in a similar manner, moving away in both directions from the most common values of the variable. In ***asymmetric (skewed) distributions*** , the frequencies change in a different manner, moving away in both directions from the most common values of the variable.

The ***variability*** of a distribution refers to the amount of differences in a distribution of data. In ***peaked distributions*** , much of the data are in a few values of a variable; in ***flat distributions*** , the data are spread evenly across the values of a variable. A ***normally distributed variable*** has a symmetric distribution that is neither peaked nor flat.

2.9 Important Terms

asymmetric (skewed) distribution

bar chart

bimodal distribution

cumulative percentage

flat distribution

frequency (*f*)

frequency distribution table

frequency polygon

grouped frequency distribution table

histogram

modality

multimodal distribution

negatively skewed distribution

normally distributed variable

outlier

peaked distribution

pie chart

positively skewed distribution

real limits

real lower limit

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

real upper limit

symmetric distribution

symmetry

unimodal distribution

variability

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

2.10 Formula Introduced in This Chapter

Percentage (%)

$$\% = \frac{f}{total\ number\ of\ scores} * 100$$

2.11 Using IBM SPSS Software *

Creating Data Files: Defining Variables and Entering Data

* IBM SPSS software. SPSS is a registered trademark of International Business Machines Corporation.

1. Open a new data file (See Figure 2.9).

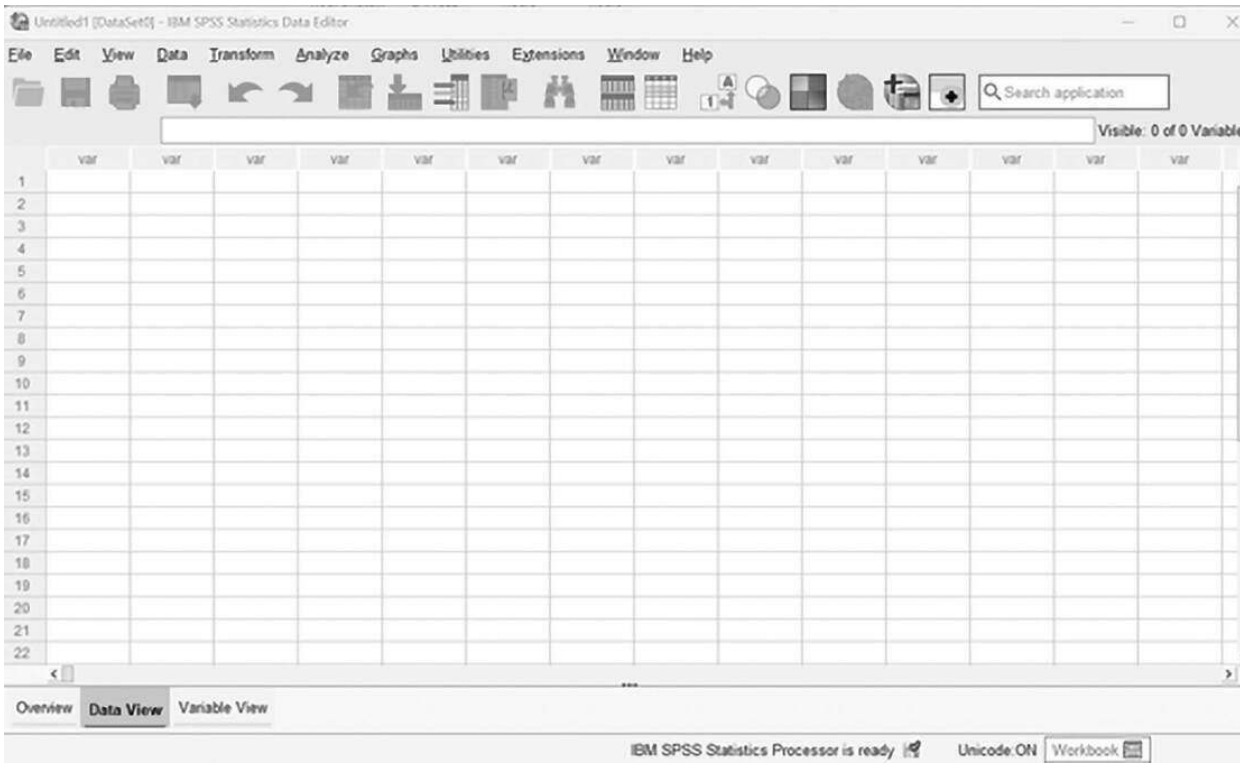


Figure 2.9 Data View

2. Define and name variable(s) (See Figure 2.10).

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

How? (1) Click **Variable View** tab, (2) click below **Name** , and (3) type name of variable in the cell.

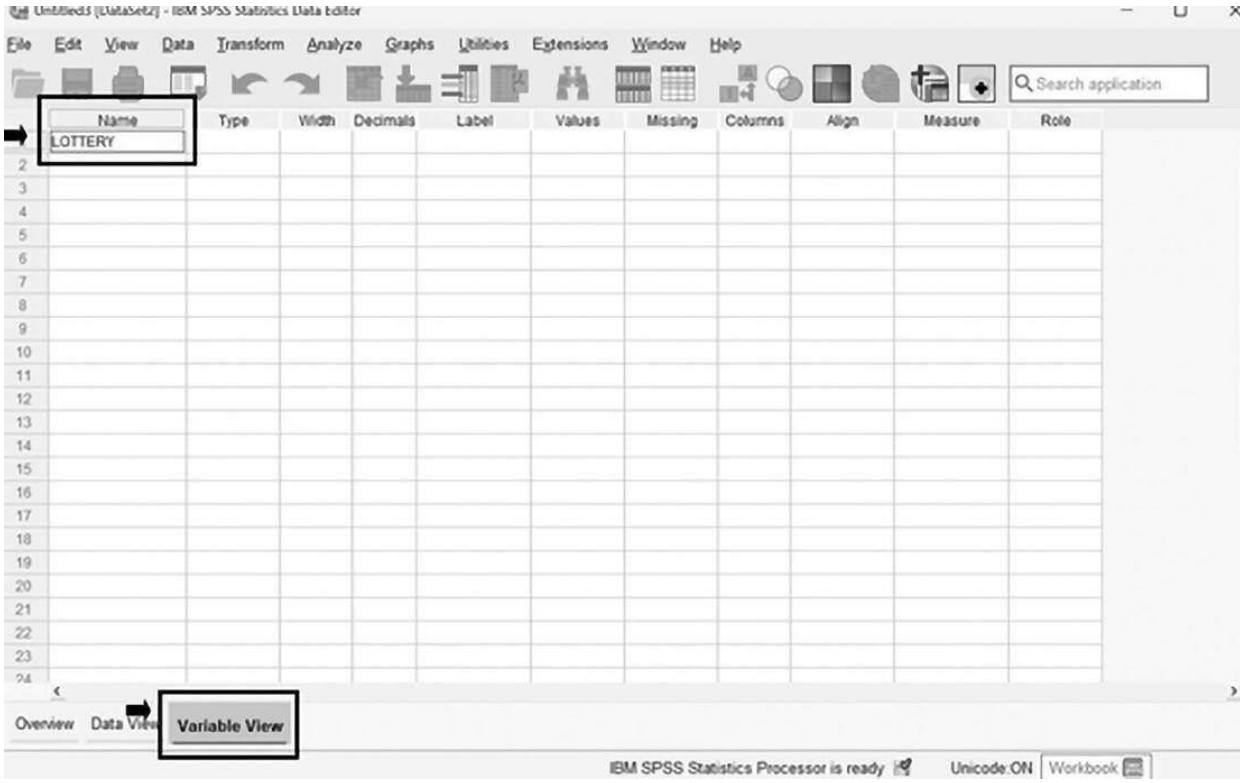


Figure 2.10 Variable View

3. Determine number of decimals for your variable(s) (See Figure 2.11).

How? Click desired number of decimal places in the **Decimals** box.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

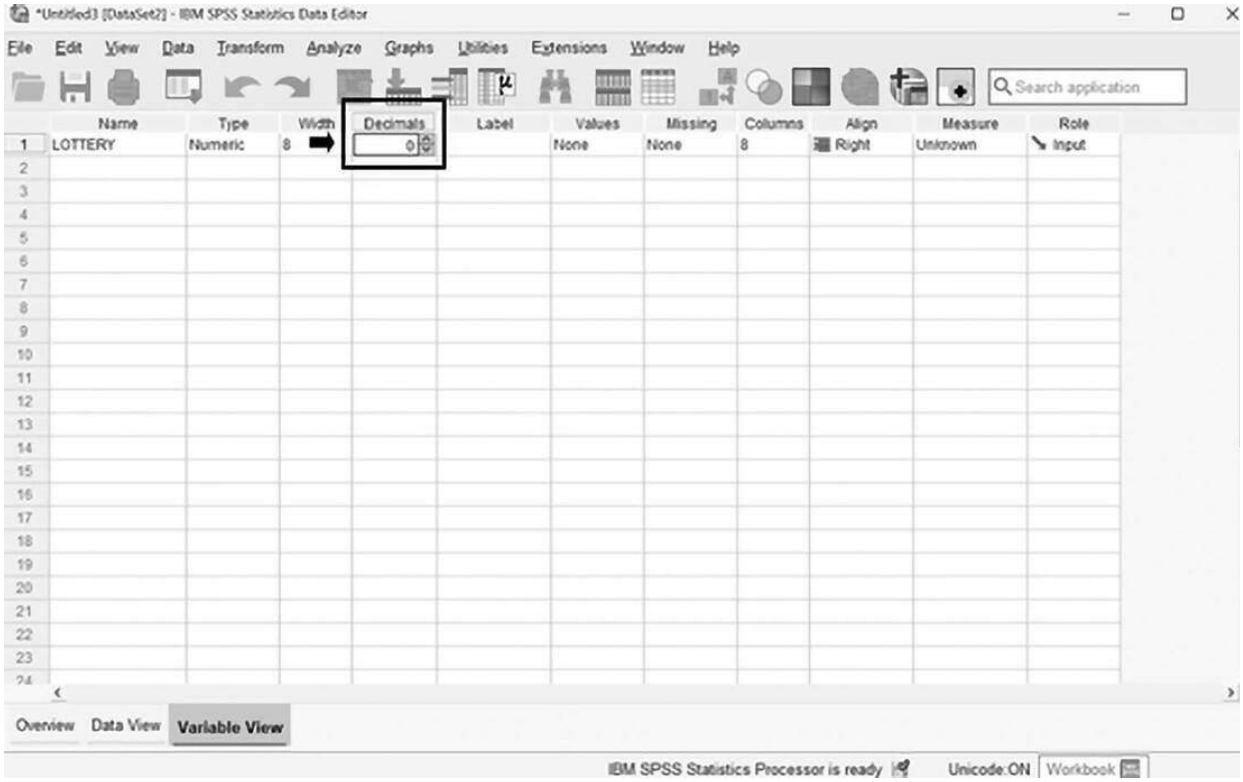


Figure 2.11 Decimals Box

4. Provide labels for your variable(s) (See Figure 2.12).

How? (1) Click below **Label** , and (2) type the label for the variable in the cell.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

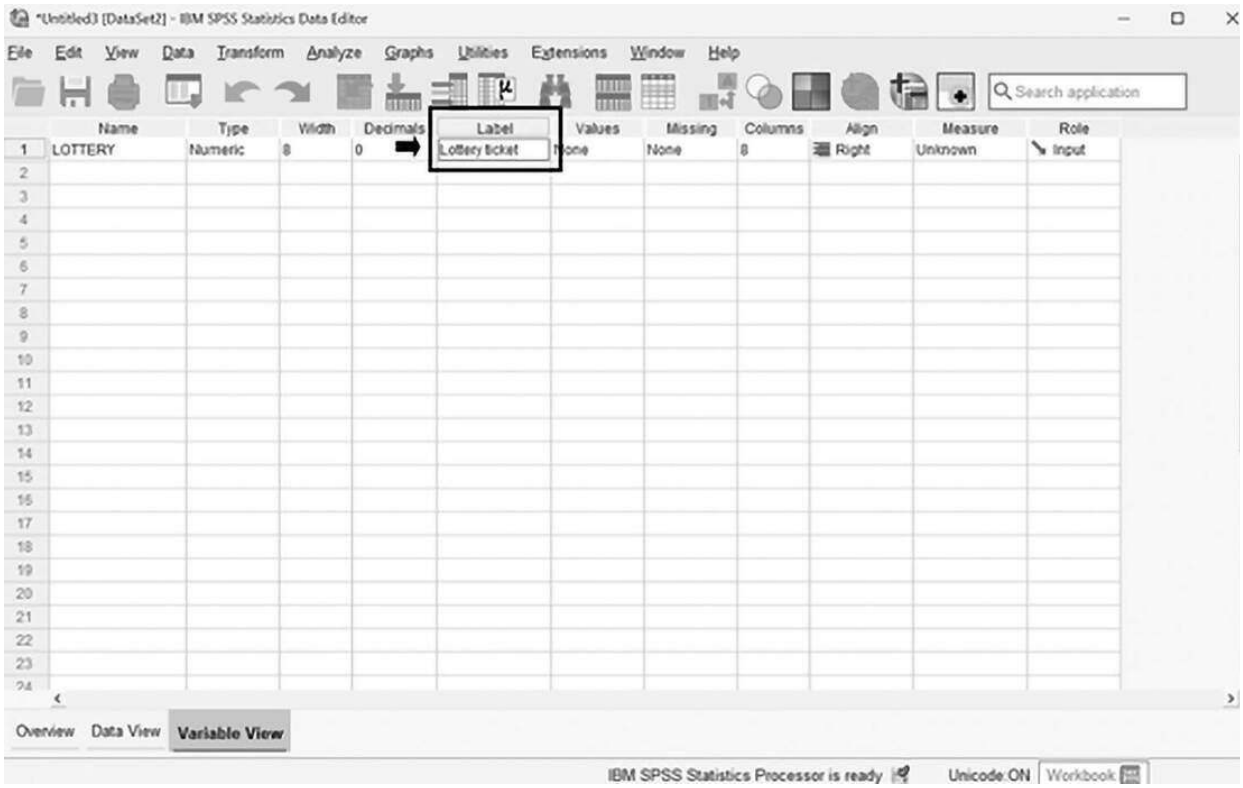


Figure 2.12 Label Box

5. Provide labels for values of nominal (categorical) variable(s) (See Figure 2.13).

How? (1) Click below **Values** , and (2) type labels for coded values of the variable.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

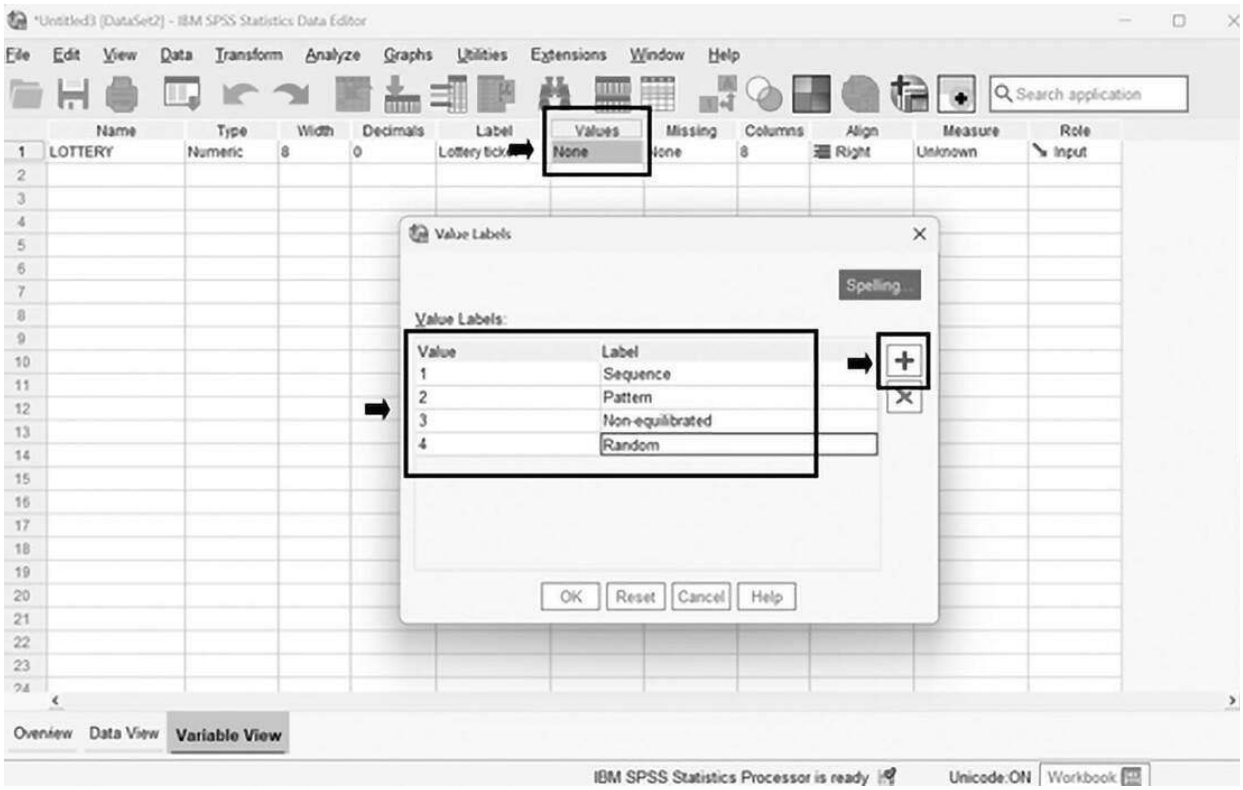


Figure 2.13 Nominal Variables

6. Enter data (See Figure 2.14).

How? (1) Click **Data View** tab and (2) enter data for the variable in the cells.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

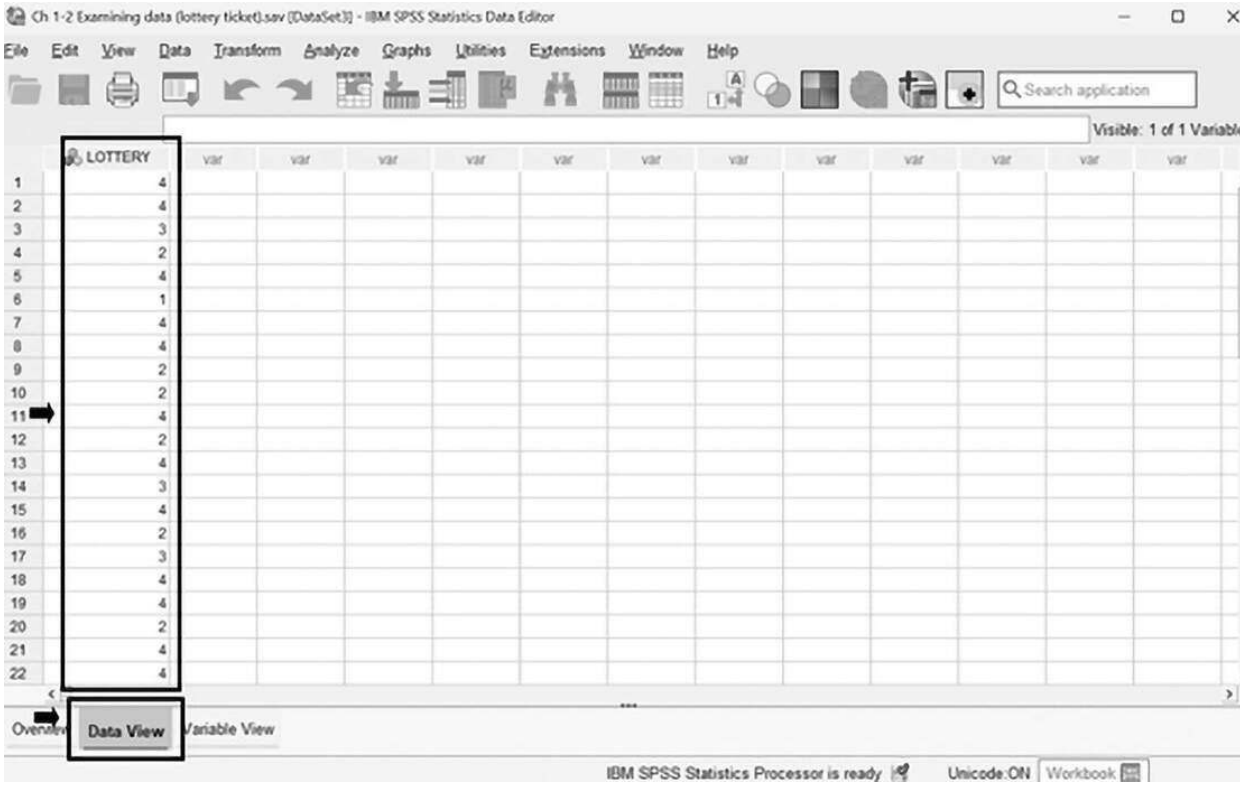


Figure 2.14 Enter Data View

Examining Data: Frequency Distribution Tables and Figures: The Lottery Study (2.1)

1. Select the frequency distribution procedure within SPSS (See Figure 2.15).

   How? (1) Click **Analyze** menu, (2) click **Descriptive Statistics** , and (3) click **Frequencies** .

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

To Jennifer , my kids Meagan and Eika , and my parents, Katsumi and Grayce Tokunaga.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

As a global academic publisher, Sage is driven by the belief that research and education are critical in shaping society. Our mission is building bridges to knowledge—supporting the development of ideas into scholarship that is certified, taught, and applied in the real world.

Sage's founder, Sara Miller McCune, transferred control of the company to an independent trust, which guarantees our independence indefinitely. This enables us to support an equitable academic future over the long term by building lasting relationships, championing diverse perspectives, and co-creating social and behavioral science resources that transform teaching and learning.

▶ Long Description:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

FUNDAMENTAL STATISTICS FOR THE SOCIAL AND BEHAVIORAL SCIENCES

Third Edition

Howard T. Tokunaga

San Jose State University



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

eISBN: 978-1-0718-6160-8

Copyright © 2027 by Sage.

All rights reserved. Except as permitted by U.S. copyright law, no part of this work may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without permission in writing from the publisher.

All third party trademarks referenced or depicted herein are included solely for the purpose of illustration and are the property of their respective owners. Reference to these trademarks in no way indicates any relationship with, or endorsement by, the trademark owner.



FOR INFORMATION:

2455 Teller Road, Thousand Oaks, California 91320, E-mail: orders@sagepub.com

3rd floor, HYLO, 103-105 Bunhill Row, London, EC1Y 8LZ, United Kingdom

10th Floor, Emaar Capital Tower 2, MG Road Sikanderpur, Sector 26, Gurugram, Haryana – 122002, India

18 Cross Street #10-10/11/12, China Square Central, Singapore 048423

Library of Congress Control Number: 2026943277

ISBN: 978-1-0718-6144-8

Printed in the United States of America

This book is printed on acid-free paper.

26 27 28 29 30 10 9 8 7 6 5 4 3 2 1

Acquisitions Editor: Anna Villarruel

Content Development Editor: Kenzie Offley

Editorial Assistant: Daria Terry

Production Editor: Syeda Aina Rahat Ali

Copy Editor: Michelle Ponce

Typesetter: Lumina Datamatics

Cover Designer: Candice Harman

Marketing Manager: Victoria Velasquez

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CONTENTS

Preface
Acknowledgments
About the Author
Chapter 1 Introduction to Statistics
Chapter 2 Examining Data: Tables and Figures
Chapter 3 Measures of Central Tendency
Chapter 4 Measures of Variability
Chapter 5 Normal Distributions
Chapter 6 Probability and Introduction to Hypothesis Testing
Chapter 7 Testing One Sample Mean
Chapter 8 Estimating the Mean of a Population
Chapter 9 Testing the Difference Between Two Means
Chapter 10 Errors in Hypothesis Testing, Statistical Power, and Effect Size
Chapter 11 One-Way Analysis of Variance (ANOVA)
Chapter 12 Two-Way Analysis of Variance (ANOVA)
Chapter 13 Correlation
Chapter 14 Linear Regression and Multiple Correlation
Chapter 15 Chi-Square
Appendix A: Tables
Appendix B: Review of Basic Mathematics
Glossary
References
Index

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

DETAILED CONTENTS

Preface

Acknowledgments

About the Author

Chapter 1 Introduction to Statistics

- 1.1 What Is Statistics?
- 1.2 Why Learn Statistics?
- 1.3 Introduction to the Stages of the Research Process
  - Developing a Research Hypothesis to Be Tested
    - Identifying a Question or Issue to Be Examined
    - Reviewing and Evaluating Relevant Theories and Research
    - Stating a Research Hypothesis: Independent and Dependent Variables
  - Collecting Data
    - Drawing a Sample From a Population
    - Determining How Variables Will Be Measured: Levels of Measurement
    - Selecting a Method to Collect the Data: Experimental and Nonexperimental Research Methods
  - Analyzing the Data
    - Calculating Descriptive Statistics
    - Calculating Inferential Statistics
  - Drawing a Conclusion Regarding the Research Hypothesis
  - Communicating the Findings of the Study
- 1.4 Plan of the Book
- 1.5 Looking Ahead
- 1.6 Summary
- 1.7 Important Terms
- 1.8 Exercises

Answers to Learning Checks

Learning Check 1

Learning Check 2

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Odd-Numbered Exercises

Chapter 2 Examining Data: Tables and Figures

2.1 An Example From the Research: Winning the Lottery

2.2 Why Examine Data?

Gaining an Initial Sense of the Data

Identifying Outliers

Evaluating Research Methodology

Determining Whether Data Meet Statistical Criteria and Assumptions

2.3 Examining Data Using Tables

What Are the Values of the Variable?

How Many Participants in the Sample Have Each Value of the Variable?

What Percentage of the Sample Has Each Value of the Variable?

2.4 Grouped Frequency Distribution Tables

Cumulative Percentages

2.5 Examining Data Using Figures

Displaying Nominal and Ordinal Variables: Bar Charts and Pie Charts

Displaying Interval and Ratio Variables: Histograms and Frequency Polygons

Drawing Inappropriate Conclusions From Figures

2.6 Examining Data: Describing Distributions

Modality

Symmetry

Variability

2.7 Looking Ahead

2.8 Summary

2.9 Important Terms

2.10 Formula Introduced in This Chapter

Percentage (%)

2.11 Using IBM SPSS Software *

Creating Data Files: Defining Variables and Entering Data

Examining Data: Frequency Distribution Tables and Figures: The Lottery Study (2.1)

2.12 Exercises

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 3 Measures of Central Tendency
3.1 An Example From the Research: The 10% Myth
3.2 Understanding Central Tendency
3.3 The Mode
Determining the Mode
3.4 The Median
Calculating the Median: Odd Number of Scores
Calculating the Median: Even Number of Scores
Calculating the Median When Multiple Scores Have the Median Value
3.5 The Mean
Calculating the Mean
Reporting Calculated Numbers
Rounding Calculated Numbers
Calculating the Mean From Frequency Distribution Tables
The Mean as a Balancing Point in a Distribution
The Sample Mean vs. the Population Mean: Statistics vs. Parameters
3.6 Comparison of the Mode, Median, and Mean
Strengths and Weaknesses of the Mode
Strengths and Weaknesses of the Median
Strengths and Weaknesses of the Mean
Choosing a Particular Measure of Central Tendency
3.7 Measures of Central Tendency: Drawing Conclusions
3.8 Looking Ahead
3.9 Summary
3.10 Important Terms
3.11 Formulas Introduced in This Chapter
Median (Odd Number of Scores)
Median (Even Number of Scores)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Mean
Mean (Frequency Distribution Table)
Mean (Balancing Point in a Distribution)
3.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 4 Measures of Variability
4.1 An Example From the Research: How Many Sometimes" in an Always"?
4.2 Understanding Variability
4.3 The Range
Strengths and Weaknesses of the Range
4.4 The Interquartile Range
Strengths and Weaknesses of the Interquartile Range
4.5 The Variance (s^2)
Definitional Formula for the Variance
Computational Formula for the Variance
Why Not Use the Absolute Value of the Deviation in Calculating the Variance?
Why Divide by N − 1 Rather Than N in Calculating the Variance?
4.6 The Standard Deviation (s)
Definitional Formula for the Standard Deviation
Computational Formula for the Standard Deviation
Eyeball Estimating" the Mean and Standard Deviation
4.7 Measures of Variability for Populations
The Population Variance (σ^2)
The Population Standard Deviation (σ)
Measures of Variability for Samples vs. Populations
4.8 Measures of Variability for Nominal Variables
The Index of Qualitative Variation
4.9 Measures of Variability: Drawing Conclusions

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

4.10 Looking Ahead
4.11 Summary
4.12 Important Terms
4.13 Formulas Introduced in This Chapter
Range
Interquartile Range
Variance (s^2) (Definitional Formula)
Variance (s^2) (Computational Formula)
Standard Deviation (s) (Definitional Formula)
Standard Deviation (s) (Computational Formula)
Population Variance (σ^2) (Definitional Formula)
Population Standard Deviation (σ) (Definitional Formula)
Index of Qualitative Variation (IQV)
4.14 Using SPSS
Calculating Measures of Central Tendency and Variability: The Frequency Rating Study (4.1)
4.15 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 5 Normal Distributions
5.1 Normal Distributions
What Are Normal Distributions?
Why Are Normal Distributions Important?
Is There Only One Normal Distribution?
5.2 The Standard Normal Distribution
What Is the Standard Normal Distribution?
Interpreting z-Scores: Distance From the Mean
Interpreting z-Scores: Relative Position Within the Distribution
Finding the Area Between the Mean and a z-Score

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Finding the Area Less or Greater Than a z-Score
Finding the Area Between Two z-Scores
5.3 Applying z-Scores to Normal Distributions
Transforming Normal Distributions Into the Standard Normal Distribution
Understanding the z-Score Formula
Example: Transforming SAT Scores Into z-Scores
Interpreting z-Scores: Distance From the Mean
Interpreting z-Scores: Relative Position Within Distribution
Finding the Area Between the Mean and a Score
Finding the Area Less or Greater Than a Score
Finding the Area Between Two Scores
Transforming Scores Into z-Scores: Drawing Conclusions
5.4 Standardizing Frequency Distributions
Example: Measuring Politicians' Behavior
Standardizing a Frequency Distribution
Interpreting Standardized Scores
Using Standardized Scores to Compare or Combine Variables
Characteristics of Standardized Distributions
When Can You Apply the Normal Curve Table to a Standardized Distribution?
5.5 Looking Ahead
5.6 Summary
5.7 Important Terms
5.8 Formulas Introduced in This Chapter
z-Score
Standardized Score
5.9 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 6 Probability and Introduction to Hypothesis Testing

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

6.1 A Brief Introduction to Probability
- What Is Probability?
- Why Is Probability Important to Researchers?
- Applying Probability to Normal Distributions
- Applying Probability to Binomial Distributions

6.2 Example: Making Heads or Tails of the Super Bowl

6.3 Introduction to Hypothesis Testing
- State the Null and Alternative Hypotheses (H_0 and H_1)
  - The Null Hypothesis (H_0)
  - The Alternative Hypothesis (H_1)
- Make a Decision About the Null Hypothesis
  - Defining a Low" Probability of a Statistic: Setting Alpha (α)
  - Identifying the Values of the Statistic With a Low Probability
  - Stating a Decision Rule
  - Calculating a Value of a Statistic
  - Making the Decision Whether to Reject the Null Hypothesis
- Draw a Conclusion From the Analysis
- Relate the Result of the Analysis to the Research Hypothesis
- Summary of Steps in Hypothesis Testing

6.4 Issues Related to Hypothesis Testing: An Introduction
- The Issue of Proof" in Hypothesis Testing
- Errors in Decision Making
- Factors Influencing the Decision About the Null Hypothesis
  - Sample Size
  - Alpha (α)
  - Directionality of the Alternative Hypothesis

6.5 Looking Ahead

6.6 Summary

6.7 Important Terms

6.8 Formula Introduced in This Chapter

Probability

6.9 Exercises

Answers to Learning Checks

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 7 Testing One Sample Mean
7.1 An Example From the Research: Do You Read Me?
7.2 The Sampling Distribution of the Mean
Characteristics of the Sampling Distribution of the Mean
7.3 Inferential Statistics: Testing One Sample Mean (σ Known)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: z-Test for One Mean
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the z-Test for One Mean
Summary
7.4 A Second Example From the Research: Unique Invulnerability
7.5 Introduction to the t-Distribution
Characteristics of the Student t-Distribution
7.6 Inferential Statistics: Testing One Sample Mean (σ Not Known)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: t-Test for One Mean
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the t-Test for One Mean

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Summary
7.7 Factors Affecting the Decision About the Null Hypothesis
Sample Size
Alpha (α)
Directionality of the Alternative Hypothesis
7.8 Looking Ahead
7.9 Summary
7.10 Important Terms
7.11 Formulas Introduced in This Chapter
z-Test for One Mean
Population Standard Error of the Mean $\left(\mathbf{i}t\sigma_{\mathbf{i}t\overline{X}}\right)$
Degrees of Freedom (df) for One Sample
t-Test for One Mean
Standard Error of the Mean $\left(\mathbf{i}ts_{\mathbf{i}t\overline{X}}\right)$
7.12 Using SPSS
Testing One Sample Mean (σ Not Known): The Unique Invulnerability Study (7.6)
7.13 Exercises
Answers to Learning Checks
Learning Check 2
Learning Check 3
Learning Check 5
Learning Check 6
Answers to Odd-Numbered Exercises
Chapter 8 Estimating the Mean of a Population
8.1 An Example From the Research: Salary Survey
8.2 Introduction to the Confidence Interval for the Mean
Using the Sampling Distribution of the Mean to Create the Confidence Interval for the Mean
8.3 The Confidence Interval for the Mean (σ Not Known)
State the Desired Level of Confidence
Calculate the Confidence Interval and Confidence Limits
Calculate the Standard Error of the Mean $\left(s_{\overline{X}}\right)$

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Calculate the Degrees of Freedom (df) and Identify the Critical Value of t

Calculate the Confidence Interval and Confidence Limits

Draw a Conclusion About the Confidence Interval

Probability of an Interval or the Probability of a Population Mean?

Summary

8.4 The Confidence Interval for the Mean (σ Known)

State the Desired Level of Confidence

Calculate the Confidence Interval and Confidence Limits

Calculate the Population Standard Error of the Mean $\left(\sigma_{\overline{X}}\right)$

Identify the Critical Value of z

Calculate the Confidence Interval and Confidence Limits

Draw a Conclusion About the Confidence Interval

Summary

8.5 Factors Affecting the Width of the Confidence Interval for the Mean

Sample Size

Estimating the Sample Size Needed for a Desired Confidence Interval

Level of Confidence

Choosing Between Different Levels of Confidence: Probability Versus Precision

8.6 Interval Estimation and Hypothesis Testing

Concerns About Hypothesis Testing

The Interpretation of Statistical Significance and Nonsignificance

The Influence of Sample Size on the Decision About the Null Hypothesis

Interval Estimation as an Alternative to Hypothesis Testing

8.7 Looking Ahead

8.8 Summary

8.9 Important Terms

8.10 Formulas Introduced in This Chapter

Confidence Interval for the Mean (σ Not Known)

Confidence Interval for the Mean (σ Known)

Estimated Sample Size for the Confidence Interval for the Mean

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

8.11 Using SPSS
Calculating the confidence interval for the mean (The salary survey study (8.1))
8.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 9 Testing the Difference Between Two Means
9.1 An Example From the Research: You Can Just Wait
9.2 The Sampling Distribution of the Difference
Characteristics of the Sampling Distribution of the Difference
9.3 Inferential Statistics: Testing the Difference Between Two Sample Means
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: t-Test for Independent Means
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the t-Test for Independent Means
Summary
9.4 Inferential Statistics: Testing the Difference Between Two Sample Means (Unequal Sample Sizes)
An Example From the Research: Kids' Motor Skills and Fitness
Inferential Statistics: Testing the Difference Between Two Sample Means (Unequal Sample Sizes)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Relate the Result of the Analysis to the Research Hypothesis

Assumptions of the t-Test for Independent Means (Unequal Sample Sizes)

Summary

9.5 Inferential Statistics: Testing the Difference Between Paired Means

Research Situations Appropriate for Within-Subjects Research Designs

An Example From the Research: Web-Based Family Interventions

Inferential Statistics: Testing the Difference Between Paired Means

Calculate the Difference Between the Paired Scores

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Draw a Conclusion From the Analysis

Relate the Result of the Analysis to the Research Hypothesis

Assumptions of the t-Test for Dependent Means

Summary

9.6 Looking Ahead

9.7 Summary

9.8 Important Terms

9.9 Formulas Introduced in This Chapter

Degrees of Freedom (df), Difference Between Two Sample Means

t-Test for Independent Means

Standard Error of the Difference $\left(s_{\overline{X}_1-\overline{X}_2}\right)$

Standard Error of the Difference (Unequal Sample Sizes)

Degrees of Freedom (df), Difference Between Paired Means

t-Test for Dependent Means

Standard Error of Difference Scores (s D)

9.10 Using SPSS

Testing the Difference Between Two Sample Means: The Parking Lot Study (9.1)

Testing the Difference Between Paired Means: The Web-Based Intervention Study (9.5)

9.11 Exercises

Answers to Learning Checks

Learning Check 1

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Answers to Odd-Numbered Exercises
Chapter 10 Errors in Hypothesis Testing, Statistical Power, and Effect Size
10.1 Hypothesis Testing vs. Criminal Trials
10.2 An Example From the Research: Truth or Consequences
Correct Decisions and Mistakes in Criminal Trials
10.3 Two Errors in Hypothesis Testing: Type I and Type II Error
Type I Error: The Risk in Rejecting the Null Hypothesis
What Is the Probability of Making a Type I Error?
What Is the Probability of Not Making a Type I Error?
Why Is Type I Error a Concern?
Type II Error: The Risk in Not Rejecting the Null Hypothesis
What Is the Probability of Making a Type II Error?
What Is the Probability of Not Making a Type II Error?
Why Is Type II Error a Concern?
Why Does Type II Error Occur?
Type I and Type II Error: Summary
10.4 Controlling Type I and Type II Error
Controlling Type I Error
Concerns About Controlling Type I Error
Controlling Type II Error
Increasing Sample Size
Raising Alpha (α)
Using a Directional Alternative Hypothesis (H1)
Increasing Between-Group Variability
Decreasing Within-Group Variability
Controlling Type I and Type II Error: Summary
10.5 Measures of Effect Size
Interpreting Inferential Statistics: The Influence of Sample Size
The Inappropriateness of “Highly Significant”

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Effect Size as Variance Accounted for

Measure of Effect Size for the Difference Between Two Sample Means: r^2

Interpreting Measures of Effect Size

Presenting Measures of Effect Size

Reasons for Calculating Measures of Effect Size

A Second Measure of Effect Size for the Difference Between Two Sample Means: Cohen's d

10.6 Looking Ahead

10.7 Summary

10.8 Important Terms

10.9 Formulas Introduced in This Chapter

Probability of Type I Error

Probability of Type II error

r Squared (r^2)

Cohen's d

10.10 Exercises

Answers to Learning Checks

Learning Check 1

Learning Check 2

Learning Check 3

Learning Check 4

Answers to Odd-Numbered Exercises

Chapter 11 One-Way Analysis of Variance (ANOVA)

11.1 An Example From the Research: It's Your Move

11.2 Introduction to Analysis of Variance (ANOVA)

Review: Testing the Difference Between the Means of Two Samples

Understanding Between-Group and Within-Group Variability

Relating Between-Group and Within-Group Variability

Introduction to Analysis of Variance (ANOVA) and the F-Ratio

Characteristics of the F-Ratio Distribution

11.3 Inferential Statistics: One-Way Analysis of Variance (ANOVA)

State the Null and Alternative Hypotheses (H_0 and H_1)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Value, and State a Decision Rule
Calculate a Statistic: F-Ratio for the One-Way ANOVA
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (R^2)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Summary
11.4 A Second Example: The Parking Lot Study Revisited
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Value, and State a Decision Rule
Calculate a Statistic: F-Ratio for the One-Way ANOVA
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (R^2)
Draw a Conclusion From the Analysis
The Relationship Between the t-Test and the F-Ratio
11.5 Analytical Comparisons Within the One-Way ANOVA
Planned Versus Unplanned Comparisons
Between-Group and Within-Group Variance in Analytical Comparisons
Conducting Planned Comparisons
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Conducting Unplanned Comparisons
Concerns Regarding Unplanned Comparisons
Methods of Controlling Familywise Error

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Controlling Familywise Error: The Tukey Test

11.6 Looking Ahead

11.7 Summary

11.8 Important Terms

11.9 Formulas Introduced in This Chapter

Degrees of Freedom for Between-Group Variance, One-Way ANOVA (df_{BG})

Degrees of Freedom for Within-Group Variance, One-Way ANOVA (df_{WG})

Between-Group Variance, One-Way ANOVA (MS_{BG})

Total Sample Mean, One-Way ANOVA $\left(\overline{X}_T\right)$

Within-Group Variance, One-Way ANOVA (MS_{WG})

F-Ratio, One-Way ANOVA (F)

Measure of Effect Size, One-Way ANOVA (R^2)

Relationship Between t-Test and F-Ratio for the One-Way ANOVA

F-Ratio, Analytical Comparison (F_{comp})

Between-Group Variance, Analytical Comparison $\left(MS_{\mathrm{BG}_{\mathrm{comp}}}\right)$

Measure of Effect Size, Analytical Comparison $\left(R^2_{\mathrm{comp}}\right)$

Critical Value, Tukey Test for Unplanned Comparisons (F_T)

11.10 Using SPSS

One-Way Analysis of Variance (ANOVA): The Chess Study (11.1)

Analytical Comparisons: Planned Comparisons—Explanation vs. Prediction (11.5)

11.11 Exercises

Answers to Learning Checks

Learning Check 2

Learning Check 3

Learning Check 4

Answers to Odd-Numbered Exercises

Chapter 12 Two-Way Analysis of Variance (ANOVA)

12.1 An Example From the Research: Vote—or Else!

12.2 Introduction to Factorial Research Designs

Testing Interaction Effects

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

An Example of Interaction Effect

An Example of No Interaction Effect

Testing Main Effects

The Relationship Between Main Effects and Interaction Effects

Factorial Designs vs. Single-Factor Designs: Advantages and Disadvantages

12.3 The Two-Factor (A × B) Research Design

Notational System for Components of the Two-Factor Research Design

Analyzing the Two-Factor Research Design: Descriptive Statistics

Notational System for Means in the Two-Factor Research Design

Examining the Descriptive Statistics

12.4 Introduction to Analysis of Variance (ANOVA) for the Two-Factor Research Design

F-Ratios in the Two-Factor Research Design

12.5 Inferential Statistics: Two-Way Analysis of Variance (ANOVA)

State the Null and Alternative Hypotheses (H_0 and H_1)

Main Effect of Factor A

Main Effect of Factor B

A × B Interaction Effect

Make Decisions About the Null Hypotheses

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Values, and State Decision Rules

Calculate Statistics: F-Ratios for the Two-Way ANOVA

Make Decisions Whether to Reject the Null Hypotheses

Determine the Levels of Significance

Calculate Measures of Effect Size (R^2)

Draw Conclusions From the Analyses

Relate the Results of the Analyses to the Research Hypothesis

12.6 Investigating a Significant A × B Interaction Effect: Analysis of Simple Effects

12.7 Looking Ahead

12.8 Summary

12.9 Important Terms

12.10 Formulas Introduced in This Chapter

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Degrees of Freedom for Main Effect of Factor A, Two-Way ANOVA (df_A)
Degrees of Freedom for Main Effect of Factor B, Two-Way ANOVA (df_B)
Degrees of Freedom for A × B Interaction Effect, Two-Way ANOVA ($df_{A\times B}$)
Degrees of Freedom for Within-Group Variance, Two-Way ANOVA (dfWG)
Between-Group Variance for Main Effect of Factor A, Two-Way ANOVA (MS_A)
Between-Group Variance for Main Effect of Factor B, Two-Way ANOVA (MS_B)
Sums of Squares for Between-Group Variance, Two-Way ANOVA (SS_{BG})
Between-Group Variance for A × B Interaction, Two-Way ANOVA ($MS_{A\times B}$)
Within-Group Variance, Two-Way ANOVA (MS_{WG})
F-Ratio for Main Effect of Factor A, Two-Way ANOVA (F_A)
F-Ratio for Main Effect of Factor B, Two-Way ANOVA (F_B)
F-Ratio for A × B Interaction, Two-Way ANOVA ($F_{A\times B}$)
Measure of Effect Size for Main Effect of Factor A, Two-Way ANOVA(R^2_A)
Measure of Effect Size for Main Effect of Factor B, Two-Way ANOVA(R^2_B)
Measure of Effect Size for A × B Interaction, Two-Way ANOVA(R^2_{AB})
12.11 Using SPSS
Two-Way Analysis of Variance (ANOVA): The Voting Message Study (12.1)
12.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 13 Correlation
13.1 An Example From the Research: It's Good for You!
13.2 Introduction to the Concept of Correlation
Describing the Relationship Between Variables
Nature of the Relationship
Direction of the Relationship
Strength of the Relationship

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Measuring the Relationship Between Variables: Correlational Statistics
Introduction to the Pearson Correlation Coefficient (r)
Calculating Correlational Statistics: The Role of Variance and Covariance
13.3 Inferential Statistics: Pearson Correlation Coefficient
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: Pearson Correlation Coefficient (r)
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (r^2)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the Pearson Correlation
A Computational Formula for the Pearson r
Represent the Covariance Between X and Y (SPXY)
Represent the Variance of Variable X (SSX)
Represent the Variance of Variable Y (SSY)
Calculate the Pearson Correlation Coefficient (r)
13.4 Correlating Two Sets of Ranks: The Spearman Rank-Order Correlation
Introduction to Ranked Variables
An Example From the Research: Do Students and Teachers Think Alike?
Inferential Statistics: The Spearman Rank-Order Correlation Coefficient
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
13.5 Correlational Statistics vs. Correlational Research
Does ANOVA = Causality?
“Correlation Is Not Causality”—What Does That Mean?
13.6 Looking Ahead

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

13.7 Summary
13.8 Important Terms
13.9 Formulas Introduced in This Chapter
Degrees of Freedom for Pearson Correlation (df)
Sum of Products (Definitional Formula) (SPXY)
Sum of Squares for the X Variable (Definitional Formula) (SSX)
Sum of Squares for the Y Variable (Definitional Formula) (SSY)
Pearson Correlation (Definitional Formula) (r)
Measure of Effect Size for the Pearson Correlation Coefficient (r^2)
Sum of Products (Computational Formula) (SPXY)
Sum of Squares for the X Variable (Computational Formula) (SSX)
Sum of Squares for the Y Variable (Computational Formula) (SSY)
Pearson Correlation (Computational Formula) (r)
Spearman Rank-Order Correlation (rs)
Measure of Effect Size for the Spearman Rank-Order Correlation $\left(r_s^2\right)$
13.10 Using SPSS
Pearson Correlation: The Picky Eater Study (13.1)
Spearman Rank-Order Correlation: The Business Student Study (13.4)
13.11 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Answers to Odd-Numbered Exercises
Chapter 14 Linear Regression and Multiple Correlation
14.1 Predicting One Variable From Another: Linear Regression
The Linear Regression Equation
Calculating the Linear Regression Equation
Calculate the Slope of the Equation (b)
Calculate the Y-Intercept of the Equation (a)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Report the Linear Regression Equation

Drawing the Linear Regression Equation

A Second Example: Put Down Your Pencils

Calculating the Linear Regression Equation

Drawing the Linear Regression Equation

14.2 Correlation With Two or More Predictors: Introduction to Multiple Correlation and Regression

The Limitations of Using Only One Predictor

Introduction to Multiple Correlation

Multiple Correlation With Two Predictors: An Example

Measuring the Relationship Between Two Predictors and a Criterion: The Multiple Correlation (R)

Testing the Multiple Correlation for Statistical Significance

Predicting a Criterion From Two Predictors: The Multiple Regression Equation

Conclusion

14.3 Looking Ahead

14.4 Summary

14.5 Important Terms

14.6 Formulas Introduced in This Chapter

Linear Regression Equation

Estimate of Slope (b)

Estimate of Y-Intercept (a)

Multiple Correlation Coefficient (R)

F-Ratio for the Multiple Correlation Coefficient (F)

Multiple Regression Equation

Regression Coefficient

Estimate of Y-Intercept (a) (Multiple Regression)

14.7 Using SPSS

Linear Regression: The Picky Eater Study (14.1)

Multiple Regression: The College Admission Example (14.2)

14.8 Exercises

Answers to Learning Checks

Learning Check 1

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 2

Answers to Odd-Numbered Exercises

Chapter 15 Chi-Square

15.1 An Example From the Research (One Categorical Variable): Are You My Type?

15.2 Introduction to the Chi-Square Statistic

Observed and Expected Frequencies: The Basis of the Chi-Square Statistic

Characteristics of the Distribution of Chi-Square Statistics

15.3 Inferential Statistic: Chi-Square Goodness-of-Fit Test

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Value, and State a Decision Rule

Calculate a Statistic: Chi-Square (χ^2)

Make a Decision Whether to Reject the Null Hypothesis

Determine the Level of Significance

Calculate a Measure of Effect Size (Cramér's φ)

Draw a Conclusion From the Analysis

Relate the Result of the Analysis to the Research Hypothesis

Assumptions Underlying the Chi-Square Statistic

Conducting the Chi-Square Goodness-of-Fit Test With Unequal Hypothesized Proportions

Calculate Expected Frequencies (f_e)

Calculate the Chi-Square Statistic (χ^2)

15.4 An Example From the Research (Two Categorical Variables): Seeing Red

15.5 Inferential Statistic: Chi-Square Test of Independence

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Value, and State a Decision Rule

Calculate a Statistic: Chi-square (χ^2)

Make a Decision Whether to Reject the Null Hypothesis

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Determine the Level of Significance
Calculate a Measure of Effect Size (Cramér's φ)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
15.6 Parametric and Nonparametric Statistical Tests
Parametric vs. Nonparametric Statistical Tests
Reasons for Using Nonparametric Statistical Tests
Examples of Nonparametric Statistical Tests
15.7 Looking Ahead
15.8 Summary
15.9 Important Terms
15.10 Formulas Introduced in This Chapter
Degrees of Freedom for the Chi-Square Goodness-of-Fit Test
Expected Frequency for the Chi-Square Goodness-of-Fit Test
Chi-Square Statistic (χ^2)
Cramér's φ for the Chi-Square Goodness-of-Fit Test
Degrees of Freedom for the Chi-Square Test of Independence
Expected Frequency for the Chi-Square Test of Independence
Cramér's φ for the Chi-Square Test of Independence
15.11 Using SPSS
Chi-Square Goodness-of-Fit Test—Equal Hypothesized Proportions: The Personality Study (15.1)
Chi-Square Goodness-of-Fit Test—Unequal Hypothesized Proportions: The Personality Study (15.3)
Chi-Square Test of Independence: The Color Preference Study (15.4)
Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Learning Check 6

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Odd-Numbered Exercises

Appendix A: Tables

Appendix B: Review of Basic Mathematics

I. Symbols

II. Terms

III. Addition, Subtraction, Multiplication, Division

A. All Positive Numbers

B. Adding All Negative Numbers

C. Adding a Combination of Positive and Negative Numbers

D. Subtracting Negative Numbers

E. Multiplying Two Negative Numbers

F. Multiplying a Combination of Positive and Negative Numbers

G. Dividing Two Negative Numbers

H. Dividing One Positive Number and One Negative Number

IV. Order of Operations

V. Fractions, Decimals, Percentages

A. Fraction

B. Decimal

C. Percentage

VI. Solving Simple Equations With One Unknown (X)

A. X Has a Value Subtracted From It (Solve for X Using Addition)

B. X Has a Value Added to It (Solve for X Using Subtraction)

C. X Is Multiplied by a Value (Solve for X Using Division)

D. X Is Divided by a Value (Solve for X Using Multiplication)

VII. Solving More Complex Equations With One Unknown (X)

A. X Is Multiplied by a Value and Then Has a Value Either Added to It or Subtracted From It (Solve for X Using Either Subtraction or Addition and Then Division)

B. X Is Divided by a Value and Then Has a Value Either Added to It or Subtracted From It (Solve for X Using Either Addition or Subtraction and Then Multiplication)

C. X Has a Value Added to It or Subtracted From It and Is Then Multiplied by a Value (Solve for X Using Division and Then Either Subtraction or Addition)

D. X Has a Value Either Added to It or Subtracted From It and Is Then Divided by a Value (Solve for X Using Multiplication and Then Either Subtraction or

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Addition)
Glossary
References
Index

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

PREFACE

A Few (?) Words About Me

When I entered college, I thought my path would take me to law school; however, taking an Intro to Psych class made me realize I enjoy the challenge of trying to understand the human mind. The college I attended, U.C. Santa Cruz, was a fairly unconventional university, but somehow in the midst of a sea of humanistic psychologists I became attracted to the empirical and methodological aspects of psychology by taking classes from instructors such as David Harrington and Dane Archer, who showed me that statistics could appeal to students if taught with a gentle, guiding approach that addresses questions relevant to students' lives. After graduating from college, I was able to get a job as a research assistant for a human resource consulting firm. Despite my lack of work experience, I was hired primarily as a result of having taken statistics and research methods courses, which taught me that learning statistics has benefits both inside and outside of the classroom.

Several years later, I started grad school at U.C. Berkeley, where two events critical to this book took place. First, serving as a teaching assistant, I found I really enjoyed helping students, particularly in statistics and research methods classes often approached with fear and suspicion. Second, I took graduate classes from Geoff Keppel, who had developed his own unique system for analyzing experimental research designs. His lectures and books showed me that statistics can be taught in a systematic way that highlights similarities rather than differences between different research situations. Geoff managed to transform something as daunting-sounding as a "3 x 2 x 4 research design" into the mathematical equivalent of playing with wooden toy alphabet blocks labeled "A," "B," and "C."

After leaving Cal, I was hired by San Jose State, where I teach an introductory statistics course aimed at students with a wide range of backgrounds, ability, and motivation. When I went to select a textbook for the course, for the first time I looked carefully at the wide range of offerings. What I found striking (and still find striking) was that the majority of books focused on providing formulas and very small sets of data designed to demonstrate how to correctly calculate the correct numbers from these formulas. Little emphasis, however, was given to what these numbers meant or implied. Given my own experiences learning and using statistics, I thought a book was needed that focusses on how statistics may be used to answer questions and test ideas within the larger research process.

Goal of this Book

This textbook, aimed primarily at students majoring in the behavioral and social sciences, includes many of the same topics and chapters as the majority of existing introductory statistics books. However, what I believe distinguishes this book is its thematically-based approach that places statistics within the research process. The primary purpose of this

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

book is to not just teach students how to calculate statistics but how to interpret the results of statistical analyses in light of a study's research hypothesis, and to communicate one's results and interpretations to a broader audience. Hopefully, this book will not only help students understand the purpose and use of statistics, but it will also give them a greater appreciation of how research studies are conceived, conducted, and communicated.

The book is intended for students with a limited knowledge of mathematics, statistics, and research design and methodology. Some common obstacles faced by students are their math anxiety, fear of conducting arithmetic calculations, and an inability to see the relevance of the material to their daily lives. To address these concerns, this books attempts to judiciously use mathematical terminology, write in a conversational manner, and utilize actual research situations rather than contrived sets of numbers.

Organization of the Book

The 15 chapters of this book may be placed into three general categories. Chapters 1 to 4 are designed to introduce students to the research process and how data that have been collected may be organized, presented, and summarized:

Chapter 1 (Introduction to Statistics) defines the concept of statistics, explains why statistics are important to learn, and describes the steps in the research process that are the framework for the rest of the book.

Chapter 2 (Examining Data: Tables and Figures) discusses the importance of examining data that have been collected and describes how tables and figures may be created to represent data.

Chapter 3 (Measures of Central Tendency) presents measures of central tendency (the mode, median, and the mean) and compares them in terms of their relative strengths and weaknesses.

Chapter 4 (Measures of Variability) presents measures of variability (range, interquartile range, variance, standard deviation, index of qualitative variation) and the concept of parameters (population variance and standard deviation).

Chapters 5 through 10 discuss the process of conducting statistical analyses to test research questions and hypotheses, as well as issues and controversies regarding this process:

Chapter 5 (Normal Distributions) introduces theoretical distributions, starting with normal distributions, before moving on to the standard normal distribution (normal curve).

Chapter 6 (Probability and Introduction to Hypothesis Testing) presents the structure for later chapters by introducing, defining, and illustrating steps in hypothesis testing.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Chapter 7 (Testing One Sample Mean) tests one sample mean using the *z* -test and *t* -test and introduces sampling distributions, the central limit theorem, and the Student *t* distribution.

Chapter 8 (Estimating the Mean of a Population) introduces interval estimation by developing the confidence interval for the mean and discusses interval estimation as an alternative to null hypothesis significance testing (NHST).

Chapter 9 (Testing the Difference Between Two Means) covers the *t* -test for independent means for both equal and unequal sample size and introduces within-subjects research designs using the *t* -test for dependent means.

Chapter 10 (Errors in Hypothesis Testing, Statistical Power, and Effect Size) discusses Type I and Type II error and introduces measures of effect size as a way of drawing appropriate conclusions about an analysis.

Chapters 11 to 15 discuss different statistical procedures used in research situations that vary in the number of independent variables as well as how the variables have been measured:

Chapter 11 (One-way Analysis of Variance (ANOVA)) discusses the one-way ANOVA, the relationship between the *t* -statistic and the F-ratio, and planned and post-hoc comparisons.

Chapter 12 (Two-way Analysis of Variance (ANOVA)) introduces factorial research designs, the calculation of F-ratios for main effects and interaction effects, and illustrates the use of simple effects analyses to test research hypotheses.

Chapter 13 (Correlation) defines and illustrates the concept of correlation, discusses the Pearson and Spearman correlation coefficients, and makes the distinction between correlational statistics and correlational research methods.

Chapter 14 (Linear Regression and Multiple Correlation) introduces the concept of regression, discusses linear regression, and works through an example of multiple regression with two predictor variables.

Chapter 15 (Chi-Square) describes the analysis of categorical variables using the chi-square goodness-of-fit test and the chi-square test of independence before ending the book with a brief introduction to nonparametric statistics.

In most cases, the chapters in this book share a uniform format and structure, centered on the research process, based on examples of actual published research studies addressing topics relevant to students' lives. For each example, the research hypothesis is clearly identified; after a brief description of how the data in the study were collected, the calculation and interpretation of both descriptive and inferential statistics are presented,

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ending with the extent to which the results of the statistical analyses support the research hypothesis and how these findings may be communicated to a broader audience.

Pedagogical Features

The chapters in this textbook contain several reoccurring features designed to enhance students' comprehension of the material. These include the following:

Periodic learning checks within chapters. Short learning checks, in the form of conceptual review questions and computational exercises, are provided throughout each chapter.

Explicit presentation of formulas and calculations. This book provides both definitional formulas to highlight their relation to the concept being measured and computational formulas to facilitate calculations. In applying these formulas, all of the steps in the calculation of statistics such as the *t* -test and F-ratio are included to help students in completing exercise and homework assignments.

Illustration of the presentation and communication of analyses. Every chapter includes a discussion and illustration of how statistical analyses may be reported as well how they may be communicated to a larger audience in both written and visual form.

Highlighted definition of terms. Every term and concept introduced in this book is boldfaced, immediately followed by an explicit definition.

End of chapter summaries and exercises. The end of each chapter contains a summary of critical points, a list of terms and formulas introduced in the chapter, and a large number of exercises, many of which refer to published research studies.

Calculation of statistics using SPSS. Many chapters include end-of-chapter content that provides an annotated presentation of how statistics may be calculated using the SPSS software. Screen shots of data files, menu selections, and output files are included and annotated to locate and identify relevant aspects of the analysis.

What's New to the 3rd Edition

I'd like to start by expressing my most heartfelt appreciation to the instructors and reviewers of the first two editions of this book as their evaluation, insights, criticisms, and suggestions were absolutely vital to the changes made in this third edition. I'd like to particularly thank the instructors who've adopted the book who were kind enough to spontaneously send me their thoughts and suggestions. It is my sincere hope that all of you see the incorporation of your input in this edition.

There were two main goals in this edition. The first goal was to make the writing as brief and concise as possible, manifested in the goal of reducing the length of each chapter by 5% to

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

10%. This goal was achieved by simplifying the language and shortening or eliminating what I, the book's adopters, and the book's reviewers felt was unnecessary detail or discussion. The second goal was to increase and update the exercises at the end of many of the chapters, including the use of recent published research as the basis of these exercises. Another goal was to update the demonstration of the use of SPSS to analyze data using the most recent version (Version 31 at the time of this book's publication).

More specifically, in terms of content, the following changes were made to these chapters:

Chapter 1 (Introduction to Statistics): The published research studies used to illustrate concepts in the text, learning checks, and exercises were added to and updated using topics relevant to students (e.g., fear of missing out (FOMO), cryptocurrency, legalization of cannabis, Gen Z, COVID, fake news).

Chapter 6 (Probability and Introduction to Hypothesis Testing): The Super Bowl example used to introduce and illustrate the process of hypothesis testing was updated.

Chapter 8 (Estimating the Mean of a Population): The main research example, which involves data from an actual national salary survey, has been updated throughout this chapter with more current salary data.

As with the first two editions, any and all questions, concerns, suggestions, and complaints you have about the book are most welcome and can be sent to me at howard.tokunaga@sjsu.edu .

A Few Tips for Students

To you, the college student about to read this book as part of taking a statistics course: "Welcome!" and "Great job!" We welcome you because you're about to embark on a semester-long journey that we hope enhances your skills and widens your perspective; we congratulate you because it's a journey not everyone is willing to take. We say to our students that by taking just one statistics course they are ahead of 95% of the population in their knowledge of statistics.

We also tell our students that our primary goal for them is not to become a statistics genius but just to become comfortable with the process of analyzing and interpreting data. In that regard, we agree with a 2012 *Forbes* magazine article entitled *The Six Classes That Will Make Any College Grad Employable* that noted "a good statistics class will help the student become comfortable pushing numbers around and avoiding false interpretations."

At the present moment, we know our encouragement may be of little comfort because you might be anxious about having to learn statistics. Some of you might be anxious about having to learn statistical concepts and formulas; others might be anxious because the research process seems pretty complicated. Talking with students who have taken our courses over the years has helped us assemble the following advice:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Master the Material Presented in the Early Chapters

Chapters 1 through 6 discuss how research is typically conducted, how data are summarized and described, and the process by which researchers conduct statistical analyses to test their ideas. It is absolutely critical for you to have a firm understanding of these chapters as they lay the groundwork for later chapters that discuss a variety of statistical procedures used by researchers to test hypotheses.

What does it mean to master this material? First, *read the chapters both before and after they're discussed in class* . Reading the chapters beforehand helps you identify anything that's unclear and have your questions addressed by your instructor; re-reading the chapters afterward helps confirm your understanding of the material. Next, *be able to define and explain key conceptspresented in these chapters* . These concepts are highlighted in bold-faced letters, and it's important to learn them when they're first introduced because they'll appear throughout the remainder of the book. Next, *do the learning checks within the chapters and the exercises at the end of the chapters* , because they include exercises and review questions you can ask yourself to assess your understanding of the material. Most importantly, *do not miss class during the early part of the semester* ! It's been our experience that students who miss critical lectures at the beginning of the term often have difficulty keeping up as the semester continues.

Review High School Algebra

If you happen to move beyond the introductory statistics course for which you're reading this book, you'll find that statistics are often calculated using computers and statistical software rather than hand calculations. Consequently, some statistics textbooks no longer include mathematical formulas but instead have students use statistical software or Microsoft Excel. However, this book takes a different approach for two main reasons. First, we believe that in order to understand the results of statistical analyses, one must understand their underlying foundation. Learning statistics via computer software can lead to a "brain-dead" approach in which students are at the mercy (rather than control) of their computers. We refer to this as the "I'm only as smart as my printout" method of learning statistics, also known as "Because this is what my computer told me."

Second, this book contains mathematical formulas and calculations for one simple reason: The statistics in this book are not hard to calculate. To assess whether you're prepared to read this book, take the following test:

1. Do you know how to add, subtract, multiply, and divide?
2. Do you know how to calculate the square or square root of a number?
3. Do you understand the order of operations and parentheses within mathematical formulas?

If your answers to these questions are "Yes," congratulations! You possess the ability to conduct *every* mathematical calculation in this book because *none* of the formulas involve geometry, trigonometry, or calculus. However, if you're unsure of your mathematical ability, we suggest you review your high school algebra. Also, there's an appendix at the end of this

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

book that includes the mathematical concepts and operations needed to conduct the statistical analyses in this book. We highly recommend you read this appendix to review and assess your mathematical skills.

Learn to Use a Statistical Calculator

Although learning the statistics in this book doesn't require the use of computers or software, you will need a hand calculator. Calculators with statistical capabilities are typically identified as "statistical" or "scientific" calculators. Although there's a wide selection of brands and models, know that the calculators most appropriate for this book cost only $10 to $15. We do *not* recommend you use or purchase a more expensive calculator! We prefer these simpler calculators because students are able to do their calculations more quickly and with fewer errors. We do suggest you purchase a calculator similar to the one your instructor uses as they will be more familiar with its features and idiosyncrasies.

In selecting a calculator, look closely at the keyboard; you'll need a calculator whose keys have labels such as $\bar{X}$, σ, and ΣX (Chapters 3 and 4 discuss what these symbols represent). When you go calculator shopping, you'll find calculators that have keys with labels such as ΣY and XY. Although calculators with these Y symbols are only slightly more expensive than calculators with just the X symbols, we do *not* recommend their purchase because these features are not needed to perform the calculations in this book. In fact, we've found students with Y calculators have a more difficult time performing simpler calculations.

Ask Questions

Conducting research is the process of asking and answering questions. Accordingly, throughout this book issues are often framed in the form of questions, which we hope encourages you to ask questions as well. The first person you should direct any questions to is yourself. As you work your way through the chapters, get into the habit of asking yourself, "What does this mean?" A good test of your level of comprehension is whether you can explain the material to yourself in a meaningful way.

You should also be sure to work with your instructor to confirm your level of understanding and clarify unanswered questions. Instructors often say to students, "The only bad question is the one you don't ask." Don't be hesitant about asking questions in class. Instructors will tell you they use one student's question to assess an entire class's level of understanding. This is what we call the "pencil test"—when one student asks a question, we look to see how many other pencils get raised, which reflects the number of students who had the same question. Asking questions enhances the learning experience for yourself, your classmates, and your instructor.

Form a Study Group

No matter how much we encourage students to ask questions in class, it seems they hesitate for fear of drawing attention to themselves. As a result, we turn to another well-

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

traveled saying: “Misery loves company.” Forming a study group with other students will help you keep up with reading assignments, confirm your comprehension of the material, clarify any confusion you have, prepare for exams, and provide social support. We do suggest that your group contain at least one person who will raise the group’s questions and concerns to the instructor during class or office hours.

Practice, Practice, Practice

The best way to learn the topics covered by this book is to obtain as much practice as possible. Throughout the chapters are examples in which the calculations have been worked out for you; in the middle and end of each chapter are a number of exercises. The answers to some of the exercises are provided at the back of this book, and your instructor has access to the answers to the other exercises. Do the end-of-chapter exercises even if your instructor does not assign them as homework in order to identify any reoccurring mistakes you make. We find that the vast majority of errors students make are simple computational errors, committed when students do the calculations too quickly.

Also note that the publisher of this book, Sage Publications, has developed a webpage for this book that contains a huge amount of supplemental tools such as videos and quizzes for each chapter. Find it by going to the Sage Publishing website **collegepublishing.sagepub.com** , searching for my name (Tokunaga), and looking for the **Resources** tab.

Enough cheerleading and preaching – let’s get started!

WELCOME!

GREAT JOB!

Additional Teaching Resources

Visit **collegepublishing.sagepub.com** , and navigate to the Resources tab on your book’s page to find the teaching materials designed to accompany this textbook. On this site you will find an array of materials that will save you time and to help you keep students engaged, including the following:

Learning management system cartridges that easily integrate with your course management system so student test results and graded assignments seamlessly flow into your gradebook

Test banks , aligned to Bloom’s Taxonomy, that provide a diverse range of test items, including multiple choice, true/false, and essay questions

Lecture notes that provide an outline and the key concepts in each chapter to aid in lecture preparation

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

PowerPoint ® slides that offer a flexible, accessible, and customizable solution for creating multimedia lectures

Tables and Figures from the book to support lecture preparation and class discussions

Sample course syllabi , including suggested models for structuring your course

Answers to even-numbered questions from the text help facilitate grading

SPSS datasets for use with exercises from the text.

Multimedia content , including original SAGE videos featuring tutorials with author Howard T. Tokunaga that bring concepts to life and appeal to diverse learners

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ACKNOWLEDGMENTS

This book has traveled a long journey, with many people playing both transient and persevering roles in its development. First and foremost, I'd like to thank my kids, Meagan and Will, for providing me perspective and inspiration amid all of the twists and turns we faced while this book was being written. I'm so very blessed to have you as my "burden." Away from home, I'm grateful to my graduate students at San Jose State—working on this book has given me much empathy for their thesis-related travails. However, they should be aware that the completion of this book ends the unspoken agreement that I not ask, "So how's your thesis?" if they don't ask, "So how's your book?" A number of friends have provided most welcome relief after long days staring at a computer screen, with particular thanks going to Ellie, Eileen, Susan, Bonnie, Ann, Lisa, and Terry. I would like to thank Charles Linsmeier and David Shirley for their many helpful comments and advice on earlier versions of this book, Julie West for her help in developing chapter exercises, and Youn-Jeong Choi and Naomi Ran for their careful examination of the chapters and exercises.

I would like to express my appreciation to everyone at SAGE who has played a role in shepherding this book to completion. First and foremost, I must express my gratitude to my editor, Vicki Knight. It has been a joy and an honor to work with someone with such a truly exceptional combination of experience, knowledge, perspective, and humor. In production I was fortunate enough to work with others on the SAGE team who made the process a pleasure: Yvonne McDuffee, editorial assistant; Katie Guarino, assistant editor; Nicole Elliott, executive marketing manager; Candice Harman, cover designer; and the copyeditor, Michelle Ponce. Furthermore, I am grateful to the reviewers who took the time to read the chapters, raise questions, and provide constructive comments and suggestions to assist me in clarifying and strengthening the manuscript. The following are the reviewers for third edition:

Erin M. Fekete, University of Indianapolis

Brooke Magnus, Boston College

Loretta Neal McGregor; Arkansas State University

Michael T. Moore, Adelphi University

Megan M. Ringel, St. Edward's University

Bettina Viereck, University of Hartford

Janice Weaver, Community College of Allegheny County

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ABOUT THE AUTHOR

Howard T. Tokunaga is professor of Psychology at San Jose State University, where he serves as coordinator of the MS Program in Industrial/Organizational (I/O) Psychology and teaches undergraduate and graduate courses in statistics, research methods, and I/O psychology. He received his bachelor's degree in psychology at U.C. Santa Cruz and his PhD in psychology at U.C. Berkeley. In addition to his teaching, he has consulted with a number of public sector and private sector organizations on a wide variety of management and human resource issues. He is author of *Moving from IBM SPSS to R and RStudio: A Statistics Companion* , and coauthor (with G. Keppel) of *Introduction to Design and Analysis* : *A Student's Handbook* . In his spare time, he enjoys the outdoors, museums, live theater, and being as physically, socially, and politically active as possible.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CHAPTER 1 INTRODUCTION TO STATISTICS

Learning Objectives

1.1 Define the concept of statistics.

1.2 Describe different reasons for learning statistics.

1.3 Describe the main steps in the research process and the tasks and concepts within each step.

In introducing this book to you, we assume you're a college student taking what is perhaps your first course in statistics to fulfill a requirement for your major or a general education requirement. If so, you may be asking yourself two questions:

What is statistics?

Why learn statistics?

The ultimate goal of this book is to help you begin to answer these two questions.

1.1 What Is Statistics?

You may not know it, but you encounter a variety of "statistics" in your day-to-day activities: the typical cost of going to college, the yearly income of the average college graduate, the average price of a car, and so on. So what exactly is "statistics"?

The Merriam-Webster dictionary defines **statistics** as *a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data* . When most people think about statistics, they focus on the "analysis" aspect of this definition—that is to say, they focus on numbers that result from analyzing data. However, statistics not only is concerned about how data are *analyzed* but also how data are *collected* and how the results of statistical analyses are *interpreted* and *communicated* . The main purpose of this book is to introduce, describe, and illustrate the role of statistics within the larger research process.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1.2 Why Learn Statistics?

There are many reasons why you should learn statistics. First, not only do you currently encounter statistics in your daily activities, but *throughout your life you are affected by the results of research and statistical analyses* . Which college you attend is based in part on test scores developed by researchers. You may have to take a personality or intelligence test to get a job. The choices of drugs and medicines available to you are based on medical research. Learning about statistics will help you become a more informed and aware consumer of research and statistical analyses that affect many aspects of your life.

A second reason for learning statistics is that *you may be required to interpret the results of statistical analyses* . Many college courses require students to read academic research journal articles. Evaluating published research is complicated by the fact that different people studying the same topic may come up with diverse or even opposing conclusions. Understanding statistics and their role in the research process will help you decide whether conclusions drawn in research articles are appropriate and justified.

Another reason for learning statistics is that *you might be required to do your own research* . College courses sometimes have students design and conduct mini-research studies; undergraduate majors might encourage students to do senior honors theses; graduate programs often require masters' theses and doctoral dissertations. Learning to collect and analyze data will help you address your own questions in an objective, systematic manner.

A final reason for learning statistics is that it may *help you in your future career* . The website Careercast.com conducts an annual survey in which they evaluate jobs on four dimensions: environmental factors, income, employment outlook, and stress. Recently, their #1 rated job was "data scientist," and the #2 rated job was "statistician." They noted, "A statistician's skill set can be used to break down and analyze large quantities of data. The demand for these skills spans a variety of industries, including marketing, banking, government, sports, retail, and even healthcare."

It's generally a good idea for students to maintain a healthy level of curiosity or even skepticism in regard to their education. However, we find that when it comes to learning statistics, students' feelings may be characterized as fear or anxiety. We understand these feelings, and we hope the benefits of learning statistics will become clear to you and help overcome any concerns you have.

1.3 Introduction to the Stages of the Research Process

Much of scientific research involves asking questions. Throughout this book, we'll examine a broad range of questions regarding human attitudes and behavior that contemporary researchers have asked and attempted to answer. The following are research questions we'll address in this chapter to introduce the stages of the research process:

> Does appealing to one's fear of missing out (FOMO) affect the intention to invest in cryptocurrencies?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Do pro-cannabis messages on social media lead to cannabis use?

Do parents differ in their awareness of their children's online activities?

Do nonbinary college students have the same experiences as cis men or cis women?

How might you try to answer questions such as these? You could base your answers on your personal beliefs, or you could adopt the answers given to you by others. But rather than relying on subjective beliefs and feelings, researchers in many academic disciplines test their ideas using the scientific method. The **scientific method** is a method of investigation that uses the objective and systematic collection and analysis of empirical data to test theories and hypotheses.

At its simplest, this book will portray the scientific method as consisting of five main steps or phases:

developing a research hypothesis to be tested,

collecting data,

analyzing the data,

drawing a conclusion regarding the research hypothesis, and

communicating the findings of the study.

Accomplishing each of these five steps requires completing a number of tasks, as shown in Figure 1.1 . Because this sequence of steps serves as the model for the research process we'll discuss throughout this book, each step is briefly introduced in the following sections. It's important to understand that the research process depicted in Figure 1.1 represents an ideal way of doing research. The "real" way, as you may discover in your own efforts, is often anything but a smooth ride but rather is filled with false starts, dead ends, and wrong turns.

Developing a Research Hypothesis to Be Tested

The initial stage—and the first step—of the research process is to develop a research hypothesis to be tested. A **research hypothesis** may be defined as a statement regarding an expected or predicted relationship between variables. A **variable** is a property or characteristic of an object, event, or person that can take on different values. One example of a variable is "U.S. state," a variable with 50 possible values (Alabama, Arkansas, etc.).

Research hypotheses are usually developed through the completion of several tasks:

identifying a question or issue to be examined,

reviewing and evaluating relevant theories and research, and

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

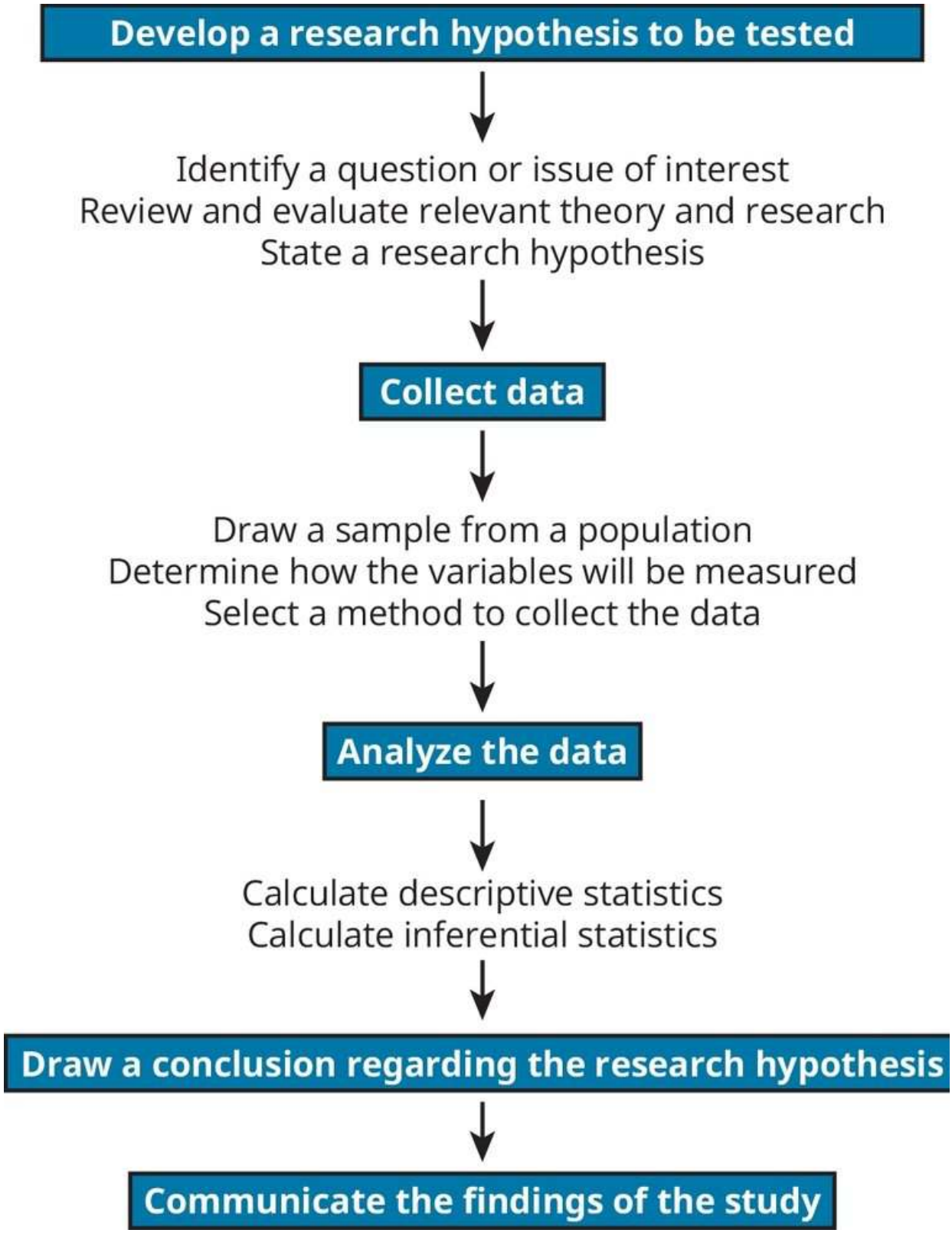
stating a research hypothesis.

Identifying a Question or Issue to Be Examined

Most research starts with a question posed by the researcher that often comes from the researcher's own ideas and daily observations. Although this may not seem terribly scientific, there's an advantage in using one's own experience as a starting point: People are generally much more motivated to explore issues that concern them personally. In teaching statistics, we advise students developing their own research projects to study something of interest to them. Conducting research can be tedious, difficult, and frustrating. At various points, you may ask yourself, "Why am I doing this?" Being able to provide a satisfactory answer to this question will help you overcome obstacles you encounter along the way.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

▶ Long Description: Figure 1.1 Steps in the Research Process Within the Scientific Method

Figure 1.1 Steps in the Research Process Within the Scientific Method

Reviewing and Evaluating Relevant Theories and Research

Beyond your own curiosity, research questions often arise from studying the theories, ideas, and research of others. A **theory** is a set of propositions used to describe or explain a phenomenon. The purpose of a theory is to summarize and explain specific facts using a logically consistent framework. Placing a question within a theoretical framework provides guidance and structure to research.

Reviewing and evaluating existing theories and research helps you decide whether it's worth the time and energy required to conduct a study. By seeing what others have done, you may conclude that the current way of thinking on a topic is incomplete or mistaken. By doing this review, you're able to ensure that the research you conduct adds to and improves upon an existing body of knowledge.

Stating a Research Hypothesis: Independent and Dependent Variables

Understanding and evaluating an existing literature not only helps develop a question of interest, but it may also lead to a predicted answer to that question. Within the scientific method, this answer is stated as a research hypothesis, defined earlier as a statement regarding an expected or predicted relationship between variables. Table 1.1 lists the research questions and research hypotheses for the studies listed at the beginning of this section. For example, the first research hypothesis states that "consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos."

One characteristic of research hypotheses such as those listed in Table 1.1 is that they identify the variables that are the focus of their research. As mentioned earlier, a variable is a property or characteristic with different values. Variables can be classified in several ways. In specifying a research hypothesis, researchers often speak in terms of "independent" and "dependent" variables. An **independent variable** may be defined as a variable manipulated or controlled by the researcher. A **dependent variable** , on the other hand, is a variable measured by the researcher that is expected to change or vary as a function of the independent variable. The independent variable is seen as the "cause" or "treatment," whereas the dependent variable is the "effect" or "outcome."

Researchers are interested in examining the effect of the independent variable on the dependent variable. Consider the first research hypothesis provided in Table 1.1 : "Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos." Here, the independent variable is the type of appeal consumers are exposed to: FOMO or non-FOMO. The dependent variable is the intention to invest in cryptos. In this study, the effect of the independent variable on the dependent variable is that

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

differences in the intention to invest in cryptos (the dependent variable) may “depend” upon which type of appeal (FOMO or non-FOMO) a consumer receives. Table 1.2 lists the independent and dependent variables for each of the research hypotheses in Table 1.1 .

In addition to identifying variables, a second characteristic of research hypotheses is that they specify the nature and direction of the relationship between variables. For example, the first research hypothesis in Table 1.1 states that “Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *higher* intention to invest in cryptos.” The word *higher* indicates the nature and direction of the relationship between the independent variable (type of appeal) and the dependent variable (intention to invest in cryptos). The direction of the relationship would not have been stated if the hypothesis had included the less specific phrase “will have a *different* intention,” which simply indicates that the two groups of consumers are not expected to be the same. Table 1.3 provides directional and nondirectional research hypotheses for the research studies from Table 1.1 .

Table 1.1 Examples of Research Questions and Hypotheses

Research Question	Research Hypothesis
Does appealing to one’s fear of missing out (FOMO) affect the intention to invest in cryptocurrencies? (Friederich et at., 2024)	Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.
Do pro-cannabis messages on social media lead to cannabis use? (Willoughby et al., 2024)	Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.
Do parents differ in their awareness of their children’s online activities? (Symons et al., 2017)	Mothers have more accurate knowledge of their children’s online activities than do fathers.
Do nonbinary college students have the same experiences as cis men or cis women? (Marx et al., 2024)	Nonbinary college students experience higher levels of victimization than do cis men or cis women.

The ability to form a directional research hypothesis is dependent on the state of the literature on the question of interest. If little or perhaps conflicting research has been conducted, researchers may not be able to form a directional hypothesis. For example, one study examining the relationship between exercise deprivation (not allowing people to exercise) and depression. reported, “We did not have a directional hypothesis, and when participants asked what we expected to find in this study, we replied: ‘We really are not sure since the results of earlier work on exercise deprivation are mixed’” (Mondin et al., 1996, p. 1200).

Table 1.2 Research Hypotheses and Their Independent and Dependent Variables

Research Hypothesis	Independent Variable	Dependent Variable
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher	Type of appeal	Intention to invest in cryptos

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Independent Variable	Dependent Variable
intention to invest in cryptos.		
Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.	Exposure to pro-cannabis messages	Intention to use cannabis
Mothers have more accurate knowledge of their children's online activities than do fathers.	Parent	Knowledge of children's online activities
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	Gender identity	Amount of victimization

Collecting Data

Once a research hypothesis has been formulated, researchers are ready to proceed to the second stage in the research process: collecting data. This step is seen as being composed of three tasks:

drawing a sample from a population,

determining how the variables will be measured, and

selecting a method by which to collect the data.

Table 1.3 Directional and Nondirectional Research Hypotheses

Directional Research Hypothesis	Nondirectional Research Hypothesis
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *higher* intention to invest in cryptos.	Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *different* intention to invest in cryptos.
Exposure to pro-cannabis messages on social media is related to *increased* intentions to use cannabis.	Exposure to pro-cannabis messages on social media is *related to* intentions to use cannabis.
Mothers have *more* accurate knowledge of their children's online activities than do fathers.	Mothers have *different* knowledge of their children's online activities than do fathers.
Nonbinary college students experience *higher* levels of victimization than do cis men or cis women.	Nonbinary college students experience *different* levels of victimization than do cis men or cis women.

Drawing a Sample From a Population

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The first step in collecting data is to identify the group of participants to which the research hypothesis applies. The group to which the results of a study may be applied or generalized is called a population. A **population** is the total number of possible units or elements that could potentially be included in a study. Although people often think the word "population" refers to everyone in the world, researchers typically define their populations more narrowly. For example, the population for the first research hypothesis in Table 1.1 could be "consumers," "U.S. consumers," "Georgia consumers," or "University of Georgia consumers." Researchers typically try to define their populations as broadly as possibly (e.g., "consumers" rather than "University of Georgia consumers") to maximize the applications or implications of their research.

Learning Check 1

Reviewing What You've Learned So Far



1. Review questions
    a. What are the five steps involved in the research process?
    b. Why is it useful to review and evaluate theories and research before conducting a study?
    c. What are the two main characteristics of research hypotheses?
    d. What is the difference between an independent variable and a dependent variable?
2. The following are several research hypotheses from published studies. For each research hypothesis, identify the independent and dependent variable.
    a. "College students will rate instructors who dress formally (i.e., business suit and tie) as having more expertise than instructors who dress casually (i.e., slacks and shirt)" (Sebastian & Bristow, 2008, p. 197).
    b. "It was expected tha. . . . greater amounts of television viewin. . . . would predict greate. . . . posttraumatic stress symptoms" (McLeish & Del Ben, 2008, p. 417).
    c. "Do borrowers who did not complete college have greater financial distress compared with . . . non-borrowers who did not complete college [or] borrowers who earned a degree?" (Jabbari et al., 2023, p. 6).
    d. "We expected that drivers would believe that texting while driving is more detrimental to their driving performance than browsing social media" (Hashash et al., 2019, p. 70).

It's extremely difficult to collect data from all members of a population. Imagine, for example, the time and money needed to collect information from every consumer in the United States! For this reason, researchers typically draw conclusions about populations based on information collected from a sample of the population. A **sample** is a subset or portion of a

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

population. Table 1.4 describes the samples used in the four studies introduced in Table 1.1 . As you can see from Table 1.4 , samples vary greatly in terms of their targeted population and size (the number of participants). The extent to which the results of a study can be generalized to the target population depends on the extent to which the sample is representative of the population.

Determining How Variables Will Be Measured: Levels of Measurement

The research hypotheses described in Table 1.1 involve variables such as type of appeal, intentions to use cannabis, and gender identity. To conduct a research study, the researcher must decide how to measure the variables stated in the research hypothesis. **Measurement** is the assignment of numbers or categories to objects or events according to rules.

For example, to measure the variable “height,” a researcher might use “number of inches from the ground in bare feet” as a form of measurement. To measure “success in college,” a student’s grade point average (GPA) may be obtained from school transcripts. “Self-esteem” might be measured by having people complete a questionnaire and be categorized as having either “low” or “high” self-esteem. As these examples demonstrate, different types of variables have different forms of measurement. Consequently, researchers have identified four distinct levels for measuring variables: nominal, ordinal, interval, and ratio.

The values of variables measured at the **nominal level of measurement** differ in category or type. The word *nominal* implies having to do with “names,” such that we use first and last names to distinguish between people. Gender identity is an example of a nominal variable, in that it consists of categories or types (such as cis male and cis female) rather than numeric values. In the first research hypothesis in Table 1.1 , the independent variable of “type of appeal” is measured at the nominal level with two categories: FOMO and non-FOMO.

Variables measured at the **ordinal level of measurement** have values that can be placed in an order relative to the other values. Rankings (such as finishing first, second, or third in a race), size (small, medium, large, or extra-large), and ratings (below average, average, or above average) are examples of ordinal variables. The values of ordinal variables indicate that one value represents relatively more or less of the variable than other values, but they don’t specify the precise difference between values. For example, although you can say that a “medium” drink is smaller than a “large” drink which is smaller than an “extra-large” drink, you can’t specify the exact difference in size between the three drinks.

Table 1.4 Research Hypotheses and Study Samples

Research Hypothesis	Study Sample
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.	“We recruited 150 participants from Amazon MTurk” (Friederich et at., 2024, p. 105).
Exposure to pro-cannabis messages on social media is	“Teens (N = 350) . . . needed to be between 13 and 17 years old and from Washington state” (Willoughby et

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Study Sample
related to increased intentions to use cannabis.	al., 2024, p. 185).
Mothers have more accurate knowledge of their children's online activities than do fathers.	"triadic data including reports from children aged 13 to 18, mothers and fathers . . . recruited in Flanders, the northern, Dutch speaking region of Belgium . . . a total of 357 valid triads was achieved" (Symons et al., 2017, p. 405).
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	"undergraduate college students aged 18–30 who were enrolled at 4-year institutions. This yielded a sample of 46,510 students" (Marx et al., 2024, p. 334).

The values of variables measured at the **interval level of measurement** are numbers equally spaced along a continuum. One example of an interval variable is the Fahrenheit scale of temperature. Here, a difference of 5 degrees has the same meaning anywhere along the scale; for example, the difference between 45°F and 50°F is the same as the difference between 65°F and 70°F. Many variables studied in the behavioral sciences (e.g., personality characteristics or attitudes) are considered to be measured at the interval level of measurement.

Variables at the **ratio level of measurement** are identical to interval variables, with one exception: Ratio scales possess what is known as a true zero point, for which the value of zero (0) represents the complete absence of the variable. Variables that describe a physical dimension (such as height, weight, distance, and time duration) typically have a true zero point. In the last research hypothesis in Table 1.1 , the variable "number of times victimized in last 12 months" is a ratio variable because it has a true zero point, where zero would indicate the complete absence of victimizations.

One advantage of ratio measurement versus interval is that ratio variables allow for a greater number of comparisons among values. Consider, for example, the value 6 for the ratio variable "inches." Not only is the difference between 4 and 6 inches the same as the difference between 6 and 8 inches (thereby involving addition and subtraction, which also can be used with interval variables), 6 inches is twice as long as 3 inches and half as long as 12 inches (involving multiplication and division). Multiplication and division comparisons cannot be made with interval variables. You cannot, for example, say that a temperature of 90°F is three times as much temperature as 30°F.

The boldfaced and italicized text in Table 1.5 illustrates how the four studies described in the earlier tables measured their independent and dependent variables. As you can see, the variables in these studies have been measured in a variety of ways at different levels of measurement.

Table 1.5 Research Hypotheses and Measurement of Variables

Research Hypothesis	Independent Variable	Dependent Variable

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Independent Variable	Dependent Variable
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.	Type of appeal **FOMO, non-FOMO (nominal)**	Intention to invest in cryptos **Likelihood of investing** **1-7 scale (1 = Highly unlikely to 7 = Highly likely) (ordinal)**
Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.	Exposure to pro-cannabis messages on social media **How often saw pro-cannabis messages (ratio)**	Intentions to use cannabis **Likelihood of using cannabis in the future (ratio)**
Mothers have more accurate knowledge of their children's online activities than do fathers.	Type of parent **Mother, Father (nominal)**	Accuracy of online activities **Accurate, somewhat accurate, inaccurate (nominal)**
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	Gender identity **Nonbinary, cis man, cis woman** **(nominal)**	Number of times victimized in last 12 months **(ratio)**

What difference does the level of measurement make?

A variable's level of measurement has important implications for researchers in that it influences how research hypotheses are stated and how data are analyzed. For example, imagine you study the variable "success in college." You could measure this variable at the *nominal* level of measurement by classifying students into one of two groups: successful or unsuccessful. Researchers typically use nominal variables when they are interested in studying differences between groups, such as, "Do successful and unsuccessful students differ in their study habits?"

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

You could instead measure success in college as an *ordinal* variable by ranking students from top to bottom, which enables a researcher to study relative differences with less concern for the precise size of these differences. For example, using ranked data, you could ask, “How similar are younger and older faculty members’ rankings of their students?”

To study success in college at the *interval* level, you could rate each student on a scale from 1 (low) to 100 (high). You could then use these ratings to address a research question such as, “Is there a relationship between faculty ratings of their students and students’ ratings of themselves?”

Finally, to measure success in college using the *ratio* level of measurement, you could use students’ salary after graduation (in dollars) as an indicator of success. Salary is a ratio variable because it contains a true zero point. In practice, ratio variables can be used to ask many of the same questions as those involving interval variables. For example, you could ask, “Is there a relationship between faculty members’ ratings of students and students’ salaries after graduation?”

As you see, how researchers measure their variables influences how they state their research questions and hypotheses. The level of measurement has another important implication: It helps determine the statistics researchers use to analyze their data. Certain statistics can be applied to variables measured at the interval or ratio levels of measurement, whereas others are appropriate for nominal or ordinal variables. As we move from chapter to chapter of this book, the level of measurement of the variables being analyzed will be noted.

Selecting a Method to Collect the Data: Experimental and Nonexperimental Research Methods

In addition to drawing a sample from a population and determining how the variables will be measured, the third step in collecting data is to determine the type of research method to use to collect the data on these variables. Research methods may be classified into two main types:

experimental research methods, and

nonexperimental research methods.

Experimental research methods.

Experimental research methods are methods designed to test causal relationships between variables—more specifically, whether changes in independent variables produce or cause changes in dependent variables. To make inferences about cause-effect relationships, researchers conducting an experiment must first eliminate other possible causes or explanations for changes in the dependent variable besides the independent variable. If it can be claimed that a variable other than the independent variable created the changes in the dependent variable, the research results are said to be confounded. A **confounding variable** is a variable related to an independent variable that provides an alternative explanation for the relationship between the independent and dependent variables.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

To illustrate the concept of confounding variables, consider the following question: Is there a relationship between a mother's ethnicity and the birth weight of her baby? Although research has shown differences in the birth weights of babies of different ethnicities, one study noted that some of this research did not eliminate the possibility that these differences may be partly due to differences between ethnic groups on factors associated with low birth weights, such as the mother's average age at the time of pregnancy and behaviors such as smoking and drinking (Ma, 2008). These other factors are considered confounding variables in that they provide alternative explanations for the relationship between ethnicity and babies' birth weights. Confounding variables make it difficult for researchers to draw firm conclusions regarding the relationship between the independent and dependent variable.

Learning Check 2

Reviewing What You've Learned So Far



1. Review questions
    a. What is the difference between a population and a sample?
    b. Within the research process, what is the relationship between populations and samples?
    c. What are the four levels of measurement? How do they differ?
2. For each of the following variables, name the scale of measurement (nominal, ordinal, interval, or ratio).
    a. Type of school attended (public, private)
    b. Probability of graduating college in 4 years (0% to 100%)
    c. Rating of television (good, better, best)
    d. Number of computers in home

Researchers minimize the influence of confounding variables in two main ways. First, researchers exercise experimental control by making the research setting (i.e., characteristics of the research participants, the instruments or measures administered, the methods used to collect data) the same for all participants. In the birth weight study, for example, the research design might control for the effect of parental smoking on birth weight by only including nonsmokers in the study sample, thereby excluding those who smoke.

Another way researchers exert control over the research setting is by including a condition known as a **control group** , a group of participants in an experiment not exposed to the independent variable. For example, if you conducted an experiment examining the effects of caffeine on one's health, one group (the experimental group) might be instructed to drink coffee while a second group (the control group) drinks decaffeinated coffee. By comparing the results of the two groups, researchers can then assess the impact of caffeinated coffee on the health of those who drink it.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Because researchers cannot possibly identify and control for all potential confounding variables, a second strategy used to minimize the influence of confounding variables is **random assignment** , which is assigning participants to each category of an independent variable in such a way that each participant has an equal chance of being assigned to each category. For example, a researcher could randomly assign a participant to the experimental or control condition by flipping a coin: heads for experimental, tails for control. The purpose of random assignment is to equalize or neutralize the effects of confounding variables by distributing them equally over the levels of the independent variable.

Nonexperimental (correlational) research methods.

Experimental research designs are one of the best tools researchers have for making causal inferences. However, the ability to make causal inferences requires a great deal of control that may not always be possible or desirable. For this reason, researchers often employ **nonexperimental research methods** , which are research methods designed to measure relationships between variables without having the ability to infer cause-effect relationships. These research methods are sometimes referred to as *correlational* research methods. It is important to understand that correlational research methods are different from correlational statistics we'll introduce in Chapter 13. Correlational research methods pertain to how data are *collected* ; correlational statistics pertain to how data are *analyzed* . Some of the most common types of nonexperimental research designs include quasi-experiments, survey research, observational research, and archival research.

Quasi-experimental research compares naturally formed or preexisting groups rather than randomly assigning participants to conditions. One example of this type of research was a study examining the effects of the COVID-19 pandemic on three groups of students: those living on campus, those living off campus but within commuting distance to school, and those living far away in another city, state, or country (Bisconer & McGill, 2023). This study is a quasi-experiment because the students were not randomly assigned to one of the three living conditions before collecting data.

Survey research methods obtain information directly from a group of people regarding their opinions, beliefs, or behavior. The goal of survey research, which can include using questionnaires and interviews, is to obtain information from a sample to represent or estimate the views of a larger population. One example of survey research examined college students' mental health by having them fill out a survey in which they reported their spirituality (the desire for truth, meaning, and significance of one's life) and their self-acceptance (knowledge and acceptance of one's self and limitations) (Nonis et al., 2024). Survey research is nonexperimental in that the variables being measured are not manipulated by the researcher.

Observational research is the systematic and objective observation of naturally occurring behavior or events. Observational research is often used to study phenomena that the researcher cannot or should not deliberately manipulate. For example, researchers studying interpersonal conflict among children would never force children to fight each other. One study resolved this issue by videotaping groups of adolescents discussing topics such as competition and peer pressure and then counting the number of times the adolescents teased or mocked each other (Connolly et al., 2015).

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Rather than observing or measuring behavior directly, **archival research** is the use of archives (records or documents of the activities of individuals, groups, or organizations) to examine research questions or hypotheses. One example of an archival research study studied the question of whether men are perceived to be more creative than women (Proudfoot et al., 2015). In this study, the 100 most viewed talks on TED.com were identified, and the researchers compared viewers' posted ratings of the talks given by men versus women.

Labeling variables in nonexperimental research.

One issue that often confuses students when learning about nonexperimental research is the practice of referring to the variables as the "independent variable" and the "dependent variable." For example, in the study described previously that involved biological sex and TED.com ratings, biological sex may be called the "independent variable." However, because biological sex isn't manipulated by the researcher, it would be inappropriate to say that biological sex "causes" differences in TED.com ratings. To avoid confusion, researchers sometimes refer to the independent variable and the dependent variable in a nonexperimental study as the **predictor variable** and **criterion variable**, respectively.

Analyzing the Data

Once data collection has been completed, the next step in the research process involves data analysis. This part of the research process addresses the primary topic of this book: statistics. Because this is the first chapter of this book, we won't describe analyzing data in detail. Instead, we introduce the notion that there are two main purposes of analyzing data: (1) to organize, summarize, and describe data that has been collected and (2) to test and draw conclusions about ideas and hypotheses. These two purposes are met by calculating two main types of statistics: descriptive statistics and inferential statistics.

Calculating Descriptive Statistics

Descriptive statistics are statistics used to summarize and describe a set of data for a variable. For example, you've probably heard of statistics used to describe or summarize aspects of our society, such as crime statistics and unemployment statistics. Using a research-related example, Caitlin Abar, a researcher at Brown University, conducted a study examining increases in students' alcohol-related behaviors after entering college (Abar, 2012). To measure students' level of alcohol use, she created a variable called "typical weekend drinking," which "was measured as the sum of drinks consumed on a typical Friday and Saturday within the past 30 days" (p. 22). One way to summarize students' responses to the typical weekend drinking variable (0, 1, 2, etc.) would be to calculate the mean, which is the mathematical average of a set of scores. The mean, described in Chapter 3, is an example of a descriptive statistic. The first part of this book describes a variety of descriptive statistics used by researchers to summarize data they've collected.

Calculating Inferential Statistics

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Besides summarizing and describing data, a second purpose of statistics is to test hypotheses. **Inferential statistics** are statistical procedures used to test hypotheses and draw conclusions from data collected during research studies. Using inferential statistics, researchers make inferences about what's believed to exist in populations based on data gathered from samples of the population.

As an example of an inferential statistic, let's return to the example of students' alcohol-related behaviors (Abar, 2012). This study was interested in seeing whether these behaviors were related to aspects of these students' relationships with their parents. One aspect was "alcohol communications," defined as "how often they discussed alcohol related topics with their parents" (p. 22). To test a hypothesis regarding the relationship between students' alcohol use and their alcohol communications, we could use an inferential statistic known as the Pearson correlation coefficient, which will be discussed in Chapter 13. The chapters in the last half of this book discuss a broad variety of inferential statistics, along with detailed instructions for analyzing and drawing conclusions from statistical data and presenting research findings.

Drawing a Conclusion Regarding the Research Hypothesis

Once statistical analyses have been completed, the next step is to interpret the results of the analyses as they relate to the study's research hypothesis. More specifically, do the findings of the analyses support or not support the research hypothesis? The word *support* in the previous sentence is very important. Students and researchers are sometimes tempted to say their findings "prove" their hypotheses are either "true" or "false." However, as we'll discuss in Chapter 5, because researchers typically don't collect data from the entire population, they can't know with 100% certainty whether their hypothesis has been proven to be true or false. A *huge* part of this book is learning how to draw appropriate conclusions from statistical analyses.

Communicating the Findings of the Study

Conducting research requires a variety of different skills: conceptual skills to develop research hypotheses, methodological skills to collect data, and mathematical skills to analyze these data. Another integral part of the research process is the communication skills needed to inform others about a study. Researchers must be not only skilled scientists but also effective writers.

Learning Check 3

Reviewing What You've Learned So Far



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1. Review questions
    a. What are the main differences between experimental and nonexperimental research methods?
    b. What are confounding variables? What do researchers do to minimize the effects of confounding variables?
    c. What are the main types of nonexperimental research methods?
    d. What are the relative strengths and weaknesses of experimental and nonexperimental research?
    e. What are the main purposes of descriptive and inferential statistics?

For many years, the American Psychological Association (APA), the professional association for psychologists, has provided researchers guidance on how to communicate the results of their research. In 1929, the APA published a seven-page article entitled, “Instructions in Regard to Preparation of Manuscript.” In contrast, the seventh edition of the *Publication Manual of the American Psychological Association* , published in 2020, consists of 428 pages! This increase in length highlights the challenge of communicating research effectively and appropriately.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1.4 Plan of the Book

The primary purpose of this chapter was to introduce statistics and place it within the larger research process. The remainder of this book discusses different statistics used by researchers to examine various questions of interest. Although it's critical for you to learn to correctly calculate statistics, it is equally important to understand the role of statistics within the research process and appreciate issues related to the use (and sometimes misuse) of statistics.

The first half of the book (Chapters 2 through 6) introduces conceptual and mathematical issues that are the foundation of statistical analyses. Chapters 2, 3, and 4 discuss how data may be examined, described, summarized, and presented in numeric, visual, and graphic form. Descriptive statistics are introduced and discussed in these chapters. Chapters 5 and 6 introduce characteristics of the inferential statistics used to test research hypotheses. The chapters in the second half of the book cover a number of different inferential statistics used with different types of variables.

In most cases, the chapters in this book share a uniform format and structure and are centered on the research process, using examples of actual published research studies that address questions with which you may be familiar. For each example, we'll clearly state the research hypothesis, briefly describe how the data in the study were collected, illustrate the calculation and interpretation of descriptive and inferential statistics, discuss the extent to which the statistical analyses support the research hypothesis, and show how to communicate one's research to a broader audience.

1.5 Looking Ahead

We began this chapter by defining statistics and providing reasons why learning about them may be of value to you. Next, we placed statistics within the stages of the larger research process, a process that will be the foundation of this textbook. As we've explained, the research process is centered on a research hypothesis, which is a predicted relationship between variables. Once a hypothesis has been stated, the next step is to collect data to be analyzed. However, before conducting statistical analyses on a set of data, there is preliminary work that must be completed. The next chapter will focus on methods used by researchers to examine data.

1.6 Summary

Statistics may be defined as a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data. As such, statistics not only is concerned about how data are analyzed but also recognizes the importance of understanding how data are collected and how the results of analyses are interpreted and communicated. There are a variety of reasons why students should learn statistics: Statistics are encountered in a wide variety of daily activities, students may be asked or required to read and interpret the results of statistical analyses in their courses, students may use statistics in conducting their own research, and statistics may help one's career.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Researchers conduct research using the **scientific method** of inquiry, a method of investigation that uses the objective and systematic collection and analysis of empirical data to test theories and hypotheses. The research process used within the scientific method of inquiry consists of five main steps: developing a research hypothesis to be tested, collecting data, analyzing the data, drawing a conclusion regarding the research hypothesis, and communicating the findings to the study.

A **research hypothesis** is a statement regarding an expected or predicted relationship between variables. Developing a research hypothesis involves identifying a question or issue to be examined, reviewing and evaluating relevant theories and research, and stating the research hypothesis to be tested in the study. A **theory** is a set of propositions used to describe or explain a phenomenon. A research hypothesis contains **variables** , which are properties or characteristics of some object, event, or person that can take on different values. More specifically, a research hypothesis states the nature and direction of a proposed relationship between an **independent variable** (a variable manipulated or controlled by the researcher) and a **dependent variable** (a variable measured by the researcher that is expected to change or vary as a function of the independent variable).

Collecting data involves drawing a sample from a population, determining how the variables will be measured, and selecting a method to collect the data. A **population** is the total number of possible units or elements that could potentially be included in a study; a **sample** is a subset or portion of a population. Variables can be measured at one of four levels of measurement: **nominal** (values differing in category or type), **ordinal** (values placed in an order relative to the other values), **interval** (values equally spaced along a numeric continuum), or **ratio** (values equally spaced along a numeric continuum with a true zero point).

There are two main types of methods used to collect data: **experimental research methods** , designed to test whether changes in independent variables produce or cause changes in dependent variables, and **nonexperimental (correlational) research methods** , designed to examine the relationship between variables without the ability to infer cause-effect relationships.

In experimental research, researchers are concerned about possible **confounding variables** , which are variables related to independent variables that provide an alternative explanation for the relationship between independent and dependent variables. To minimize the impact of confounding variables, a research study may exercise experimental control over various aspects of the situation, including the use of a **control group** , which is a group of participants not exposed to the independent variable the research is designed to test, or **random assignment** , which involves assigning participants to each category of an independent variable in such a way that each participant has an equal chance of being assigned to each category.

Examples of nonexperimental research methods are **survey research** , in which information is directly obtained from a group of people regarding their opinions, beliefs, or behavior; **observational research** , which involves the systematic and objective observation of naturally occurring events; and **archival research** , which uses archives (records or documents of the activities of individuals, groups, or organizations) to examine research questions or hypotheses. In nonexperimental research, the independent variable and

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

dependent variable may be referred to as the **predictor variable** and **criterion variable**, respectively.

Once data have been collected, the next step is to analyze them using two main types of statistics: **descriptive statistics**, which summarize and describe a set of data, and **inferential statistics**, which are statistical techniques used to test hypotheses and draw conclusions from data collected during research studies.

Once the statistical analyses have been completed, the results of the analyses are interpreted regarding whether they support or not support the study's research hypothesis. The final step in the research process is to communicate the study to a broader audience.

1.7 Important Terms

archival research

confounding variable

control group

criterion variable

dependent variable

descriptive statistics

experimental research methods

independent variable

inferential statistics

interval level of measurement

level of measurement (nominal, ordinal, interval, ratio)

measurement

nonexperimental research methods

observational research

population

predictor variable

quasi-experimental research

random assignment

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

research hypothesis

sample

scientific method

statistics

survey research

theory

variable

1.8 Exercises

1. Identify the independent and dependent variable for each of the following research questions and hypotheses from published articles.
    a. “We hypothesize that finance majors will embody narcissistic traits more than other business majors” (Wood et al., 2021, p. 924).
    b. “Men are expected to employ physical characteristics (particularly those that are directly related to sex) more often than women in selecting dating candidates” (Hetsroni, 2000, p. 91).
    c. “It is hypothesized that perceptions of emotional pain and support needs will be higher for the community of the Black victim than those of the officers’” (Kahn et al., 2024).
    d. “Gen Z students are more likely to believe that ‘Group projects bring out the worst in people’ than Millennials” (Schlee et al., 2020, p. 145).
    e. “We hypothesized to find . . . higher stress and anxiety post COVID in the Portuguese group in contrast with the British ” (Penalba-Sanchez et al., 2022, p. 322).
    f. “We hypothesized that consumers would think that a ‘sale price’ presentation would generate a greater monetary savings than an ‘everyday low price’ presentation” (Tom & Ruiz, 1997, p. 403).
2. The following are a number of research questions and hypotheses from actual published articles. For each hypothesis, identify the independent and dependent variable.
    a. “The use of color in a Yellow Pages advertisement will increase the perception of quality of the products for a particular business when compared with noncolor advertisements” (Lohse & Rosen, 2001, p. 75).
    b. “We hypothesized that parents who use more frequent corporal and verbal punishmen. . . . will report more problem behaviors in their children” (Brenner & Fox, 1998, p. 252).
    c. “It was hypothesized that adolescents with anorexia nervosa woul. . . . be more respectful when compared with peers with bulimia nervosa” (Pryor & Wiederman, 1998, p. 292).
    d. “The purpose of the present research was to assess brand name recognition as a function of humor in advertisements. . . . It was predicted that participants would recognize product brand names that had been presented with humorous

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

advertisements more often than brand names presented with nonhumorous advertisements" (Berg & Lippman, 2001, p. 197).

e. "The purpose of this study was to explore the relation between time spent in daycare and the quality of exploratory behaviors in 9-month-old infant. . . . it was hypothesized tha. . . . infants who spent greater amounts of time in center-based care would demonstrate more advanced exploratory behaviors than infants who did not spend as much time in center-based care" (Schuetze et al., 1999, p. 269).

f. "It was hypothesized that students would score higher on test items in which the narrative contains topics and elements that resonate with their daily experiences versus test items comprised of material that is unfamiliar" (Erdodi, 2012, p. 172).

3. Identify the independent and dependent variable for each of the following research questions and hypotheses from published articles.
    a. "It was hypothesized that the number of selfies posted . . . would be positively correlated with dimensions of narcissism" (Barry et al., 2017, p. 51).
    b. "We hypothesized that preschoolers' self-reported preferences for aggressive media . . . would be positively related to parent reports of children's aggressive behaviors" (Jamnik & DiLalla, 2018, p. 159).
    c. "We hypothesized that helicopter parenting would lead to higher levels of depression" (Darlow et al., 2017, p. 3).
    d. "We predicted that weight-related shame would be associated with greater binge eating" (Craven & Fekete, 2019, p. 45).
    e. "We hypothesize that men with strong gender identities will be more sensitive to manipulated threats to their masculinity compared with men with gender identities that are less central to their identity" (Harrison & Michelson, 2019, p. 65).
    f. "After exposure to fake news story with user comments identifying the news as fake, people are less likely to share the fake news than after exposure to the fake news story with user comments supportive of the content" (Colliander, 2019, p. 204).
4. Name the scale of measurement (nominal, ordinal, interval, ratio) for each of the following variables:
    a. The amount of time needed to react to a sound
    b. Gender
    c. Score on the Scholastic Aptitude Test (SAT)
    d. Political orientation (not at all conservative, conservative, very conservative)
    e. Political affiliation (Democrat, Republican, Independent)
5. Name the scale of measurement (nominal, ordinal, interval, ratio) for each of the following variables:
    a. One's age (in years)
    b. Size of soft drink (small, medium, large, extra large)
    c. Voting behavior (in favor vs. against)
    d. IQ score
    e. Parent (mother vs. father)
6. A faculty member wishes to assess the relationship between students' scores on the SAT and their performance in college.
    a. What is a possible research hypothesis in this situation?
    b. What are the independent and dependent variables?
    c. How could you measure the variable "performance in college" at each of the four levels of measurement?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

7. A researcher hypothesizes that drivers who use cellular phones will get into a greater number of traffic accidents than drivers who do not use these phones. For this example,
    a. What are the independent and dependent variables?
    b. How could you measure the dependent variable?
    c. How could you conduct this study using an experimental research method? How could you conduct this study using a nonexperimental method?

Answers to Learning Checks

Learning Check 1

1.

    a. IV: Instructor's dress; DV: Expertise

    b. IV: Amount of television viewing; DV: Posttraumatic stress symptoms

    c. IV: Student group; DV: Financial distress

    d. IV: Activity while driving; DV: Driving performance

Learning Check 2

2.

    a. Nominal

    b. Ratio

    c. Ordinal

    d. Ratio

Answers to Odd-Numbered Exercises

1.

    a. IV: Type of business major; DV: Narcissistic traits

    b. IV: Biological sex; DV: Employment of physical characteristics

    c. IV: Community; DV: Perceptions of emotional pain and support needs

    d. IV: Generation; DV: Beliefs about group projects

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

e. IV: Country; DV: Stress and anxiety

f. IV: Price presentation; DV: Perception of savings

3.

a. IV: Number of selfies posted; DV: Narcissism

b. IV: Preference for aggressive media; DV: Parent reports of children's aggressive behaviors

c. IV: Helicopter parenting; DV: Depression

d. IV: Weight-related shame; DV: Binge eating

e. IV: Strength of gender identity; DV: Sensitivity to threat to masculinity

f. IV: Type of user comments to news story; DV: Sharing of news story

5.

a. Ratio

b. Ordinal

c. Nominal

d. Interval

e. Nominal

7.

a. IV: Cellular phone use; DV: Number of traffic accidents

b. Example: Determine the number of traffic accidents each person has experienced in the past year.

c. Example of experimental: Use a driving simulation with the "driver" talking on a phone and measure the number of potential driving mistakes or accidents.

d. Example of nonexperimental: Take a survey of the number of accidents people have been in and whether or not they were using a phone during the accident.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CHAPTER 2 EXAMINING DATA: TABLES AND FIGURES

Learning Objectives

2.1 Describe the main purpose and the research hypothesis of the “lottery” study.

2.2 Describe different reasons for examining data.

2.3 Describe the steps used to create a frequency distribution table.

2.4 Describe the steps used to create a grouped frequency distribution table.

2.5 Explain the differences between bar charts, pie charts, histograms, and frequency polygons.

2.6 Describe the three aspects and types of distributions of data.

Chapter 1 described a process by which research may be conducted, which begins by stating a research hypothesis. The next step involves collecting data that will ultimately be used in statistical analyses designed to test the research hypothesis. However, researchers don’t conduct analyses until they’ve first examined the data to make sure appropriate conclusions can be drawn from the analyses. This chapter describes *why* researchers examine data and *how* data may be examined using tables, charts, and graphs. As with many of the chapters in this book, this discussion will be guided by the findings from a published research study.

2.1 An Example From the Research: Winning the Lottery

A friend of yours flips a coin and asks you to guess whether it’s “heads” or “tails.” You guess “tails,” but the answer is “heads.” Your friend flips the coin a second time, and again you predict “tails.” But again the answer is “heads.” On the third coin flip, you guess “tails” once more—but, once again, you’re wrong. As your friend prepares to flip the coin a fourth time, you tell yourself there’s little chance the coin could be “heads” again. But the reality is that the fourth coin flip is equally likely to be “heads” or “tails” because the outcome of the first three coin flips has absolutely no impact on the fourth flip. In other words, the outcome of a coin flip is a random event that can’t be controlled or predetermined. But do you really believe this?

The impact of random events can take on greater significance than the outcome of a coin flip. For example, government-run lotteries provide the opportunity to win huge amounts of money through the random selection of numbers. In 2001, Dr. Karen Hardoon, a researcher at McGill University in Canada, studied the thought processes people use when purchasing lottery tickets (Hardoon et al., 2001). As previous research had found gamblers believe they can control random events such as the rolling of dice or the spinning of a roulette wheel, Dr. Hardoon and her colleagues were interested in

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

determining “whether individuals with gambling problems perceive the purchase of lottery tickets in a similar manner as non-problem gamblers” (p. 752). On the basis of earlier findings, Dr. Hardoon hypothesized that problem gamblers are less likely to believe lottery outcomes are random than are nonproblem gamblers.

To test their research hypothesis, Dr. Hardoon and her associates located a group of gamblers and asked whether they had ever experienced problems as a result of their gambling. Each participant was classified into one of two groups: problem gamblers and nonproblem gamblers.

The participants in the study, which we’ll refer to as the lottery study, were shown four lottery tickets and asked, “If you were to buy a ticket to play in the lottery, which one would you select?” Each ticket contained six numbers selected from among the numbers 1 through 49; the four tickets were given a particular label. The *sequence* ticket had numbers in consecutive order (30-31-32-33-34-35); numbers on the *pattern* ticket increased by 5s (5-10-15-20-25-30); the *nonequilibrated* numbers (35-37-40-43-44-49) were clustered at the upper end of the 49 numbers, and the *random* ticket (7-8-23-34-36-42) shared none of the characteristics of the other three tickets. We’ll call the variable in this study “lottery ticket,” a variable with four categories: sequence, pattern, nonequilibrated, and random.

The researchers recorded which of the four tickets each participant chose. Table 2.1 lists the data for the lottery ticket variable for the nonproblem gamblers (the data for the problem gamblers are provided in the exercises at the end of this chapter). Looking at this table, we see that the nonproblem gamblers differed in their lottery ticket choices; however, it’s difficult to describe these differences given the unorganized nature of the collected data. Consequently, before conducting statistical analyses on their data, researchers typically organize and examine them. In the following sections, we will discuss why and how data are organized and examined.

2.2 Why Examine Data?

There are a variety of reasons why researchers examine data they’ve collected before analyzing the data. These reasons include the following:

To gain an initial sense of the data

To identify outliers

To evaluate research methodology

To determine whether data meet statistical criteria and assumptions

Each of these reasons is discussed in the following sections.

Table 2.1 The Lottery Ticket Choices of 22 Nonproblem Gamblers

Participant	Lottery Ticket
1	**Random**
2	**Random**
3	**Nonequilibrated**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Participant	Lottery Ticket
4	**Pattern**
5	**Random**
6	**Sequence**
7	**Random**
8	**Random**
9	**Pattern**
10	**Pattern**
11	**Random**
12	**Pattern**
13	**Random**
14	**Nonequilibrated**
15	**Random**
16	**Pattern**
17	**Nonequilibrated**
18	**Random**
19	**Random**
20	**Pattern**
21	**Random**
22	**Random**

Gaining an Initial Sense of the Data

Because researchers spend a great amount of time and energy designing their research studies, they're eager to analyze the data they've collected. Students taking courses in statistics, on the other hand, face a different challenge: They're given sets of data with which they have no prior experience. Although both researchers and students may be motivated to analyze their data as quickly as possible, it's important to first examine the data in order to gain an initial sense of them.

As an example, imagine you conduct a study regarding the fuel efficiency of automobiles and collect data on the miles per gallon (MPG) of different types of cars. You might expect the scores in your data to range from 10 to 50 MPG, with the majority between 25 and 35 MPG. Examining data helps researchers gain an initial sense of whether the data are in line with their expectations. Examining data also provides a way of confirming the results of later statistical analyses. We've found that students' errors in their calculations could have been avoided had they first looked at the set of data. For example, examining data will help you avoid making statements such as, "The average MPG in my sample was 127 MPG."

Identifying Outliers

Researchers also examine their data to identify **outliers** , defined as rare, extreme scores that lie outside of the range of the majority of scores in a set of data. Imagine, for example, that a group of 15 students is asked to take a 20-item test of mathematical calculations; each student's score is the number of test items answered correctly. In scoring the tests, a researcher finds that 14 students

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

correctly answered between 8 and 17 questions. However, one student provided the wrong answer to every question for a score of 0. In this case, the score of 0 would be considered an outlier.

If not identified, outliers may distort conclusions researchers draw about their sample and their data, particularly when the sample is relatively small. In the previous example, the single score of 0 might lead an observer to conclude the class as a whole performed more poorly on the test than they actually did.

Evaluating Research Methodology

Examining data is also useful in assessing the effectiveness of the methods used to collect the data. For the 20-item math test, for example, it might be reasonable to expect the majority of scores to fall in the middle of the 0 to 20 range. But what if every student answered at least 17 of the 20 questions correctly? You might conclude the test was too easy and that more difficult questions are needed to accurately assess students' math skills. By comparing the collected data for a variable with the range of possible values, researchers are able to evaluate and modify their measurement tools.

Determining Whether Data Meet Statistical Criteria and Assumptions

Researchers also examine their data to see whether they meet certain statistical criteria and assumptions. Many of the statistics in this book are based on assumptions regarding the shape and nature of a set of data that has been collected. For example, these procedures assume that in most data sets, the majority of scores will fall in the middle of the range of possible values, with a small amount of scores at the highest and lowest possible values (you may have heard this referred to as "the bell curve"). In the example of the 20-item math test, it might be assumed that most scores will be between 7 and 13, with smaller number of scores less than 7 or greater than 13. When assumptions such as this are not met, it's possible to draw inappropriate conclusions from statistical analyses that are conducted.

2.3 Examining Data Using Tables

One part of examining data that have been collected for a variable involves answering a series of questions:

> What are the values of the variable?
>
> How many participants in the sample have each value of the variable?
>
> What percentage of the sample has each value of the variable?

As it may be difficult to answer questions such as these just by looking at the data in Table 2.1 , researchers typically organize their data by creating a table known as a **frequency distribution table** , a table that summarizes the number and percentage of participants for the different values of a variable. This section discusses the steps involved in creating frequency distribution tables using these questions.

What Are the Values of the Variable?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The first step in organizing the data for a variable is to identify all of the possible values for the variable. Table 2.2 begins the construction of a frequency distribution table by identifying the four values for the lottery ticket variable: sequence, pattern, nonequilibrated, and random.

How Many Participants in the Sample Have Each Value of the Variable?

Once the values of the variable have been identified, the next step in organizing a set of data is to determine the **frequency** of each value, which is the number of participants in a sample corresponding to a value of a variable. These frequencies help indicate the nature and shape of the distribution of scores for a variable in a sample.

Table 2.2 Identifying the Values of the Lottery Ticket Variable

Number	Lottery Ticket
1	Sequence
2	Pattern
3	Nonequilibrated
4	Random

Table 2.3 Using the Tick Mark Method to Determine the Frequencies of the Lottery Ticket Variable

Lottery Ticket	Tick Marks	*f*
Sequence	|	1
Pattern	||||— |	6
Nonequilibrated	|||	3
Random	||||— ||||— ||	12
Total	||||— ||||— ||||—||||—||	22

One way to determine the frequencies for a variable is to sort the data into the different values and then count the number of participants having each value using a “tick mark method,” where a tick mark is placed next to a value of the variable each time that value appears in the sample. As shown in Table 2.3 , the tick marks are then counted to determine the frequency for each value. For example, the number of tick marks and therefore the frequency (f) for the sequence ticket is 1 because it was picked by one participant (Participant 6 in Table 2.1). Similarly, the tick marks for each of the other three types of tickets can be recorded and counted until all of the data in the sample are accounted for.

What Percentage of the Sample Has Each Value of the Variable?

Although it’s important to determine the frequency for each value of a variable, these frequencies by themselves are not always particularly informative. For example, how you interpret “There are 35 Democrats in my sample” is different when the sample consists of 40 voters rather than 400 voters.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Because the interpretation of frequencies depends on the size of the sample, it's useful to calculate the *percentage* of the sample having each value of a variable. These percentages can be calculated using the following formula:

$$\% = \frac{f}{total\,number\,of\,scores} * 100$$

where *f* is the frequency for a value of a variable.

Looking at Table 2.3 , 1 of the 22 nonproblem gamblers chose the sequence ticket. Therefore, the percentage of the sample choosing this ticket is

$$\begin{aligned} \% &= \frac{f}{total\,number\,of\,scores} * 100 \\ &= \frac{1}{22} * 100 = .05 * 100 \\ &= 5\% \end{aligned}$$

After calculating the percentage for each of the four tickets, Table 2.4 provides the frequency distribution table for the lottery ticket variable.

Once a frequency distribution table has been constructed for a variable, researchers can begin to draw conclusions about their sample. For the lottery ticket variable, the random ticket had the highest frequency and percentage (*f* = 12, 54%) compared to the pattern (*f* = 6, 27%), nonequilibrated (*f* = 3, 14%), and sequence (*f* = 1, 5%) tickets. As a result, the researchers in the lottery study made the following observation:

"The results of the present study indicated that, for the entire sample . . . the greatest percentage of tickets selected by the entire sampl. . . . were random tickets." (Hardoon et al., 2001, p. 760)

Table 2.4 Frequency Distribution Table for the Lottery Ticket Variable

Lottery Ticket	***f***	**%**
Sequence	1	**5%**
Pattern	6	**27%**
Nonequilibrated	3	**14%**
Random	12	**54%**
Total	22	**100%**

Learning Check 1

Reviewing What You've Learned So Far

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



1. Review questions
    a. What are some reasons for examining data before conducting statistical analyses?
    b. Why are outliers of concern to researchers?
    c. What questions do you address about a set of data for a variable in constructing frequency distribution tables?
    d. Why is it useful to calculate percentages in addition to frequencies for each value of a variable?
2. For each of the following situations, create a frequency distribution table.
    a. A college counselor asks a group of 116 seniors what their plans are after graduating from college; 71 say they are going to “work,” 34 are planning on going to “graduate school,” and 11 say they are “not sure” what their plans are.
    b. A high school instructor teaching a course finds that of the students she teaches, 69 are Freshmen, 18 are Sophomores, 12 are Juniors, and 9 are Seniors.
    c. A pollster stops 12 people and asks if they are either “in favor” or “against” a local proposition in Table 2.5 :

Table 2.5 Example of People’s Position

Person	Position
1	**In favor**
2	**In favor**
3	**Against**
4	**In favor**
5	**Against**
6	**Against**
7	**In favor**
8	**Against**
9	**In favor**
10	**Against**
11	**In favor**
12	**In favor**

    d. The company that makes M&Ms candy (www.mms.com) conducted a survey asking people which new color they would like to have added: purple, aqua, or pink. Table 2.6 are the votes of a hypothetical sample:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Table 2.6
Hypothetical Sample of New M&M Color

Person	Color
1	**Pink**
2	**Purple**
3	**Pink**
4	**Aqua**
5	**Purple**
6	**Purple**
7	**Purple**
8	**Aqua**
9	**Purple**
10	**Pink**
11	**Aqua**
12	**Pink**
13	**Pink**
14	**Purple**
15	**Aqua**
16	**Pink**
17	**Purple**
18	**Purple**
19	**Aqua**
20	**Purple**
21	**Purple**
22	**Pink**
23	**Pink**
24	**Purple**

2.4 Grouped Frequency Distribution Tables

The frequency distribution table in Table 2.4 was sufficient for presenting the frequencies for the lottery ticket variable. However, other situations may involve variables with a large number of possible values. Consider, for example, the variable grade point average (GPA). The values for GPA typically range from 0.00 to 4.00, with hundreds of possible values in between. Creating a frequency distribution table for GPA could result in an extremely large table, with many values having frequencies of zero.

To examine a variable with a large number of values, it is useful to construct a **grouped frequency distribution table** , a table that groups the values of a variable measured at the interval or ratio level

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

of measurement into a small number of intervals and then provides the frequency and percentage for each interval. To illustrate how grouped frequency distribution tables are constructed, Table 2.7 (a) provides a sample of 42 hypothetical GPAs. In Table 2.7 (b), each of these 42 GPAs has been placed into one of eight intervals that comprise a grouped frequency distribution table.

To illustrate how to read and interpret a grouped frequency distribution table, the lowest interval in Table 2.7 (b) consists of students with GPAs less than (<) 2.00. The frequency in this interval (*f* = 1) consists of one student: Student 29 (GPA = 1.96). The next interval consists of students with GPAs ranging from 2.00 to 2.29 and contains three students (4, 22, and 38). This grouping continues until all of the GPAs in the sample have been accounted for.

To create a grouped frequency distribution table, the real limits of each interval should be identified. The **real limits** of an interval are the values of the variable that are halfway between the top of one interval and the bottom of the next interval. For example, because of how values of GPA are rounded, the lowest value of GPA that would lead someone to be placed in the interval 2.30–2.59 is 2.295, which is halfway between the top of the 2.00–2.29 interval and the bottom of the 2.30–2.59 interval. The **real lower limit** is the smallest value of a variable that would be grouped into a particular interval. In this case, 2.295 represents the real lower limit for the 2.30–2.59 interval. On the other hand, the **real upper limit** is the largest value of a variable that would be grouped into a particular interval. For the 2.30–2.59 interval, the real upper limit is 2.594.

Once the grouped frequency table for the GPA example has been created (Table 2.7(b)), what preliminary conclusions might be drawn about these students? First, because the 2.90–3.19 interval has the greatest frequency (*f* = 10), you might conclude that the typical student in this sample has a GPA close to 3.00. Second, moving in both directions away from the 2.90–3.19 interval, the frequencies of the different intervals grow progressively smaller, with fewer and fewer students having either very high or very low GPAs. Table 2.8 provides a list of guidelines for creating grouped frequency distribution tables.

Cumulative Percentages

In addition to determining the percentage of the sample that has each individual value or interval for a variable, it is sometimes useful to combine these percentages. For the GPA example, you may be interested in knowing, “What percentage of the sample had GPAs less than 2.60?” To answer this question, the percentages in the “%” column of the frequency distribution table must be combined. This type of percentage is referred to as a **cumulative percentage** , defined as the percentage of a sample at or below a particular value of a variable.

Table 2.7a Grade Point Average (GPA), 42 Students: Individual GPAs

Student	GPA
1	**3.05**
2	**2.83**
3	**3.26**
4	**2.19**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Student	GPA
5	**3.52**
6	**3.34**
7	**3.02**
8	**2.97**
9	**3.71**
10	**2.50**
11	**2.77**
12	**3.10**
13	**3.06**
14	**3.83**
15	**2.39**
16	**3.16**
17	**3.92**
18	**2.37**
19	**3.65**
20	**3.70**
21	**3.00**
22	**2.26**
23	**2.65**
24	**3.40**
25	**2.87**
26	**2.92**
27	**2.89**
28	**3.41**
29	**1.96**
30	**3.77**
31	**3.20**
32	**2.86**
33	**2.94**
34	**2.71**
35	**3.43**
36	**3.95**
37	**3.32**
38	**2.28**
39	**3.08**
40	**3.25**
41	**2.40**
42	**3.22**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Table 2.7b Grade Point Average (GPA), 42 Students: Grouped Frequency Distribution Table

GPA	*f*	%
3.80+	3	7%
3.50−3.79	5	12%
3.20−3.49	9	21%
2.90−3.19	10	24%
2.60−2.89	7	17%
2.30−2.59	4	10%
2.00−2.29	3	7%
<2.00	1	2%
Total	42	100%

Table 2.8 Guidelines for Creating Grouped Frequency Distribution Tables

Guideline	Guideline Description
1. *Variables are grouped into approximately 10 intervals.*	The exact number of intervals depends on the data in each sample. For example, data that fall within a small range of values may be accurately represented with a relatively small number of intervals, whereas data extending across a wide range may require a larger number of intervals.
2. *The number of intervals should accurately represent the data.*	The intervals should represent the nature of the data as accurately as possible. For example, there should not be so large a number of intervals that many intervals have frequencies of zero (e.g., 3.00−3.05) or so few intervals that a large majority of the sample falls into only one or two intervals (e.g., 3.00−3.99).
3. *Intervals should be of equal size.*	To make the intervals comparable, they should be of equal size or width. For example, you would not want one interval of 2.01−2.50 (a width of .50) and another of 2.51−2.70 (a width of .20). One exception to this guideline is the lowest or highest interval (e.g., less than 2.00 or greater than 3.70), which is typically wider than the other intervals because it contains a relatively small frequency.
4. *Intervals should not overlap.*	The intervals should be distinct from each other, such that each score may be placed in only one interval. For example, intervals of 2.30− *2.60* and *2.60*− 2.90 are a problem because the GPA 2.60 is in both intervals leading to confusion regarding within which interval to place someone with a GPA of 2.60.

Table 2.9 includes the cumulative percentages for the GPA example. As an example of a cumulative percentage, the cumulative percentage for the 2.30–2.59 interval (19%) has been calculated by combining the percentages for three intervals: < 2.00, 2.00–2.29, and 2.30–2.59: 2% + 7% + 10% = 19%. As a result, we could conclude that 19% of this sample had GPAs less than 2.60. The accumulating of percentages continues until all (100%) of the scores in the sample are combined. As

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

shown in Table 2.9 , cumulative percentages are typically located in the final column of a frequency distribution table and labeled “Cum %.”

Table 2.9 Frequency Distribution Table With Cumulative Percentages for the GPA Variable

GPA	*f*	%	Cum %
3.80+	3	7%	**100%**
3.50−3.79	5	12%	**93%**
3.20−3.49	9	21%	**81%**
2.90−3.19	10	24%	**60%**
2.60−2.89	7	17%	**36%**
2.30−2.59	4	10%	**19%**
2.00−2.29	3	7%	**9%**
< 2.00	1	2%	**2%**
Total	42	100%	

2.5 Examining Data Using Figures

In addition to organizing data into frequency distribution tables, researchers also examine data visually using figures such as charts or graphs. There are four types of figures often used to visually portray data for variables: bar charts, pie charts, histograms, and frequency polygons. Which type of figure a researcher may use depends on the variable’s level of measurement (nominal, ordinal, interval, ratio). Data for variables measured at the nominal or ordinal level of measurement are typically displayed using bar charts or pie charts. The values of interval and ratio variables are graphed using histograms and frequency polygons. Each of these types of visual illustration is discussed in the following sections.

Displaying Nominal and Ordinal Variables: Bar Charts and Pie Charts

The values of variables measured at the *nominal* level of measurement differ in category or type; one example of a nominal variable is the lottery ticket variable discussed earlier, which consisted of four categories: sequence, pattern, nonequilibrated, and random. The values of variables measured at the *ordinal* level of measurement can be placed in an order relative to the other values; age (young, middle age, elderly) is one example of an ordinal variable.

One way to display nominal and ordinal variables is to use a bar chart. A **bar chart** is a figure that uses bars to represent the frequency or percentage of a sample corresponding to each value of a variable, with the different bars not touching each other. Figure 2.1 shows a bar chart for the data from the lottery ticket variable. In this figure, each bar corresponds to one of the four types of lottery tickets; the height of each bar corresponds to the frequency (*f*) of nonproblem gamblers who selected that type of ticket.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

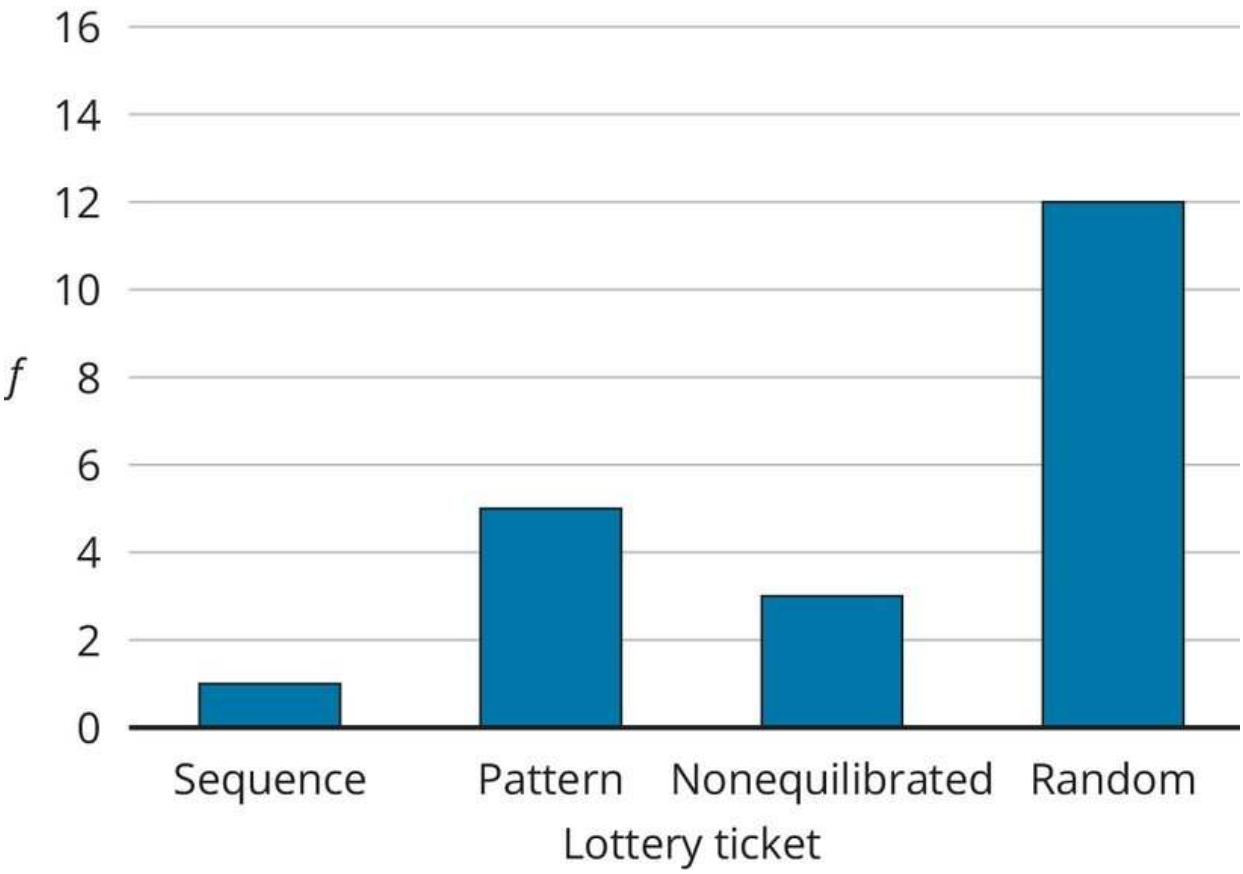


Figure 2.1 Bar Chart Showing the Frequencies of the Lottery Ticket Variable for the Nonproblem Gamblers

When creating a bar chart, the horizontal (X) axis contains the name of the variable and the values of the variable; the vertical (Y) axis contains the frequency (*f*) or percentage (%) for each value. Notice that the bars in the bar chart in Figure 2.1 do not touch each other; the gap between the bars indicates that the values of the variable represent distinct groups or categories that can't be placed along a numeric continuum.

A second type of figure used to display nominal or ordinal variables is a pie chart. A **pie chart** is a figure that uses a circle divided into proportions (slices) that represent the percentage of the sample for each value of a variable. Pie charts are created by dividing the 360 degrees of a circle into proportions corresponding to the percentage of each value.

Figure 2.2 shows a pie chart for the data for the lottery ticket variable. To illustrate how pie charts are constructed, the frequency distribution table in Table 2.4 shows that 5% of nonproblem gamblers chose the sequence ticket. Given that a pie chart consists of 360 degrees, 5% of the 360 degrees are needed to represent the sequence ticket. As 5% * 360 = 18, 18 degrees of the pie is assigned to the sequence ticket. Similarly, 97 degrees of the pie chart (27% of 360) are used to represent the pattern ticket. This process continues until all 360 degrees of the pie chart are accounted for.

The decision whether to use a bar chart or pie chart to illustrate a variable may depend on the number of values for the variable. For example, a pie chart for a variable that has 12 values may

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

result in many of the 12 “slices” being very thin and small, making them hard to see, meaning that a bar chart may be a better choice to illustrate this variable.

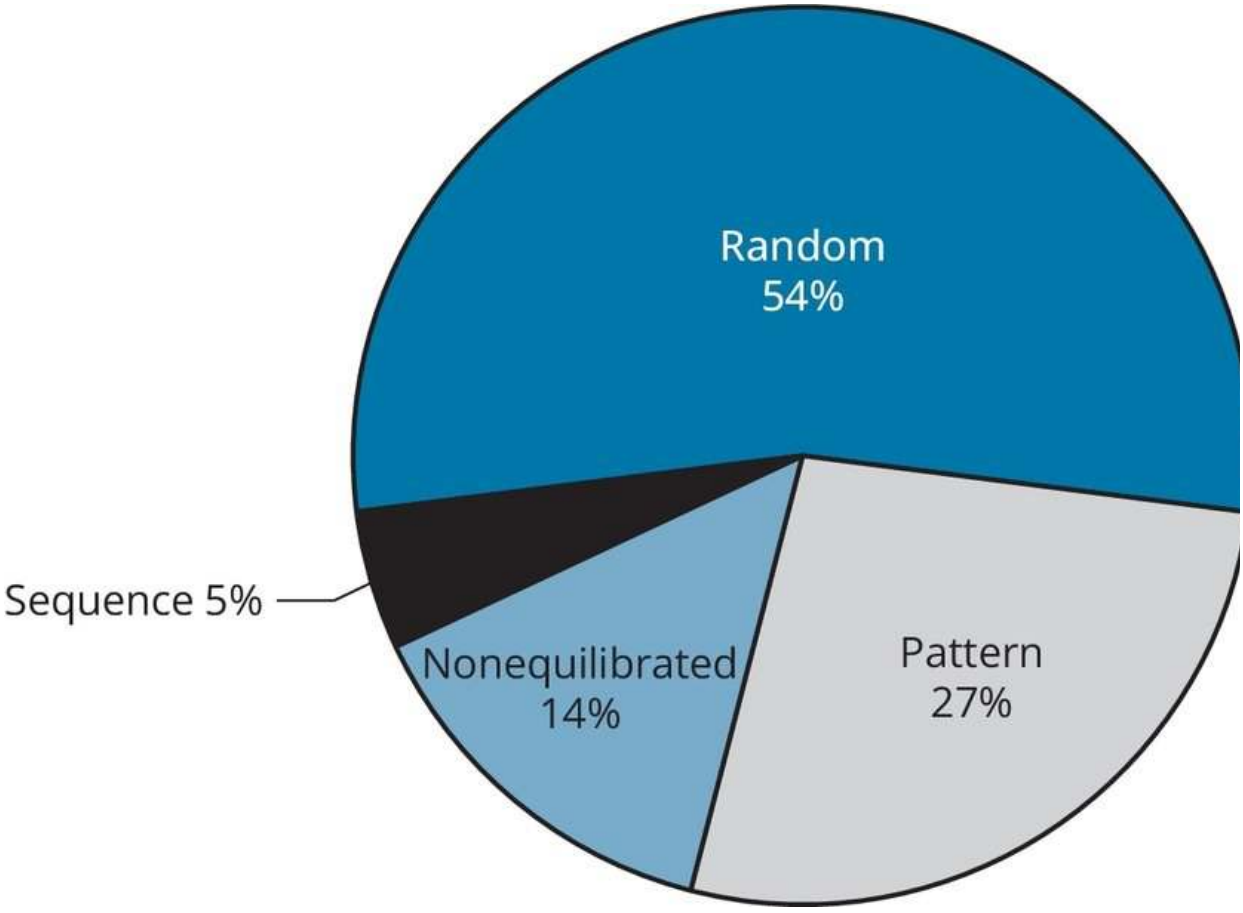


Figure 2.2 Pie Chart Showing the Percentages of the Lottery Ticket Variable for the Non−Problem Gamblers

Displaying Interval and Ratio Variables: Histograms and Frequency Polygons

Interval and ratio variables have values equally spaced along a numeric continuum and are identical to each other with one exception: Ratio scales possess a true zero point, for which the value of zero (0) represents the complete absence of the variable. The values of interval and ratio variables may be illustrated using histograms and frequency polygons, the choice of which depends on the number of values for the variable.

A **histogram** is a figure in which bars are used to represent the frequency of each value of a variable, with the different bars touching each other. The name and values of the variable are located on the X-axis, and the frequency or percentage for each value is on the Y-axis. Histograms are similar to bar charts, with one important difference: The bars in a histogram touch each other, indicating that the values of the variable are connected to each other along a numeric continuum.

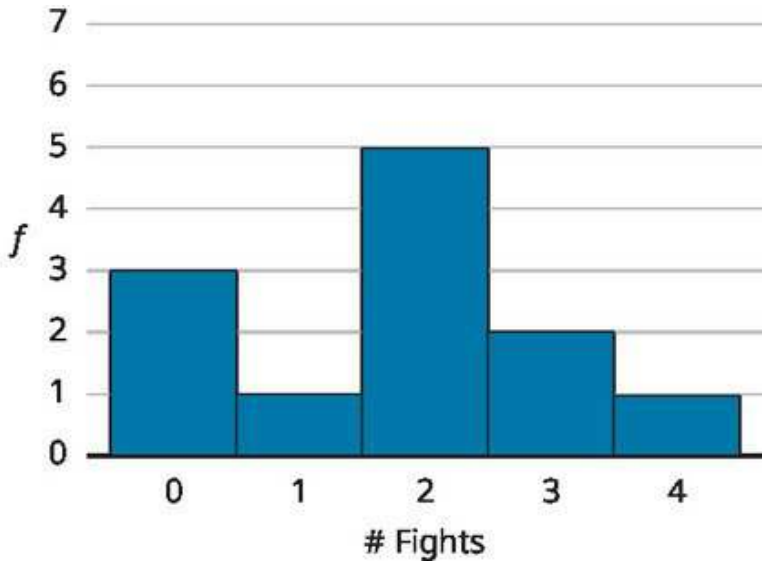
join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

As an example, imagine a teacher is interested in studying aggressive behavior in children and measures this variable by counting the number of fights each of her students is involved in over a 3-day period. “Number of fights” is a ratio variable rather than an interval variable because the value of 0 represents the complete absence of fights. Figure 2.3 provides a frequency distribution table and histogram for this variable.

A histogram is used when an interval or ratio variable consists of a relatively small number of values. But what if, for example, the number of fights could be anywhere from 0 to 20 rather than 0 to 5? In order to have the histogram fit on one page, the 21 bars would need to be very thin and crowded together, making the histogram confusing and difficult to read. In situations such as this, the values of the variable may be more clearly illustrated using a figure known as a **frequency polygon**, which is a line graph that uses data points to represent the frequency of each value of a variable, with straight lines connecting the data points. In a frequency polygon, the X-axis represents the name and values of the variable, while the Y-axis represents the frequency for each value.

# Fights	*f*
4	1
3	2
2	5
1	1
0	3
Total	12



▶ Long Description: Figure 2.3 Frequency Distribution Table and Histogram for the Number of Fights Variable

Figure 2.3 Frequency Distribution Table and Histogram for the Number of Fights Variable

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

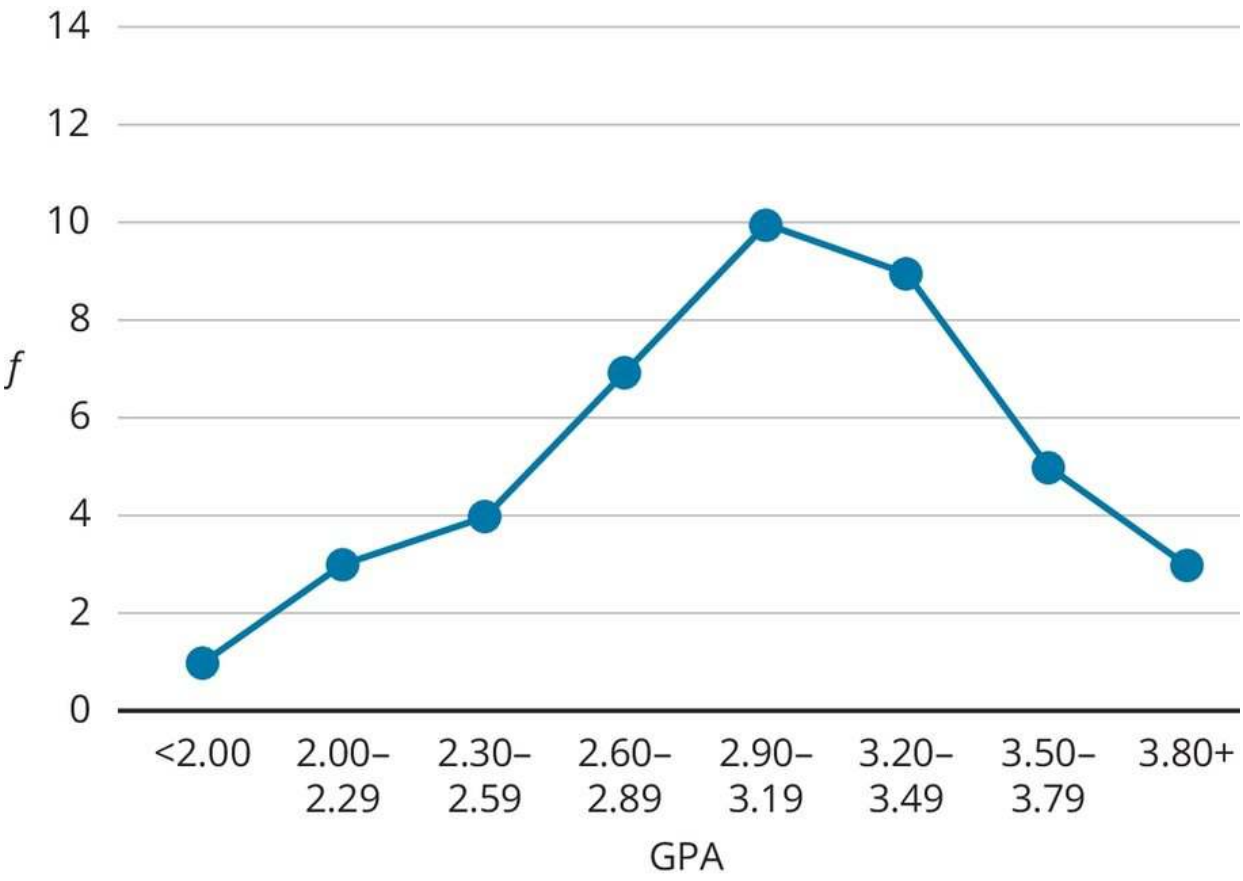


Figure 2.4 Frequency Polygon Showing the Frequency of Students in Each Interval of the Grade Point Average (GPA) Variable

Figure 2.4 shows a frequency polygon based on the grouped frequency distribution table of the earlier example of students' GPA (Table 2.7(b)). In this frequency polygon, the data points representing the frequencies of the eight GPA intervals are connected using straight, sharp lines. These lines indicate that the values of the variable can be connected together along a numeric continuum.

Although a frequency polygon like the one in Figure 2.4 may look jagged as it moves from one data point to the next, it provides a general sense of the shape of the distribution of scores for a variable. For example, looking at Figure 2.4 , you might conclude that the majority of GPAs in this sample are in the middle of the 1.00 to 4.00 range, with relatively low frequencies of students with either low or high GPAs.

Learning Check 2

Reviewing What You've Learned So Far

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



1. Review questions
    a. What is the difference between a frequency distribution table and a grouped frequency distribution table?
    b. For what levels of measurement would you create a pie chart, bar chart, histogram, or a frequency polygon?
    c. What is the difference between a bar chart and a histogram?
    d. Why would it be inappropriate to create a histogram or frequency polygon for a nominal or ordinal variable?
    e. Under what circumstances might you use a frequency polygon rather than a histogram to graph an interval or ratio variable?
2. For each of the following variables, identify whether you could use a bar chart, pie chart, histogram, and/or a frequency polygon to visually display the data.
    a. Miles per gallon (MPG)
    b. Living situation (on-campus, off-campus)
    c. Number of brothers and sisters
    d. Marital status (single, married, divorced)

Drawing Inappropriate Conclusions From Figures

A well-constructed figure can greatly facilitate the understanding of complex information. However, a poorly designed figure may not only be confusing, it can lead to inappropriate or misleading conclusions. Box 2.1 provides guidelines for constructing figures based on recommendations from the American Psychological Association (APA).

A simple example may demonstrate how altering the features of a figure can lead to different interpretations of the same set of data. Imagine a pollster assesses how voters with different political affiliations plan to vote on a ballot measure. She asks a sample of Democrats, Republicans, and Independents the question, “Do you plan on voting in favor of Measure A?” In examining her data, the pollster finds 56% of Democrats, 47% of Republicans, and 52% of Independents intend to vote “yes” on the measure.

Figure 2.5 graphs these three percentages using two bar charts; bar charts are used to display the data because the variable of interest (political affiliation) is a nominal variable (Democrat, Republican, or Independent). The bars don’t touch each other because the three political affiliations are distinct from each other and can’t be placed along a numeric continuum.

Although the two bar charts are similar, there is one important difference in how they are constructed. Following the guidelines in Box 2.1, the bar chart in Figure 2.5 (a) starts the Y-axis with the value 0% and ends with 100%. However, the bar chart in Figure 2.5 (b) starts with the value 45% and ends with 65%. By limiting the range of values of the percentages, the bar chart in Figure 2.5 (b) exaggerates differences in the heights of the three bars. As a result, someone looking at Figure 2.5 (b) may interpret the differences between the three political affiliations as being larger than someone looking at Figure 2.5 (a).

The saying “a picture is worth a thousand words” is particularly relevant for the presentation of statistical data. Because figures can provide a vivid representation of data, it’s the responsibility of researchers to portray their data in an accurate manner and to avoid creating images that may mislead or confuse.

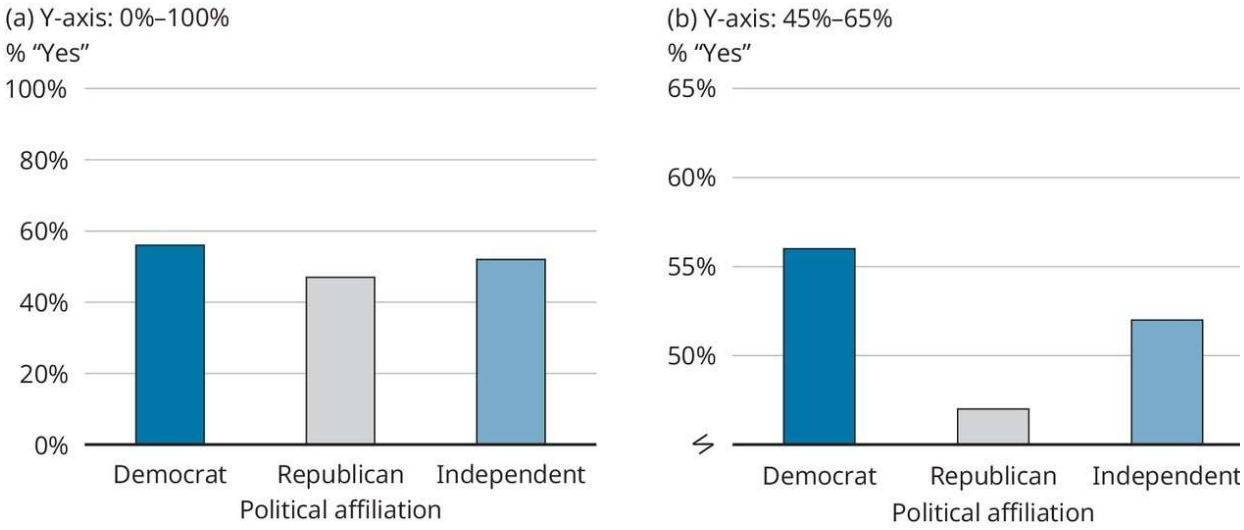
join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Box 2.1 Some APA Guidelines for Creating Figures

1. The figure should be formatted in a portrait rather than landscape page orientation (the reader should not turn the page sideways to read the table).
2. The figure is numbered using Arabic numbers (Figure 1, not Figure *I* or Figure *A*).
3. The figure includes a title that describes the contents of the figure. For example, it is not enough to name a figure simply " Figure 2.1 " or " Figure 2.1 . Bar chart."
4. Figures have a rectangular shape, with the height of the vertical (Y) axis two thirds to three fourths the length of the horizontal (X) axis.
5. The horizontal (X) axis contains the values of the variable; the vertical (Y) axis contains the frequency (*f*) or percentage (%) of the sample having each value of the variable.
6. The *X* - and *Y* -axes include descriptive labels written in both upper- and lowercase letters (as opposed to all capital letters).
7. The labels for the *X* - and *Y* -axes are short enough to fit on one line along the axes.
8. The label of the Y-axis is often placed parallel to the Y-axis (the letters face sideways rather than vertically) but can sometimes be seen on top of the y-axis for accessibility.
9. The Y-axis starts with the value zero (0) and is divided into equally spaced values or intervals.
10. The Y-axis is divided into 6 to 10 values (just enough values to accommodate bar length). The bars should neither be flattened along the bottom of the Y-axis nor bumped up against the top of the Y-axis.

Political affiliation	% "Yes"
Democrat	56%
Republican	47%
Independent	52%



▶ Long Description: Figure 2.5 Different Displays of the Same Data

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Figure 2.5 Different Displays of the Same Data

2.6 Examining Data: Describing Distributions

We've discussed how researchers construct tables and figures to organize their data. Looking at a table or figure enables researchers to describe the distribution of values or scores for a variable. Describing a distribution involves addressing questions such as, "Which values of the variable occur most often?," "How do the frequencies of the values of the variable change in relation to the most common values?," and "What is the amount of differences among the different values?"

Questions such as these pertain to three aspects of distributions: modality, symmetry, and variability. These three aspects of distributions are introduced in the following sections; they'll also be discussed in greater detail in the next two chapters of this book.

Modality

The first feature of a distribution of data for a variable is its modality. The **modality** of a distribution refers to the value or values of the variable that have the highest frequency or occur most often in a set of data. Using the lottery study as an example, identifying the modality of the lottery ticket variable involves answering the question, "Which of the four lottery tickets was chosen the most often?"

The modality of a distribution can be identified by looking at a frequency distribution table; the modality is the value(s) of the variable with the highest frequency (*f*) or percentage (%). The modality may also be determined by examining a figure and locating the tallest bar in a bar chart or histogram, the highest point in a frequency polygon, or the biggest "slice" in a pie chart. For the lottery ticket variable, the modality is the random ticket.

In identifying the modality of a distribution, it may be that one value or a limited number of values adjacent to one another clearly have the highest frequency. In such cases, the distribution is referred to as a unimodal distribution. A **unimodal distribution** is a distribution where one value of a variable occurs with the greatest frequency. Looking at the frequency distribution table (Table 2.4) or figure (Figure 2.1) for the lottery ticket variable, this distribution would be considered unimodal as one value, the random ticket, clearly occurred more often than the other three ticket choices.

In contrast to a unimodal distribution, it's possible that two distinct values may both have relatively high frequencies; if so, the distribution is referred to as a **bimodal distribution** , which is a distribution where two values clearly distinct from each other have the greatest frequency. A bimodal distribution sometimes resembles the two humps on a camel's back. As an example of a bimodal distribution, in April 2024 the Pew Research Center (www.pewresearch.org) asked a sample of 7,000 voters whether they felt Donald Trump had the mental fitness needed to be president, with five responses ranging from "extremely" to "not at all." The majority of voters chose one of the two extreme choices, either "extremely" (23%) or "not at all" (35%), with only 14% choosing the middle response of "somewhat."

And, in addition to unimodal and bimodal distributions, it's possible for a distribution to have more than two values with relatively equally high frequencies; this type of distribution is known as a **multimodal distribution** —a distribution where more than two values have the greatest frequency. In examining people's attitudes toward gun control, in May 2023, the news network CNN asked 1,227 people the question, "Do you think allowing gun owners to carry their guns in public places makes those places safer, less safe, or makes no difference in how safe a public place is?" They found people differed greatly in their responses, with 32% saying "makes public places safer," 36% saying

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

"makes public places less safe," and 32% saying "makes no difference," Figure 2.6 presents an example of a unimodal, bimodal, and multimodal distribution.

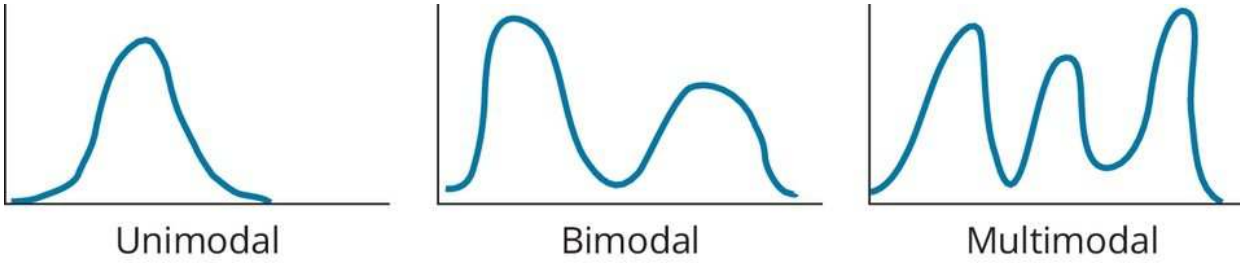


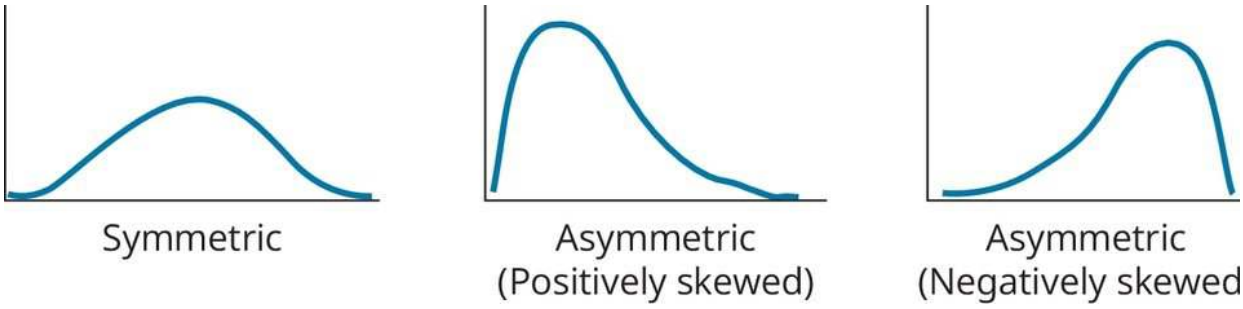
Figure 2.6 Distributions With Different Modalities (Unimodal, Bimodal, and Multimodal)

Symmetry

The second feature of a distribution of data is its symmetry. The **symmetry** of a distribution refers to how the frequencies of values of a variable change in relation to the most common or frequently occurring values. Describing the symmetry of a distribution means answering the question, "What is the shape of the distribution?"

The symmetry of a distribution can be determined by looking at the frequencies in a frequency distribution table or the bars or points in a figure. By doing so, a distribution may be described as being either symmetric or asymmetric. A **symmetric distribution** is one in which the frequencies change in a similar manner moving away in both directions from the most frequently occurring values. An example of a symmetric distribution is illustrated in Figure 2.7 . One way to determine whether a distribution is symmetric is to draw the distribution on a piece of paper and then fold the distribution vertically down the middle so that the two halves are on top of each other. A distribution is symmetric when the two halves are similar in shape such that they resemble mirror images of each other.

However, a distribution may also be asymmetric. An **asymmetric distribution** (also known as a *skewed* distribution) is one in which the frequencies change in a different manner moving away in both directions from the most frequently occurring values. In asymmetric distributions, the highest frequencies are located at one end of the distribution rather than in the middle, such that if you fold an asymmetric distribution down the middle, the shapes of the two halves do not look the same.



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Figure 2.7 Distributions With Different Symmetries (Symmetric, Positively Skewed, and Negatively Skewed)

In an asymmetric distribution, the frequencies take on the shape of a long tail as the values move toward the end of the distribution; scores in the tail of the distribution are often considered to be outliers. For example, in the 2022 Current Population Survey, the U.S. Census Bureau (www.census.gov) estimated that most Americans (63%) had annual incomes less than $50,000. However, a very small percentage (3%) were greater than $200,000; these higher incomes would be considered outliers.

Depending on the location of its tail, asymmetric distributions are referred to either as positively skewed or negatively skewed. A **positively skewed distribution** is one in which the higher frequencies are at the lower end of the distribution, with the tail on the upper (right) end of the distribution; a **negatively skewed distribution** is one in which the higher frequencies are at the upper end of the distribution, with the tail on the lower (left) end of the distribution. Figure 2.7 provides an illustration of a symmetric distribution and two asymmetric distributions (one positively skewed and one negatively skewed). Annual income, illustrated by the U.S. Census Bureau example mentioned earlier, is an example of a positively skewed distribution.

Variability

The third way in which distributions may be described is in terms of the amount or nature of differences among the different values for a variable. The **variability** of a distribution refers to the amount of differences in a distribution of data for a variable. The variability in a distribution is described by answering the question, “To what extent are the scores in the distribution different from each other?”

The amount of variability in a distribution can be estimated by looking at a frequency distribution table or figure. Imagine, for example, an instructor gives an 11-item quiz to her class and finds almost all of the students answer six, seven, or eight of the items correctly. In this situation, the distribution would resemble the shape of a tall mountain: thin and tall, with many of the scores residing in a few values of the variable. A distribution with a small amount of variability such as this one is referred to as a **peaked distribution** , defined as a distribution where much of the data are in a small number of values of a variable.

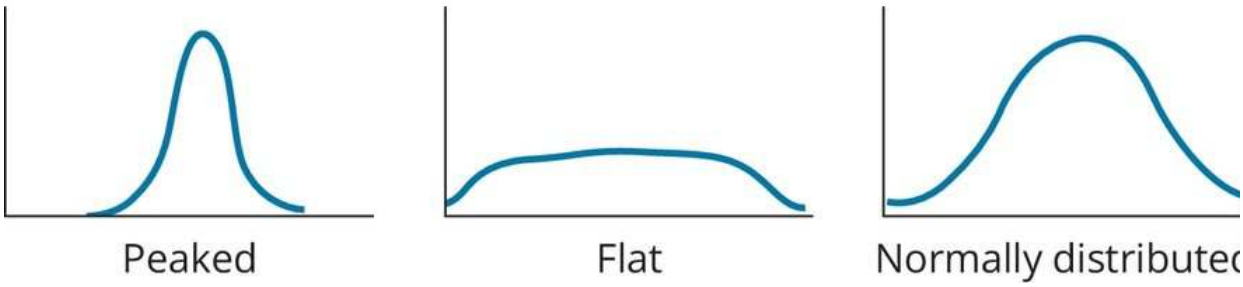


Figure 2.8 Distributions With Different Variability (Peaked, Flat, and Normally Distributed)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

On the other hand, what would the distribution look like if she had found her students' scores were distributed relatively evenly across the entire range of possible scores? In this situation, the distribution would not be tall and thin like a mountain but rather would be low and wide like a plateau. This type of distribution is referred to as a **flat distribution** , which is a distribution where the data are spread evenly across the values of a variable. The CNN gun control example described earlier is an example of a flat distribution.

In discussing modality, symmetry, and variability, distributions may have more than one mode, may be asymmetric (skewed), and may be peaked or flat. But what if a distribution has none of these qualities? A distribution that is unimodal, symmetric, and neither peaked nor flat may be described as "bell shaped." If you look at a bell, you'll notice that it is the highest in the middle, and as you move away from the top of a bell in both the left and right directions, the shape of the bell slopes downward in the same manner. A bell-shaped distribution for a variable is given a special name: a normally distributed variable.

A **normally distributed variable** is a distribution for a variable that is considered unimodal, symmetric, and neither peaked nor flat. Note that the word *normal* is a statistical concept and not a value judgment. In other words, asymmetric, peaked, or flat distributions are not considered "abnormal." The concept of normal distributions will be discussed in greater detail in Chapter 5. Figure 2.8 provides an illustration of three distributions with different amounts of variability: peaked, flat, and normally distributed.

Learning Check 3

Reviewing What You've Learned So Far



1. Review questions
    a. What are the three aspects used to describe distributions of variables?
    b. What is the difference between unimodal, bimodal, and multimodal distributions?
    c. What is the difference between symmetric and asymmetric (skewed) distributions?
    d. What is the difference between positively and negatively skewed distributions?
    e. What is the difference between peaked and flat distributions?
    f. What are the characteristics of a normally distributed variable?
2. For each of the following sets of data, determine whether the distribution of scores is unimodal, bimodal, or multimodal.
    a. 4, 2, 11, 6, 10, 2, 6, 13, 2, 10, 6, 10, 1
    b. 15, 17, 13, 18, 17, 14, 20, 18, 17, 14, 17
    c. 20, 10, 50, 10, 80, 30, 70, 10, 60, 10, 80, 0, 30, 80, 10, 100, 80
3. For each of the following sets of data, determine whether the distribution of scores is symmetric or asymmetric (skewed).
    a. 4, 5, 1, 4, 2, 4, 3, 5, 3, 4, 4, 3
    b. 220, 230, 250, 230, 200, 240, 210, 230, 270, 220, 240, 230, 240, 230
    c. 3, 1, 9, 2, 3, 6, 2, 7, 3, 5, 2, 6, 3, 2, 2
4. For each of the following sets of data, determine whether the distribution of scores is peaked or flat.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

a. 16, 10, 19, 12, 13, 15, 14, 17, 15, 20, 11, 13, 19, 16, 15
b. 12, 15, 12, 11, 13, 11, 12, 12, 14, 10, 12, 11, 16, 12, 13, 12, 11
c. 14, 12, 16, 14, 13, 15, 12, 14, 14, 15, 14, 14, 13, 14

2.7 Looking Ahead

This chapter has discussed why and how researchers examine data. Examining data by creating tables and figures requires a number of different skills, including organizational skills (to prepare a set of data for statistical analysis), communication skills (to present data in tables or figures), and conceptual skills (to describe a distribution). The next two chapters focus on mathematical skills related to the statistical analysis of data, describing how researchers describe distributions numerically rather than visually. Given that this is a statistics textbook, we will focus on calculating statistics to describe and analyze data. However, and we cannot say this strongly enough, it is *extremely* important to examine a set of data before conducting statistical analyses as this will prevent errors both in your calculations of statistics as well as your interpretation of the results of these calculations.

2.8 Summary

Before conducting statistical analyses, researchers typically examine their data to gain an initial sense of the data, detect data coding or data entry errors, identify ***outliers*** (rare, extreme scores that lie outside of the range of the majority of scores in a set of data), evaluate research methodology, and determine whether the data meet statistical criteria and assumptions.

One common way of examining variables involves creating tables known as ***frequency distribution tables*** (tables that summarize the number and percentage of participants that express the different values for a variable) and ***grouped frequency distribution tables*** (tables that group the values of a variable measured at the interval or ratio level of measurement into a small number of intervals and then provide the frequency and percentage within each interval). To create a grouped frequency table, the ***real limits*** of each interval (the values of the variable that fall halfway between the top of one interval and the bottom of the next interval) need to be identified.

Data for a variable may also be examined by creating figures such as bar charts, pie charts, histograms, and frequency polygons. ***Bar charts*** and ***pie charts*** are used for variables measured at the nominal or ordinal level of measurement; a bar chart uses bars to represent the frequency or percentage of a sample corresponding to each value of a variable, with the different bars not touching each other; a pie chart is a circle divided into proportions that represent the percentage of the sample corresponding to each value of a variable. ***Histograms*** and ***frequency polygons*** are used for variables measured at the interval or ratio level of measurement; a histogram is a figure in which bars are used to represent the frequency of each value of a variable, with the different bars touching each other; a frequency polygon is a line graph that uses data points to represent the frequency of each value of a variable, with lines connecting the data points.

From examining their data, researchers can describe the distribution of values or scores for a variable. There are three main aspects of distributions: modality, symmetry, and variability.

The ***modality*** of a distribution is the value(s) of the variable that have the highest frequencies or occur most often in a set of data. ***Unimodal distributions*** are distributions where one value occurs with the greatest frequency, ***bimodal distributions*** have two values with the greatest frequency, and ***multimodal distributions*** have more than two values with the greatest frequencies.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The ***symmetry*** of a distribution refers to how the frequencies of values of a variable change in relation to the most common or frequently occurring values of the variable. In ***symmetric distributions*** , the frequencies change in a similar manner, moving away in both directions from the most common values of the variable. In ***asymmetric (skewed) distributions*** , the frequencies change in a different manner, moving away in both directions from the most common values of the variable.

The ***variability*** of a distribution refers to the amount of differences in a distribution of data. In ***peaked distributions*** , much of the data are in a few values of a variable; in ***flat distributions*** , the data are spread evenly across the values of a variable. A ***normally distributed variable*** has a symmetric distribution that is neither peaked nor flat.

2.9 Important Terms

asymmetric (skewed) distribution

bar chart

bimodal distribution

cumulative percentage

flat distribution

frequency (*f*)

frequency distribution table

frequency polygon

grouped frequency distribution table

histogram

modality

multimodal distribution

negatively skewed distribution

normally distributed variable

outlier

peaked distribution

pie chart

positively skewed distribution

real limits

real lower limit

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

real upper limit

symmetric distribution

symmetry

unimodal distribution

variability

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

2.10 Formula Introduced in This Chapter

Percentage (%)

$$\% = \frac{f}{total\ number\ of\ scores} * 100$$

2.11 Using IBM SPSS Software *

Creating Data Files: Defining Variables and Entering Data

* IBM SPSS software. SPSS is a registered trademark of International Business Machines Corporation.

1. Open a new data file (See Figure 2.9).

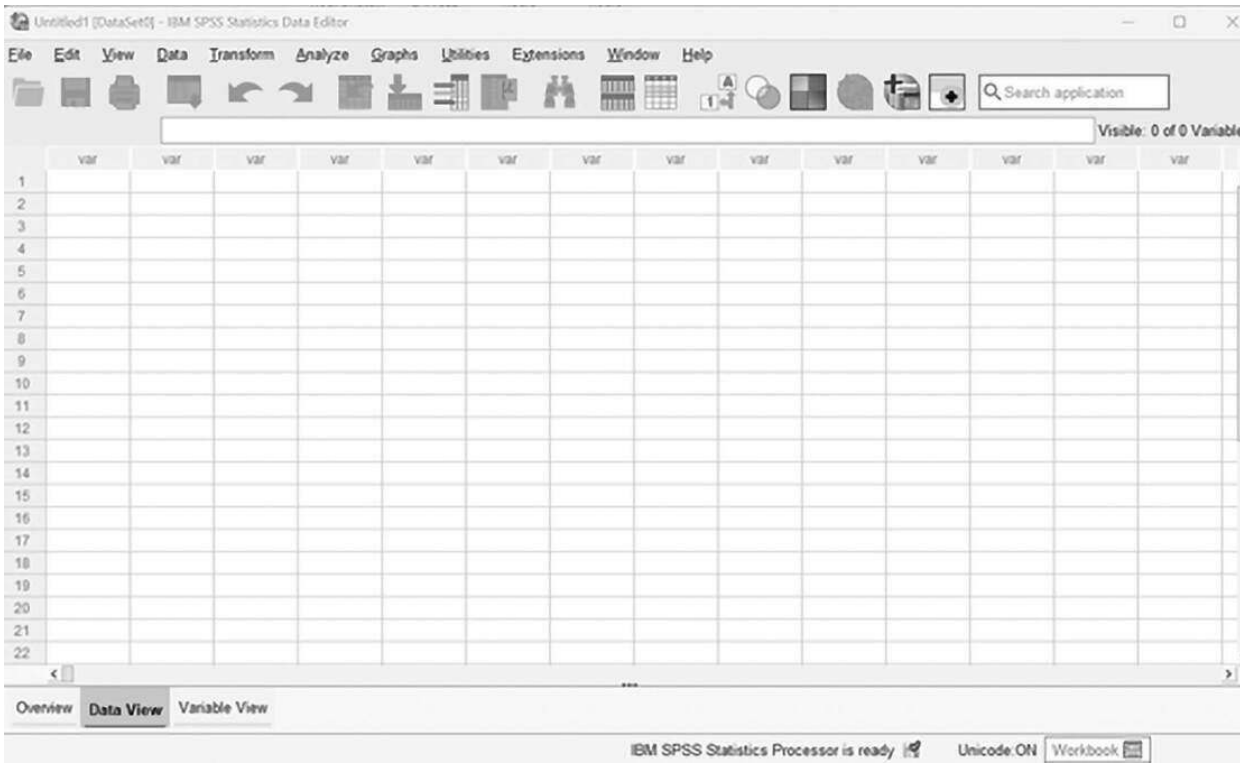


Figure 2.9 Data View

2. Define and name variable(s) (See Figure 2.10).

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

How? (1) Click **Variable View** tab, (2) click below **Name** , and (3) type name of variable in the cell.

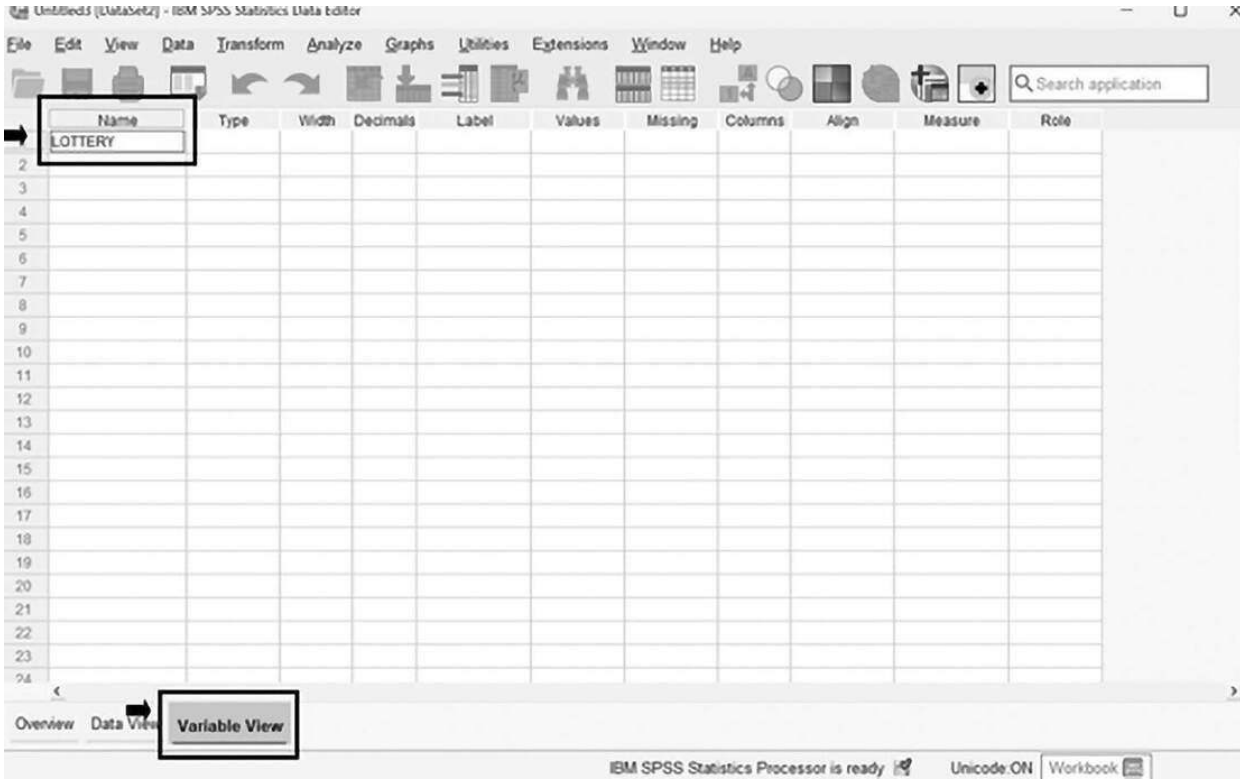


Figure 2.10 Variable View

3. Determine number of decimals for your variable(s) (See Figure 2.11).

How? Click desired number of decimal places in the **Decimals** box.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

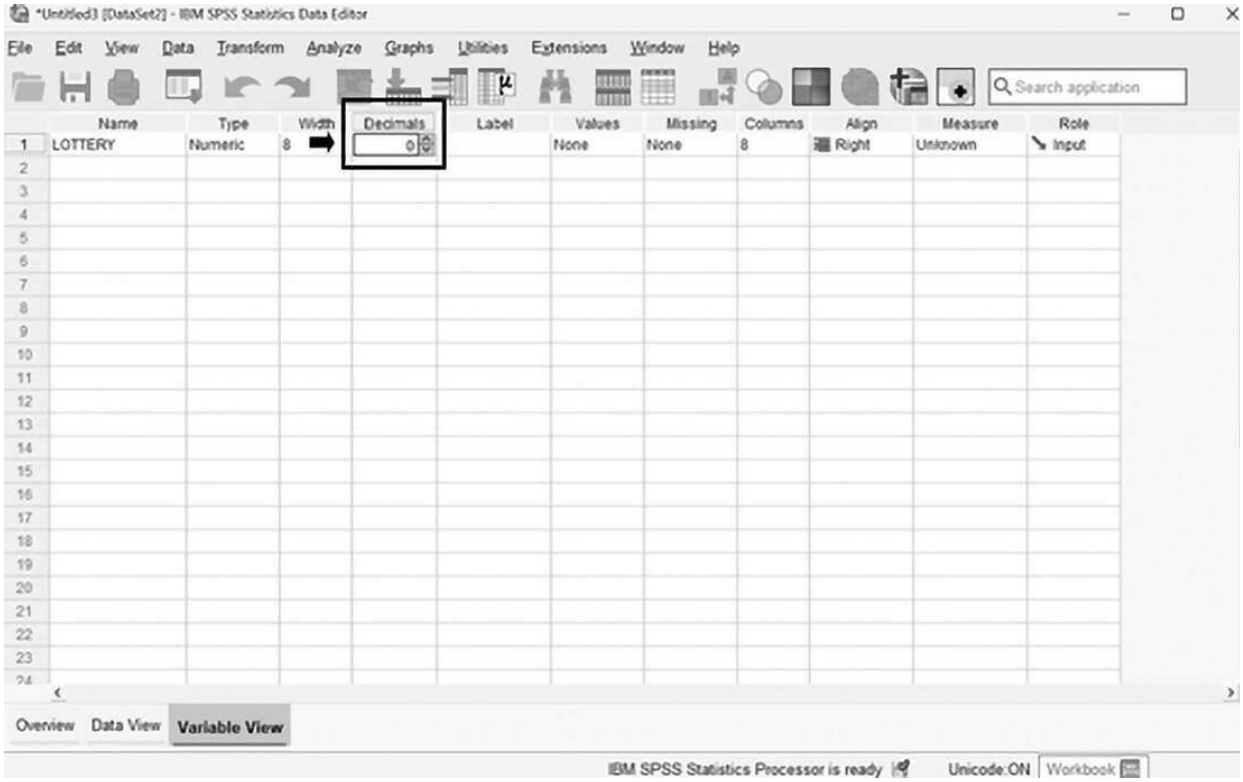


Figure 2.11 Decimals Box

4. Provide labels for your variable(s) (See Figure 2.12).

How? (1) Click below **Label** , and (2) type the label for the variable in the cell.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

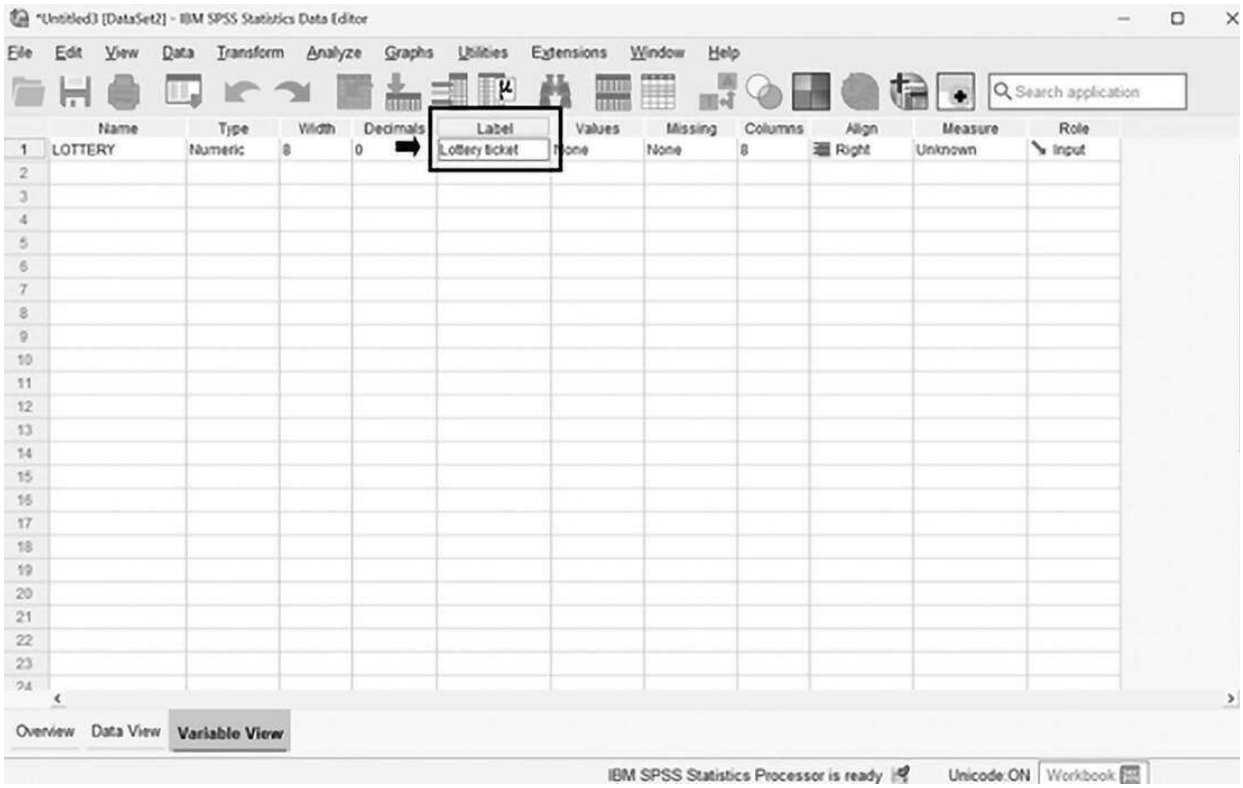


Figure 2.12 Label Box

5. Provide labels for values of nominal (categorical) variable(s) (See Figure 2.13).

How? (1) Click below **Values** , and (2) type labels for coded values of the variable.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

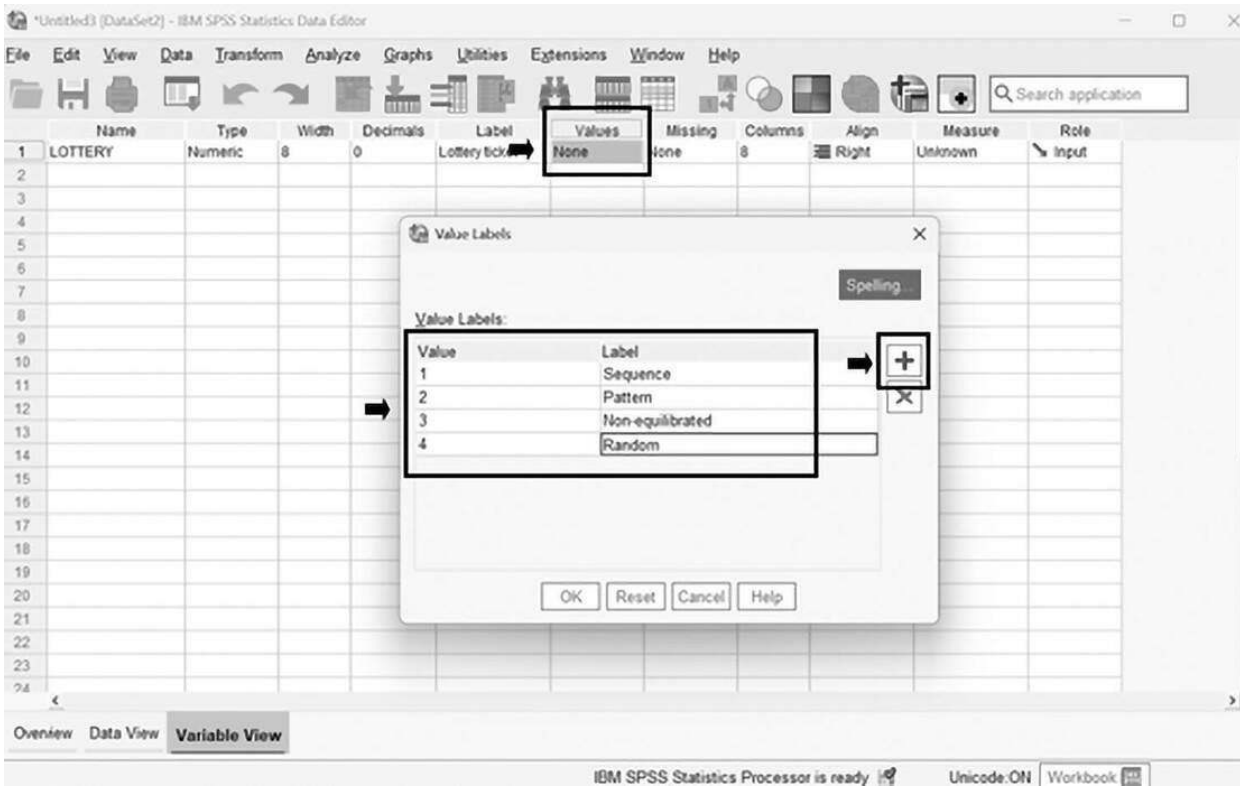


Figure 2.13 Nominal Variables

6. Enter data (See Figure 2.14).

How? (1) Click **Data View** tab and (2) enter data for the variable in the cells.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

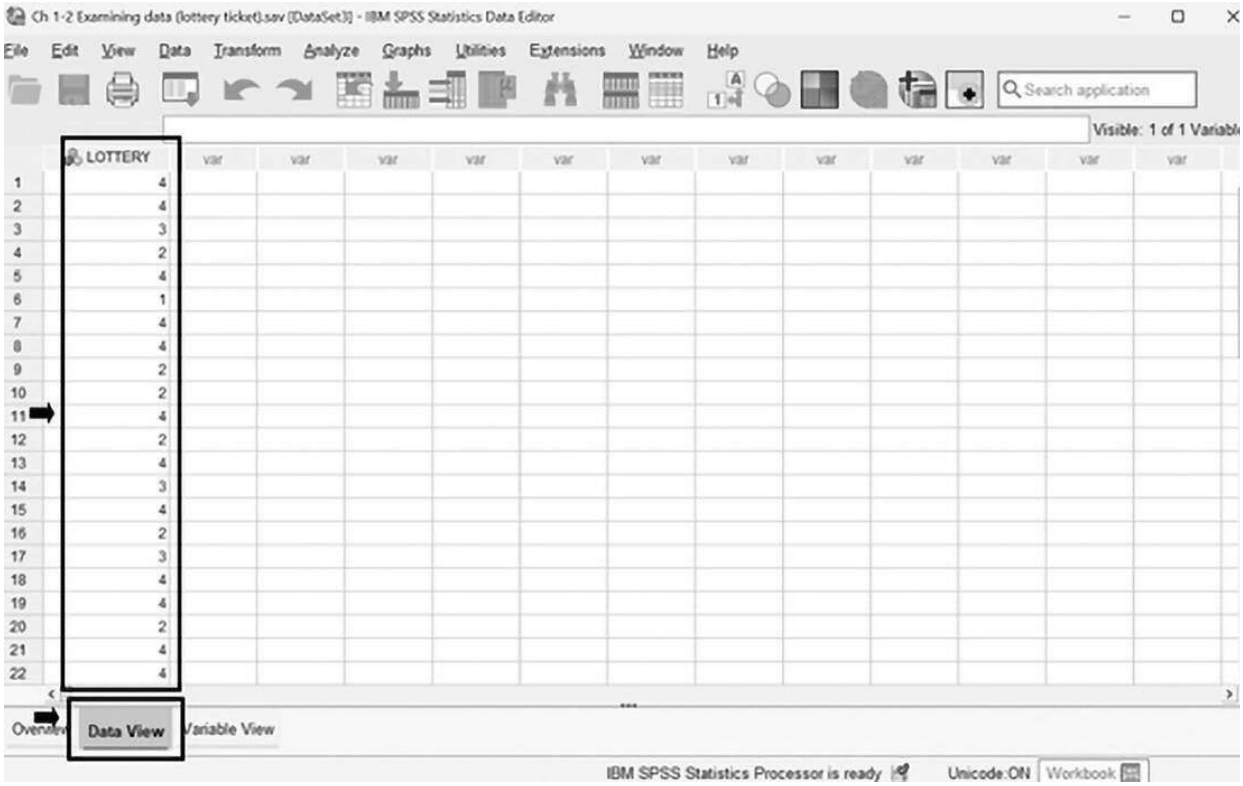


Figure 2.14 Enter Data View

Examining Data: Frequency Distribution Tables and Figures: The Lottery Study (2.1)

1. Select the frequency distribution procedure within SPSS (See Figure 2.15).

   How? (1) Click **Analyze** menu, (2) click **Descriptive Statistics** , and (3) click **Frequencies** .

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

To Jennifer , my kids Meagan and Eika , and my parents, Katsumi and Grayce Tokunaga.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

As a global academic publisher, Sage is driven by the belief that research and education are critical in shaping society. Our mission is building bridges to knowledge—supporting the development of ideas into scholarship that is certified, taught, and applied in the real world.

Sage's founder, Sara Miller McCune, transferred control of the company to an independent trust, which guarantees our independence indefinitely. This enables us to support an equitable academic future over the long term by building lasting relationships, championing diverse perspectives, and co-creating social and behavioral science resources that transform teaching and learning.

▶ Long Description:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

FUNDAMENTAL STATISTICS FOR THE SOCIAL AND BEHAVIORAL SCIENCES

Third Edition

Howard T. Tokunaga

San Jose State University



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

eISBN: 978-1-0718-6160-8

Copyright © 2027 by Sage.

All rights reserved. Except as permitted by U.S. copyright law, no part of this work may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without permission in writing from the publisher.

All third party trademarks referenced or depicted herein are included solely for the purpose of illustration and are the property of their respective owners. Reference to these trademarks in no way indicates any relationship with, or endorsement by, the trademark owner.



FOR INFORMATION:

2455 Teller Road, Thousand Oaks, California 91320, E-mail: orders@sagepub.com

3rd floor, HYLO, 103-105 Bunhill Row, London, EC1Y 8LZ, United Kingdom

10th Floor, Emaar Capital Tower 2, MG Road Sikanderpur, Sector 26, Gurugram, Haryana – 122002, India

18 Cross Street #10-10/11/12, China Square Central, Singapore 048423

Library of Congress Control Number: 2026943277

ISBN: 978-1-0718-6144-8

Printed in the United States of America

This book is printed on acid-free paper.

26 27 28 29 30 10 9 8 7 6 5 4 3 2 1

Acquisitions Editor: Anna Villarruel

Content Development Editor: Kenzie Offley

Editorial Assistant: Daria Terry

Production Editor: Syeda Aina Rahat Ali

Copy Editor: Michelle Ponce

Typesetter: Lumina Datamatics

Cover Designer: Candice Harman

Marketing Manager: Victoria Velasquez

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CONTENTS

Preface
Acknowledgments
About the Author
Chapter 1 Introduction to Statistics
Chapter 2 Examining Data: Tables and Figures
Chapter 3 Measures of Central Tendency
Chapter 4 Measures of Variability
Chapter 5 Normal Distributions
Chapter 6 Probability and Introduction to Hypothesis Testing
Chapter 7 Testing One Sample Mean
Chapter 8 Estimating the Mean of a Population
Chapter 9 Testing the Difference Between Two Means
Chapter 10 Errors in Hypothesis Testing, Statistical Power, and Effect Size
Chapter 11 One-Way Analysis of Variance (ANOVA)
Chapter 12 Two-Way Analysis of Variance (ANOVA)
Chapter 13 Correlation
Chapter 14 Linear Regression and Multiple Correlation
Chapter 15 Chi-Square
Appendix A: Tables
Appendix B: Review of Basic Mathematics
Glossary
References
Index

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

DETAILED CONTENTS

Preface

Acknowledgments

About the Author

Chapter 1 Introduction to Statistics

- 1.1 What Is Statistics?
- 1.2 Why Learn Statistics?
- 1.3 Introduction to the Stages of the Research Process
  - Developing a Research Hypothesis to Be Tested
    - Identifying a Question or Issue to Be Examined
    - Reviewing and Evaluating Relevant Theories and Research
    - Stating a Research Hypothesis: Independent and Dependent Variables
  - Collecting Data
    - Drawing a Sample From a Population
    - Determining How Variables Will Be Measured: Levels of Measurement
    - Selecting a Method to Collect the Data: Experimental and Nonexperimental Research Methods
  - Analyzing the Data
    - Calculating Descriptive Statistics
    - Calculating Inferential Statistics
  - Drawing a Conclusion Regarding the Research Hypothesis
  - Communicating the Findings of the Study
- 1.4 Plan of the Book
- 1.5 Looking Ahead
- 1.6 Summary
- 1.7 Important Terms
- 1.8 Exercises

Answers to Learning Checks

Learning Check 1

Learning Check 2

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Odd-Numbered Exercises

Chapter 2 Examining Data: Tables and Figures

2.1 An Example From the Research: Winning the Lottery

2.2 Why Examine Data?

Gaining an Initial Sense of the Data

Identifying Outliers

Evaluating Research Methodology

Determining Whether Data Meet Statistical Criteria and Assumptions

2.3 Examining Data Using Tables

What Are the Values of the Variable?

How Many Participants in the Sample Have Each Value of the Variable?

What Percentage of the Sample Has Each Value of the Variable?

2.4 Grouped Frequency Distribution Tables

Cumulative Percentages

2.5 Examining Data Using Figures

Displaying Nominal and Ordinal Variables: Bar Charts and Pie Charts

Displaying Interval and Ratio Variables: Histograms and Frequency Polygons

Drawing Inappropriate Conclusions From Figures

2.6 Examining Data: Describing Distributions

Modality

Symmetry

Variability

2.7 Looking Ahead

2.8 Summary

2.9 Important Terms

2.10 Formula Introduced in This Chapter

Percentage (%)

2.11 Using IBM SPSS Software *

Creating Data Files: Defining Variables and Entering Data

Examining Data: Frequency Distribution Tables and Figures: The Lottery Study (2.1)

2.12 Exercises

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 3 Measures of Central Tendency
3.1 An Example From the Research: The 10% Myth
3.2 Understanding Central Tendency
3.3 The Mode
Determining the Mode
3.4 The Median
Calculating the Median: Odd Number of Scores
Calculating the Median: Even Number of Scores
Calculating the Median When Multiple Scores Have the Median Value
3.5 The Mean
Calculating the Mean
Reporting Calculated Numbers
Rounding Calculated Numbers
Calculating the Mean From Frequency Distribution Tables
The Mean as a Balancing Point in a Distribution
The Sample Mean vs. the Population Mean: Statistics vs. Parameters
3.6 Comparison of the Mode, Median, and Mean
Strengths and Weaknesses of the Mode
Strengths and Weaknesses of the Median
Strengths and Weaknesses of the Mean
Choosing a Particular Measure of Central Tendency
3.7 Measures of Central Tendency: Drawing Conclusions
3.8 Looking Ahead
3.9 Summary
3.10 Important Terms
3.11 Formulas Introduced in This Chapter
Median (Odd Number of Scores)
Median (Even Number of Scores)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Mean
Mean (Frequency Distribution Table)
Mean (Balancing Point in a Distribution)
3.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 4 Measures of Variability
4.1 An Example From the Research: How Many Sometimes" in an Always"?
4.2 Understanding Variability
4.3 The Range
Strengths and Weaknesses of the Range
4.4 The Interquartile Range
Strengths and Weaknesses of the Interquartile Range
4.5 The Variance (s^2)
Definitional Formula for the Variance
Computational Formula for the Variance
Why Not Use the Absolute Value of the Deviation in Calculating the Variance?
Why Divide by N − 1 Rather Than N in Calculating the Variance?
4.6 The Standard Deviation (s)
Definitional Formula for the Standard Deviation
Computational Formula for the Standard Deviation
Eyeball Estimating" the Mean and Standard Deviation
4.7 Measures of Variability for Populations
The Population Variance (σ^2)
The Population Standard Deviation (σ)
Measures of Variability for Samples vs. Populations
4.8 Measures of Variability for Nominal Variables
The Index of Qualitative Variation
4.9 Measures of Variability: Drawing Conclusions

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

4.10 Looking Ahead
4.11 Summary
4.12 Important Terms
4.13 Formulas Introduced in This Chapter
Range
Interquartile Range
Variance (s^2) (Definitional Formula)
Variance (s^2) (Computational Formula)
Standard Deviation (s) (Definitional Formula)
Standard Deviation (s) (Computational Formula)
Population Variance (σ^2) (Definitional Formula)
Population Standard Deviation (σ) (Definitional Formula)
Index of Qualitative Variation (IQV)
4.14 Using SPSS
Calculating Measures of Central Tendency and Variability: The Frequency Rating Study (4.1)
4.15 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 5 Normal Distributions
5.1 Normal Distributions
What Are Normal Distributions?
Why Are Normal Distributions Important?
Is There Only One Normal Distribution?
5.2 The Standard Normal Distribution
What Is the Standard Normal Distribution?
Interpreting z-Scores: Distance From the Mean
Interpreting z-Scores: Relative Position Within the Distribution
Finding the Area Between the Mean and a z-Score

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Finding the Area Less or Greater Than a z-Score
Finding the Area Between Two z-Scores
5.3 Applying z-Scores to Normal Distributions
Transforming Normal Distributions Into the Standard Normal Distribution
Understanding the z-Score Formula
Example: Transforming SAT Scores Into z-Scores
Interpreting z-Scores: Distance From the Mean
Interpreting z-Scores: Relative Position Within Distribution
Finding the Area Between the Mean and a Score
Finding the Area Less or Greater Than a Score
Finding the Area Between Two Scores
Transforming Scores Into z-Scores: Drawing Conclusions
5.4 Standardizing Frequency Distributions
Example: Measuring Politicians' Behavior
Standardizing a Frequency Distribution
Interpreting Standardized Scores
Using Standardized Scores to Compare or Combine Variables
Characteristics of Standardized Distributions
When Can You Apply the Normal Curve Table to a Standardized Distribution?
5.5 Looking Ahead
5.6 Summary
5.7 Important Terms
5.8 Formulas Introduced in This Chapter
z-Score
Standardized Score
5.9 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 6 Probability and Introduction to Hypothesis Testing

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

6.1 A Brief Introduction to Probability
- What Is Probability?
- Why Is Probability Important to Researchers?
- Applying Probability to Normal Distributions
- Applying Probability to Binomial Distributions

6.2 Example: Making Heads or Tails of the Super Bowl

6.3 Introduction to Hypothesis Testing
- State the Null and Alternative Hypotheses (H_0 and H_1)
  - The Null Hypothesis (H_0)
  - The Alternative Hypothesis (H_1)
- Make a Decision About the Null Hypothesis
  - Defining a Low" Probability of a Statistic: Setting Alpha (α)
  - Identifying the Values of the Statistic With a Low Probability
  - Stating a Decision Rule
  - Calculating a Value of a Statistic
  - Making the Decision Whether to Reject the Null Hypothesis
- Draw a Conclusion From the Analysis
- Relate the Result of the Analysis to the Research Hypothesis
- Summary of Steps in Hypothesis Testing

6.4 Issues Related to Hypothesis Testing: An Introduction
- The Issue of Proof" in Hypothesis Testing
- Errors in Decision Making
- Factors Influencing the Decision About the Null Hypothesis
  - Sample Size
  - Alpha (α)
  - Directionality of the Alternative Hypothesis

6.5 Looking Ahead

6.6 Summary

6.7 Important Terms

6.8 Formula Introduced in This Chapter

Probability

6.9 Exercises

Answers to Learning Checks

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 7 Testing One Sample Mean
7.1 An Example From the Research: Do You Read Me?
7.2 The Sampling Distribution of the Mean
Characteristics of the Sampling Distribution of the Mean
7.3 Inferential Statistics: Testing One Sample Mean (σ Known)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: z-Test for One Mean
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the z-Test for One Mean
Summary
7.4 A Second Example From the Research: Unique Invulnerability
7.5 Introduction to the t-Distribution
Characteristics of the Student t-Distribution
7.6 Inferential Statistics: Testing One Sample Mean (σ Not Known)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: t-Test for One Mean
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the t-Test for One Mean

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Summary
7.7 Factors Affecting the Decision About the Null Hypothesis
Sample Size
Alpha (α)
Directionality of the Alternative Hypothesis
7.8 Looking Ahead
7.9 Summary
7.10 Important Terms
7.11 Formulas Introduced in This Chapter
z-Test for One Mean
Population Standard Error of the Mean $\left(\mathbf{i}t\sigma_{\mathbf{i}t\overline{X}}\right)$
Degrees of Freedom (df) for One Sample
t-Test for One Mean
Standard Error of the Mean $\left(\mathbf{i}ts_{\mathbf{i}t\overline{X}}\right)$
7.12 Using SPSS
Testing One Sample Mean (σ Not Known): The Unique Invulnerability Study (7.6)
7.13 Exercises
Answers to Learning Checks
Learning Check 2
Learning Check 3
Learning Check 5
Learning Check 6
Answers to Odd-Numbered Exercises
Chapter 8 Estimating the Mean of a Population
8.1 An Example From the Research: Salary Survey
8.2 Introduction to the Confidence Interval for the Mean
Using the Sampling Distribution of the Mean to Create the Confidence Interval for the Mean
8.3 The Confidence Interval for the Mean (σ Not Known)
State the Desired Level of Confidence
Calculate the Confidence Interval and Confidence Limits
Calculate the Standard Error of the Mean $\left(s_{\overline{X}}\right)$

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Calculate the Degrees of Freedom (df) and Identify the Critical Value of t

Calculate the Confidence Interval and Confidence Limits

Draw a Conclusion About the Confidence Interval

Probability of an Interval or the Probability of a Population Mean?

Summary

8.4 The Confidence Interval for the Mean (σ Known)

State the Desired Level of Confidence

Calculate the Confidence Interval and Confidence Limits

Calculate the Population Standard Error of the Mean $\left(\sigma_{\overline{X}}\right)$

Identify the Critical Value of z

Calculate the Confidence Interval and Confidence Limits

Draw a Conclusion About the Confidence Interval

Summary

8.5 Factors Affecting the Width of the Confidence Interval for the Mean

Sample Size

Estimating the Sample Size Needed for a Desired Confidence Interval

Level of Confidence

Choosing Between Different Levels of Confidence: Probability Versus Precision

8.6 Interval Estimation and Hypothesis Testing

Concerns About Hypothesis Testing

The Interpretation of Statistical Significance and Nonsignificance

The Influence of Sample Size on the Decision About the Null Hypothesis

Interval Estimation as an Alternative to Hypothesis Testing

8.7 Looking Ahead

8.8 Summary

8.9 Important Terms

8.10 Formulas Introduced in This Chapter

Confidence Interval for the Mean (σ Not Known)

Confidence Interval for the Mean (σ Known)

Estimated Sample Size for the Confidence Interval for the Mean

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

8.11 Using SPSS
Calculating the confidence interval for the mean (The salary survey study (8.1))
8.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 9 Testing the Difference Between Two Means
9.1 An Example From the Research: You Can Just Wait
9.2 The Sampling Distribution of the Difference
Characteristics of the Sampling Distribution of the Difference
9.3 Inferential Statistics: Testing the Difference Between Two Sample Means
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: t-Test for Independent Means
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the t-Test for Independent Means
Summary
9.4 Inferential Statistics: Testing the Difference Between Two Sample Means (Unequal Sample Sizes)
An Example From the Research: Kids' Motor Skills and Fitness
Inferential Statistics: Testing the Difference Between Two Sample Means (Unequal Sample Sizes)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Relate the Result of the Analysis to the Research Hypothesis

Assumptions of the t-Test for Independent Means (Unequal Sample Sizes)

Summary

9.5 Inferential Statistics: Testing the Difference Between Paired Means

Research Situations Appropriate for Within-Subjects Research Designs

An Example From the Research: Web-Based Family Interventions

Inferential Statistics: Testing the Difference Between Paired Means

Calculate the Difference Between the Paired Scores

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Draw a Conclusion From the Analysis

Relate the Result of the Analysis to the Research Hypothesis

Assumptions of the t-Test for Dependent Means

Summary

9.6 Looking Ahead

9.7 Summary

9.8 Important Terms

9.9 Formulas Introduced in This Chapter

Degrees of Freedom (df), Difference Between Two Sample Means

t-Test for Independent Means

Standard Error of the Difference $\left(s_{\overline{X}_1-\overline{X}_2}\right)$

Standard Error of the Difference (Unequal Sample Sizes)

Degrees of Freedom (df), Difference Between Paired Means

t-Test for Dependent Means

Standard Error of Difference Scores (s D)

9.10 Using SPSS

Testing the Difference Between Two Sample Means: The Parking Lot Study (9.1)

Testing the Difference Between Paired Means: The Web-Based Intervention Study (9.5)

9.11 Exercises

Answers to Learning Checks

Learning Check 1

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Answers to Odd-Numbered Exercises
Chapter 10 Errors in Hypothesis Testing, Statistical Power, and Effect Size
10.1 Hypothesis Testing vs. Criminal Trials
10.2 An Example From the Research: Truth or Consequences
Correct Decisions and Mistakes in Criminal Trials
10.3 Two Errors in Hypothesis Testing: Type I and Type II Error
Type I Error: The Risk in Rejecting the Null Hypothesis
What Is the Probability of Making a Type I Error?
What Is the Probability of Not Making a Type I Error?
Why Is Type I Error a Concern?
Type II Error: The Risk in Not Rejecting the Null Hypothesis
What Is the Probability of Making a Type II Error?
What Is the Probability of Not Making a Type II Error?
Why Is Type II Error a Concern?
Why Does Type II Error Occur?
Type I and Type II Error: Summary
10.4 Controlling Type I and Type II Error
Controlling Type I Error
Concerns About Controlling Type I Error
Controlling Type II Error
Increasing Sample Size
Raising Alpha (α)
Using a Directional Alternative Hypothesis (H1)
Increasing Between-Group Variability
Decreasing Within-Group Variability
Controlling Type I and Type II Error: Summary
10.5 Measures of Effect Size
Interpreting Inferential Statistics: The Influence of Sample Size
The Inappropriateness of “Highly Significant”

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Effect Size as Variance Accounted for

Measure of Effect Size for the Difference Between Two Sample Means: r^2

Interpreting Measures of Effect Size

Presenting Measures of Effect Size

Reasons for Calculating Measures of Effect Size

A Second Measure of Effect Size for the Difference Between Two Sample Means: Cohen's d

10.6 Looking Ahead

10.7 Summary

10.8 Important Terms

10.9 Formulas Introduced in This Chapter

Probability of Type I Error

Probability of Type II error

r Squared (r^2)

Cohen's d

10.10 Exercises

Answers to Learning Checks

Learning Check 1

Learning Check 2

Learning Check 3

Learning Check 4

Answers to Odd-Numbered Exercises

Chapter 11 One-Way Analysis of Variance (ANOVA)

11.1 An Example From the Research: It's Your Move

11.2 Introduction to Analysis of Variance (ANOVA)

Review: Testing the Difference Between the Means of Two Samples

Understanding Between-Group and Within-Group Variability

Relating Between-Group and Within-Group Variability

Introduction to Analysis of Variance (ANOVA) and the F-Ratio

Characteristics of the F-Ratio Distribution

11.3 Inferential Statistics: One-Way Analysis of Variance (ANOVA)

State the Null and Alternative Hypotheses (H_0 and H_1)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Value, and State a Decision Rule
Calculate a Statistic: F-Ratio for the One-Way ANOVA
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (R^2)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Summary
11.4 A Second Example: The Parking Lot Study Revisited
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Value, and State a Decision Rule
Calculate a Statistic: F-Ratio for the One-Way ANOVA
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (R^2)
Draw a Conclusion From the Analysis
The Relationship Between the t-Test and the F-Ratio
11.5 Analytical Comparisons Within the One-Way ANOVA
Planned Versus Unplanned Comparisons
Between-Group and Within-Group Variance in Analytical Comparisons
Conducting Planned Comparisons
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Conducting Unplanned Comparisons
Concerns Regarding Unplanned Comparisons
Methods of Controlling Familywise Error

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Controlling Familywise Error: The Tukey Test

11.6 Looking Ahead

11.7 Summary

11.8 Important Terms

11.9 Formulas Introduced in This Chapter

Degrees of Freedom for Between-Group Variance, One-Way ANOVA (df_{BG})

Degrees of Freedom for Within-Group Variance, One-Way ANOVA (df_{WG})

Between-Group Variance, One-Way ANOVA (MS_{BG})

Total Sample Mean, One-Way ANOVA $\left(\overline{X}_T\right)$

Within-Group Variance, One-Way ANOVA (MS_{WG})

F-Ratio, One-Way ANOVA (F)

Measure of Effect Size, One-Way ANOVA (R^2)

Relationship Between t-Test and F-Ratio for the One-Way ANOVA

F-Ratio, Analytical Comparison (F_{comp})

Between-Group Variance, Analytical Comparison $\left(MS_{\mathrm{BG}_{\mathrm{comp}}}\right)$

Measure of Effect Size, Analytical Comparison $\left(R^2_{\mathrm{comp}}\right)$

Critical Value, Tukey Test for Unplanned Comparisons (F_T)

11.10 Using SPSS

One-Way Analysis of Variance (ANOVA): The Chess Study (11.1)

Analytical Comparisons: Planned Comparisons—Explanation vs. Prediction (11.5)

11.11 Exercises

Answers to Learning Checks

Learning Check 2

Learning Check 3

Learning Check 4

Answers to Odd-Numbered Exercises

Chapter 12 Two-Way Analysis of Variance (ANOVA)

12.1 An Example From the Research: Vote—or Else!

12.2 Introduction to Factorial Research Designs

Testing Interaction Effects

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

An Example of Interaction Effect

An Example of No Interaction Effect

Testing Main Effects

The Relationship Between Main Effects and Interaction Effects

Factorial Designs vs. Single-Factor Designs: Advantages and Disadvantages

12.3 The Two-Factor (A × B) Research Design

Notational System for Components of the Two-Factor Research Design

Analyzing the Two-Factor Research Design: Descriptive Statistics

Notational System for Means in the Two-Factor Research Design

Examining the Descriptive Statistics

12.4 Introduction to Analysis of Variance (ANOVA) for the Two-Factor Research Design

F-Ratios in the Two-Factor Research Design

12.5 Inferential Statistics: Two-Way Analysis of Variance (ANOVA)

State the Null and Alternative Hypotheses (H_0 and H_1)

Main Effect of Factor A

Main Effect of Factor B

A × B Interaction Effect

Make Decisions About the Null Hypotheses

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Values, and State Decision Rules

Calculate Statistics: F-Ratios for the Two-Way ANOVA

Make Decisions Whether to Reject the Null Hypotheses

Determine the Levels of Significance

Calculate Measures of Effect Size (R^2)

Draw Conclusions From the Analyses

Relate the Results of the Analyses to the Research Hypothesis

12.6 Investigating a Significant A × B Interaction Effect: Analysis of Simple Effects

12.7 Looking Ahead

12.8 Summary

12.9 Important Terms

12.10 Formulas Introduced in This Chapter

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Degrees of Freedom for Main Effect of Factor A, Two-Way ANOVA (df_A)
Degrees of Freedom for Main Effect of Factor B, Two-Way ANOVA (df_B)
Degrees of Freedom for A × B Interaction Effect, Two-Way ANOVA ($df_{A\times B}$)
Degrees of Freedom for Within-Group Variance, Two-Way ANOVA (dfWG)
Between-Group Variance for Main Effect of Factor A, Two-Way ANOVA (MS_A)
Between-Group Variance for Main Effect of Factor B, Two-Way ANOVA (MS_B)
Sums of Squares for Between-Group Variance, Two-Way ANOVA (SS_{BG})
Between-Group Variance for A × B Interaction, Two-Way ANOVA ($MS_{A\times B}$)
Within-Group Variance, Two-Way ANOVA (MS_{WG})
F-Ratio for Main Effect of Factor A, Two-Way ANOVA (F_A)
F-Ratio for Main Effect of Factor B, Two-Way ANOVA (F_B)
F-Ratio for A × B Interaction, Two-Way ANOVA ($F_{A\times B}$)
Measure of Effect Size for Main Effect of Factor A, Two-Way ANOVA(R^2_A)
Measure of Effect Size for Main Effect of Factor B, Two-Way ANOVA(R^2_B)
Measure of Effect Size for A × B Interaction, Two-Way ANOVA(R^2_{AB})
12.11 Using SPSS
Two-Way Analysis of Variance (ANOVA): The Voting Message Study (12.1)
12.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 13 Correlation
13.1 An Example From the Research: It's Good for You!
13.2 Introduction to the Concept of Correlation
Describing the Relationship Between Variables
Nature of the Relationship
Direction of the Relationship
Strength of the Relationship

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Measuring the Relationship Between Variables: Correlational Statistics
Introduction to the Pearson Correlation Coefficient (r)
Calculating Correlational Statistics: The Role of Variance and Covariance
13.3 Inferential Statistics: Pearson Correlation Coefficient
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: Pearson Correlation Coefficient (r)
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (r^2)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the Pearson Correlation
A Computational Formula for the Pearson r
Represent the Covariance Between X and Y (SPXY)
Represent the Variance of Variable X (SSX)
Represent the Variance of Variable Y (SSY)
Calculate the Pearson Correlation Coefficient (r)
13.4 Correlating Two Sets of Ranks: The Spearman Rank-Order Correlation
Introduction to Ranked Variables
An Example From the Research: Do Students and Teachers Think Alike?
Inferential Statistics: The Spearman Rank-Order Correlation Coefficient
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
13.5 Correlational Statistics vs. Correlational Research
Does ANOVA = Causality?
“Correlation Is Not Causality”—What Does That Mean?
13.6 Looking Ahead

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

13.7 Summary
13.8 Important Terms
13.9 Formulas Introduced in This Chapter
Degrees of Freedom for Pearson Correlation (df)
Sum of Products (Definitional Formula) (SPXY)
Sum of Squares for the X Variable (Definitional Formula) (SSX)
Sum of Squares for the Y Variable (Definitional Formula) (SSY)
Pearson Correlation (Definitional Formula) (r)
Measure of Effect Size for the Pearson Correlation Coefficient (r^2)
Sum of Products (Computational Formula) (SPXY)
Sum of Squares for the X Variable (Computational Formula) (SSX)
Sum of Squares for the Y Variable (Computational Formula) (SSY)
Pearson Correlation (Computational Formula) (r)
Spearman Rank-Order Correlation (rs)
Measure of Effect Size for the Spearman Rank-Order Correlation $\left(r_s^2\right)$
13.10 Using SPSS
Pearson Correlation: The Picky Eater Study (13.1)
Spearman Rank-Order Correlation: The Business Student Study (13.4)
13.11 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Answers to Odd-Numbered Exercises
Chapter 14 Linear Regression and Multiple Correlation
14.1 Predicting One Variable From Another: Linear Regression
The Linear Regression Equation
Calculating the Linear Regression Equation
Calculate the Slope of the Equation (b)
Calculate the Y-Intercept of the Equation (a)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Report the Linear Regression Equation

Drawing the Linear Regression Equation

A Second Example: Put Down Your Pencils

Calculating the Linear Regression Equation

Drawing the Linear Regression Equation

14.2 Correlation With Two or More Predictors: Introduction to Multiple Correlation and Regression

The Limitations of Using Only One Predictor

Introduction to Multiple Correlation

Multiple Correlation With Two Predictors: An Example

Measuring the Relationship Between Two Predictors and a Criterion: The Multiple Correlation (R)

Testing the Multiple Correlation for Statistical Significance

Predicting a Criterion From Two Predictors: The Multiple Regression Equation

Conclusion

14.3 Looking Ahead

14.4 Summary

14.5 Important Terms

14.6 Formulas Introduced in This Chapter

Linear Regression Equation

Estimate of Slope (b)

Estimate of Y-Intercept (a)

Multiple Correlation Coefficient (R)

F-Ratio for the Multiple Correlation Coefficient (F)

Multiple Regression Equation

Regression Coefficient

Estimate of Y-Intercept (a) (Multiple Regression)

14.7 Using SPSS

Linear Regression: The Picky Eater Study (14.1)

Multiple Regression: The College Admission Example (14.2)

14.8 Exercises

Answers to Learning Checks

Learning Check 1

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 2

Answers to Odd-Numbered Exercises

Chapter 15 Chi-Square

15.1 An Example From the Research (One Categorical Variable): Are You My Type?

15.2 Introduction to the Chi-Square Statistic

Observed and Expected Frequencies: The Basis of the Chi-Square Statistic

Characteristics of the Distribution of Chi-Square Statistics

15.3 Inferential Statistic: Chi-Square Goodness-of-Fit Test

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Value, and State a Decision Rule

Calculate a Statistic: Chi-Square (χ^2)

Make a Decision Whether to Reject the Null Hypothesis

Determine the Level of Significance

Calculate a Measure of Effect Size (Cramér's φ)

Draw a Conclusion From the Analysis

Relate the Result of the Analysis to the Research Hypothesis

Assumptions Underlying the Chi-Square Statistic

Conducting the Chi-Square Goodness-of-Fit Test With Unequal Hypothesized Proportions

Calculate Expected Frequencies (f_e)

Calculate the Chi-Square Statistic (χ^2)

15.4 An Example From the Research (Two Categorical Variables): Seeing Red

15.5 Inferential Statistic: Chi-Square Test of Independence

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Value, and State a Decision Rule

Calculate a Statistic: Chi-square (χ^2)

Make a Decision Whether to Reject the Null Hypothesis

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Determine the Level of Significance
Calculate a Measure of Effect Size (Cramér's φ)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
15.6 Parametric and Nonparametric Statistical Tests
Parametric vs. Nonparametric Statistical Tests
Reasons for Using Nonparametric Statistical Tests
Examples of Nonparametric Statistical Tests
15.7 Looking Ahead
15.8 Summary
15.9 Important Terms
15.10 Formulas Introduced in This Chapter
Degrees of Freedom for the Chi-Square Goodness-of-Fit Test
Expected Frequency for the Chi-Square Goodness-of-Fit Test
Chi-Square Statistic (χ^2)
Cramér's φ for the Chi-Square Goodness-of-Fit Test
Degrees of Freedom for the Chi-Square Test of Independence
Expected Frequency for the Chi-Square Test of Independence
Cramér's φ for the Chi-Square Test of Independence
15.11 Using SPSS
Chi-Square Goodness-of-Fit Test—Equal Hypothesized Proportions: The Personality Study (15.1)
Chi-Square Goodness-of-Fit Test—Unequal Hypothesized Proportions: The Personality Study (15.3)
Chi-Square Test of Independence: The Color Preference Study (15.4)
Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Learning Check 6

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Odd-Numbered Exercises

Appendix A: Tables

Appendix B: Review of Basic Mathematics

I. Symbols

II. Terms

III. Addition, Subtraction, Multiplication, Division

A. All Positive Numbers

B. Adding All Negative Numbers

C. Adding a Combination of Positive and Negative Numbers

D. Subtracting Negative Numbers

E. Multiplying Two Negative Numbers

F. Multiplying a Combination of Positive and Negative Numbers

G. Dividing Two Negative Numbers

H. Dividing One Positive Number and One Negative Number

IV. Order of Operations

V. Fractions, Decimals, Percentages

A. Fraction

B. Decimal

C. Percentage

VI. Solving Simple Equations With One Unknown (X)

A. X Has a Value Subtracted From It (Solve for X Using Addition)

B. X Has a Value Added to It (Solve for X Using Subtraction)

C. X Is Multiplied by a Value (Solve for X Using Division)

D. X Is Divided by a Value (Solve for X Using Multiplication)

VII. Solving More Complex Equations With One Unknown (X)

A. X Is Multiplied by a Value and Then Has a Value Either Added to It or Subtracted From It (Solve for X Using Either Subtraction or Addition and Then Division)

B. X Is Divided by a Value and Then Has a Value Either Added to It or Subtracted From It (Solve for X Using Either Addition or Subtraction and Then Multiplication)

C. X Has a Value Added to It or Subtracted From It and Is Then Multiplied by a Value (Solve for X Using Division and Then Either Subtraction or Addition)

D. X Has a Value Either Added to It or Subtracted From It and Is Then Divided by a Value (Solve for X Using Multiplication and Then Either Subtraction or

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Addition)
Glossary
References
Index

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

PREFACE

A Few (?) Words About Me

When I entered college, I thought my path would take me to law school; however, taking an Intro to Psych class made me realize I enjoy the challenge of trying to understand the human mind. The college I attended, U.C. Santa Cruz, was a fairly unconventional university, but somehow in the midst of a sea of humanistic psychologists I became attracted to the empirical and methodological aspects of psychology by taking classes from instructors such as David Harrington and Dane Archer, who showed me that statistics could appeal to students if taught with a gentle, guiding approach that addresses questions relevant to students' lives. After graduating from college, I was able to get a job as a research assistant for a human resource consulting firm. Despite my lack of work experience, I was hired primarily as a result of having taken statistics and research methods courses, which taught me that learning statistics has benefits both inside and outside of the classroom.

Several years later, I started grad school at U.C. Berkeley, where two events critical to this book took place. First, serving as a teaching assistant, I found I really enjoyed helping students, particularly in statistics and research methods classes often approached with fear and suspicion. Second, I took graduate classes from Geoff Keppel, who had developed his own unique system for analyzing experimental research designs. His lectures and books showed me that statistics can be taught in a systematic way that highlights similarities rather than differences between different research situations. Geoff managed to transform something as daunting-sounding as a "3 x 2 x 4 research design" into the mathematical equivalent of playing with wooden toy alphabet blocks labeled "A," "B," and "C."

After leaving Cal, I was hired by San Jose State, where I teach an introductory statistics course aimed at students with a wide range of backgrounds, ability, and motivation. When I went to select a textbook for the course, for the first time I looked carefully at the wide range of offerings. What I found striking (and still find striking) was that the majority of books focused on providing formulas and very small sets of data designed to demonstrate how to correctly calculate the correct numbers from these formulas. Little emphasis, however, was given to what these numbers meant or implied. Given my own experiences learning and using statistics, I thought a book was needed that focusses on how statistics may be used to answer questions and test ideas within the larger research process.

Goal of this Book

This textbook, aimed primarily at students majoring in the behavioral and social sciences, includes many of the same topics and chapters as the majority of existing introductory statistics books. However, what I believe distinguishes this book is its thematically-based approach that places statistics within the research process. The primary purpose of this

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

book is to not just teach students how to calculate statistics but how to interpret the results of statistical analyses in light of a study's research hypothesis, and to communicate one's results and interpretations to a broader audience. Hopefully, this book will not only help students understand the purpose and use of statistics, but it will also give them a greater appreciation of how research studies are conceived, conducted, and communicated.

The book is intended for students with a limited knowledge of mathematics, statistics, and research design and methodology. Some common obstacles faced by students are their math anxiety, fear of conducting arithmetic calculations, and an inability to see the relevance of the material to their daily lives. To address these concerns, this books attempts to judiciously use mathematical terminology, write in a conversational manner, and utilize actual research situations rather than contrived sets of numbers.

Organization of the Book

The 15 chapters of this book may be placed into three general categories. Chapters 1 to 4 are designed to introduce students to the research process and how data that have been collected may be organized, presented, and summarized:

Chapter 1 (Introduction to Statistics) defines the concept of statistics, explains why statistics are important to learn, and describes the steps in the research process that are the framework for the rest of the book.

Chapter 2 (Examining Data: Tables and Figures) discusses the importance of examining data that have been collected and describes how tables and figures may be created to represent data.

Chapter 3 (Measures of Central Tendency) presents measures of central tendency (the mode, median, and the mean) and compares them in terms of their relative strengths and weaknesses.

Chapter 4 (Measures of Variability) presents measures of variability (range, interquartile range, variance, standard deviation, index of qualitative variation) and the concept of parameters (population variance and standard deviation).

Chapters 5 through 10 discuss the process of conducting statistical analyses to test research questions and hypotheses, as well as issues and controversies regarding this process:

Chapter 5 (Normal Distributions) introduces theoretical distributions, starting with normal distributions, before moving on to the standard normal distribution (normal curve).

Chapter 6 (Probability and Introduction to Hypothesis Testing) presents the structure for later chapters by introducing, defining, and illustrating steps in hypothesis testing.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Chapter 7 (Testing One Sample Mean) tests one sample mean using the *z* -test and *t* -test and introduces sampling distributions, the central limit theorem, and the Student *t* distribution.

Chapter 8 (Estimating the Mean of a Population) introduces interval estimation by developing the confidence interval for the mean and discusses interval estimation as an alternative to null hypothesis significance testing (NHST).

Chapter 9 (Testing the Difference Between Two Means) covers the *t* -test for independent means for both equal and unequal sample size and introduces within-subjects research designs using the *t* -test for dependent means.

Chapter 10 (Errors in Hypothesis Testing, Statistical Power, and Effect Size) discusses Type I and Type II error and introduces measures of effect size as a way of drawing appropriate conclusions about an analysis.

Chapters 11 to 15 discuss different statistical procedures used in research situations that vary in the number of independent variables as well as how the variables have been measured:

Chapter 11 (One-way Analysis of Variance (ANOVA)) discusses the one-way ANOVA, the relationship between the *t* -statistic and the F-ratio, and planned and post-hoc comparisons.

Chapter 12 (Two-way Analysis of Variance (ANOVA)) introduces factorial research designs, the calculation of F-ratios for main effects and interaction effects, and illustrates the use of simple effects analyses to test research hypotheses.

Chapter 13 (Correlation) defines and illustrates the concept of correlation, discusses the Pearson and Spearman correlation coefficients, and makes the distinction between correlational statistics and correlational research methods.

Chapter 14 (Linear Regression and Multiple Correlation) introduces the concept of regression, discusses linear regression, and works through an example of multiple regression with two predictor variables.

Chapter 15 (Chi-Square) describes the analysis of categorical variables using the chi-square goodness-of-fit test and the chi-square test of independence before ending the book with a brief introduction to nonparametric statistics.

In most cases, the chapters in this book share a uniform format and structure, centered on the research process, based on examples of actual published research studies addressing topics relevant to students' lives. For each example, the research hypothesis is clearly identified; after a brief description of how the data in the study were collected, the calculation and interpretation of both descriptive and inferential statistics are presented,

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ending with the extent to which the results of the statistical analyses support the research hypothesis and how these findings may be communicated to a broader audience.

Pedagogical Features

The chapters in this textbook contain several reoccurring features designed to enhance students' comprehension of the material. These include the following:

Periodic learning checks within chapters. Short learning checks, in the form of conceptual review questions and computational exercises, are provided throughout each chapter.

Explicit presentation of formulas and calculations. This book provides both definitional formulas to highlight their relation to the concept being measured and computational formulas to facilitate calculations. In applying these formulas, all of the steps in the calculation of statistics such as the *t* -test and F-ratio are included to help students in completing exercise and homework assignments.

Illustration of the presentation and communication of analyses. Every chapter includes a discussion and illustration of how statistical analyses may be reported as well how they may be communicated to a larger audience in both written and visual form.

Highlighted definition of terms. Every term and concept introduced in this book is boldfaced, immediately followed by an explicit definition.

End of chapter summaries and exercises. The end of each chapter contains a summary of critical points, a list of terms and formulas introduced in the chapter, and a large number of exercises, many of which refer to published research studies.

Calculation of statistics using SPSS. Many chapters include end-of-chapter content that provides an annotated presentation of how statistics may be calculated using the SPSS software. Screen shots of data files, menu selections, and output files are included and annotated to locate and identify relevant aspects of the analysis.

What's New to the 3rd Edition

I'd like to start by expressing my most heartfelt appreciation to the instructors and reviewers of the first two editions of this book as their evaluation, insights, criticisms, and suggestions were absolutely vital to the changes made in this third edition. I'd like to particularly thank the instructors who've adopted the book who were kind enough to spontaneously send me their thoughts and suggestions. It is my sincere hope that all of you see the incorporation of your input in this edition.

There were two main goals in this edition. The first goal was to make the writing as brief and concise as possible, manifested in the goal of reducing the length of each chapter by 5% to

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

10%. This goal was achieved by simplifying the language and shortening or eliminating what I, the book's adopters, and the book's reviewers felt was unnecessary detail or discussion. The second goal was to increase and update the exercises at the end of many of the chapters, including the use of recent published research as the basis of these exercises. Another goal was to update the demonstration of the use of SPSS to analyze data using the most recent version (Version 31 at the time of this book's publication).

More specifically, in terms of content, the following changes were made to these chapters:

Chapter 1 (Introduction to Statistics): The published research studies used to illustrate concepts in the text, learning checks, and exercises were added to and updated using topics relevant to students (e.g., fear of missing out (FOMO), cryptocurrency, legalization of cannabis, Gen Z, COVID, fake news).

Chapter 6 (Probability and Introduction to Hypothesis Testing): The Super Bowl example used to introduce and illustrate the process of hypothesis testing was updated.

Chapter 8 (Estimating the Mean of a Population): The main research example, which involves data from an actual national salary survey, has been updated throughout this chapter with more current salary data.

As with the first two editions, any and all questions, concerns, suggestions, and complaints you have about the book are most welcome and can be sent to me at howard.tokunaga@sjsu.edu .

A Few Tips for Students

To you, the college student about to read this book as part of taking a statistics course: "Welcome!" and "Great job!" We welcome you because you're about to embark on a semester-long journey that we hope enhances your skills and widens your perspective; we congratulate you because it's a journey not everyone is willing to take. We say to our students that by taking just one statistics course they are ahead of 95% of the population in their knowledge of statistics.

We also tell our students that our primary goal for them is not to become a statistics genius but just to become comfortable with the process of analyzing and interpreting data. In that regard, we agree with a 2012 *Forbes* magazine article entitled *The Six Classes That Will Make Any College Grad Employable* that noted "a good statistics class will help the student become comfortable pushing numbers around and avoiding false interpretations."

At the present moment, we know our encouragement may be of little comfort because you might be anxious about having to learn statistics. Some of you might be anxious about having to learn statistical concepts and formulas; others might be anxious because the research process seems pretty complicated. Talking with students who have taken our courses over the years has helped us assemble the following advice:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Master the Material Presented in the Early Chapters

Chapters 1 through 6 discuss how research is typically conducted, how data are summarized and described, and the process by which researchers conduct statistical analyses to test their ideas. It is absolutely critical for you to have a firm understanding of these chapters as they lay the groundwork for later chapters that discuss a variety of statistical procedures used by researchers to test hypotheses.

What does it mean to master this material? First, *read the chapters both before and after they're discussed in class* . Reading the chapters beforehand helps you identify anything that's unclear and have your questions addressed by your instructor; re-reading the chapters afterward helps confirm your understanding of the material. Next, *be able to define and explain key conceptspresented in these chapters* . These concepts are highlighted in bold-faced letters, and it's important to learn them when they're first introduced because they'll appear throughout the remainder of the book. Next, *do the learning checks within the chapters and the exercises at the end of the chapters* , because they include exercises and review questions you can ask yourself to assess your understanding of the material. Most importantly, *do not miss class during the early part of the semester* ! It's been our experience that students who miss critical lectures at the beginning of the term often have difficulty keeping up as the semester continues.

Review High School Algebra

If you happen to move beyond the introductory statistics course for which you're reading this book, you'll find that statistics are often calculated using computers and statistical software rather than hand calculations. Consequently, some statistics textbooks no longer include mathematical formulas but instead have students use statistical software or Microsoft Excel. However, this book takes a different approach for two main reasons. First, we believe that in order to understand the results of statistical analyses, one must understand their underlying foundation. Learning statistics via computer software can lead to a "brain-dead" approach in which students are at the mercy (rather than control) of their computers. We refer to this as the "I'm only as smart as my printout" method of learning statistics, also known as "Because this is what my computer told me."

Second, this book contains mathematical formulas and calculations for one simple reason: The statistics in this book are not hard to calculate. To assess whether you're prepared to read this book, take the following test:

1. Do you know how to add, subtract, multiply, and divide?
2. Do you know how to calculate the square or square root of a number?
3. Do you understand the order of operations and parentheses within mathematical formulas?

If your answers to these questions are "Yes," congratulations! You possess the ability to conduct *every* mathematical calculation in this book because *none* of the formulas involve geometry, trigonometry, or calculus. However, if you're unsure of your mathematical ability, we suggest you review your high school algebra. Also, there's an appendix at the end of this

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

book that includes the mathematical concepts and operations needed to conduct the statistical analyses in this book. We highly recommend you read this appendix to review and assess your mathematical skills.

Learn to Use a Statistical Calculator

Although learning the statistics in this book doesn't require the use of computers or software, you will need a hand calculator. Calculators with statistical capabilities are typically identified as "statistical" or "scientific" calculators. Although there's a wide selection of brands and models, know that the calculators most appropriate for this book cost only $10 to $15. We do *not* recommend you use or purchase a more expensive calculator! We prefer these simpler calculators because students are able to do their calculations more quickly and with fewer errors. We do suggest you purchase a calculator similar to the one your instructor uses as they will be more familiar with its features and idiosyncrasies.

In selecting a calculator, look closely at the keyboard; you'll need a calculator whose keys have labels such as $\bar{X}$, σ, and ΣX (Chapters 3 and 4 discuss what these symbols represent). When you go calculator shopping, you'll find calculators that have keys with labels such as ΣY and XY. Although calculators with these Y symbols are only slightly more expensive than calculators with just the X symbols, we do *not* recommend their purchase because these features are not needed to perform the calculations in this book. In fact, we've found students with Y calculators have a more difficult time performing simpler calculations.

Ask Questions

Conducting research is the process of asking and answering questions. Accordingly, throughout this book issues are often framed in the form of questions, which we hope encourages you to ask questions as well. The first person you should direct any questions to is yourself. As you work your way through the chapters, get into the habit of asking yourself, "What does this mean?" A good test of your level of comprehension is whether you can explain the material to yourself in a meaningful way.

You should also be sure to work with your instructor to confirm your level of understanding and clarify unanswered questions. Instructors often say to students, "The only bad question is the one you don't ask." Don't be hesitant about asking questions in class. Instructors will tell you they use one student's question to assess an entire class's level of understanding. This is what we call the "pencil test"—when one student asks a question, we look to see how many other pencils get raised, which reflects the number of students who had the same question. Asking questions enhances the learning experience for yourself, your classmates, and your instructor.

Form a Study Group

No matter how much we encourage students to ask questions in class, it seems they hesitate for fear of drawing attention to themselves. As a result, we turn to another well-

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

traveled saying: “Misery loves company.” Forming a study group with other students will help you keep up with reading assignments, confirm your comprehension of the material, clarify any confusion you have, prepare for exams, and provide social support. We do suggest that your group contain at least one person who will raise the group’s questions and concerns to the instructor during class or office hours.

Practice, Practice, Practice

The best way to learn the topics covered by this book is to obtain as much practice as possible. Throughout the chapters are examples in which the calculations have been worked out for you; in the middle and end of each chapter are a number of exercises. The answers to some of the exercises are provided at the back of this book, and your instructor has access to the answers to the other exercises. Do the end-of-chapter exercises even if your instructor does not assign them as homework in order to identify any reoccurring mistakes you make. We find that the vast majority of errors students make are simple computational errors, committed when students do the calculations too quickly.

Also note that the publisher of this book, Sage Publications, has developed a webpage for this book that contains a huge amount of supplemental tools such as videos and quizzes for each chapter. Find it by going to the Sage Publishing website **collegepublishing.sagepub.com** , searching for my name (Tokunaga), and looking for the **Resources** tab.

Enough cheerleading and preaching – let’s get started!

WELCOME!

GREAT JOB!

Additional Teaching Resources

Visit **collegepublishing.sagepub.com** , and navigate to the Resources tab on your book’s page to find the teaching materials designed to accompany this textbook. On this site you will find an array of materials that will save you time and to help you keep students engaged, including the following:

Learning management system cartridges that easily integrate with your course management system so student test results and graded assignments seamlessly flow into your gradebook

Test banks , aligned to Bloom’s Taxonomy, that provide a diverse range of test items, including multiple choice, true/false, and essay questions

Lecture notes that provide an outline and the key concepts in each chapter to aid in lecture preparation

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

PowerPoint ® slides that offer a flexible, accessible, and customizable solution for creating multimedia lectures

Tables and Figures from the book to support lecture preparation and class discussions

Sample course syllabi , including suggested models for structuring your course

Answers to even-numbered questions from the text help facilitate grading

SPSS datasets for use with exercises from the text.

Multimedia content , including original SAGE videos featuring tutorials with author Howard T. Tokunaga that bring concepts to life and appeal to diverse learners

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ACKNOWLEDGMENTS

This book has traveled a long journey, with many people playing both transient and persevering roles in its development. First and foremost, I'd like to thank my kids, Meagan and Will, for providing me perspective and inspiration amid all of the twists and turns we faced while this book was being written. I'm so very blessed to have you as my "burden." Away from home, I'm grateful to my graduate students at San Jose State—working on this book has given me much empathy for their thesis-related travails. However, they should be aware that the completion of this book ends the unspoken agreement that I not ask, "So how's your thesis?" if they don't ask, "So how's your book?" A number of friends have provided most welcome relief after long days staring at a computer screen, with particular thanks going to Ellie, Eileen, Susan, Bonnie, Ann, Lisa, and Terry. I would like to thank Charles Linsmeier and David Shirley for their many helpful comments and advice on earlier versions of this book, Julie West for her help in developing chapter exercises, and Youn-Jeong Choi and Naomi Ran for their careful examination of the chapters and exercises.

I would like to express my appreciation to everyone at SAGE who has played a role in shepherding this book to completion. First and foremost, I must express my gratitude to my editor, Vicki Knight. It has been a joy and an honor to work with someone with such a truly exceptional combination of experience, knowledge, perspective, and humor. In production I was fortunate enough to work with others on the SAGE team who made the process a pleasure: Yvonne McDuffee, editorial assistant; Katie Guarino, assistant editor; Nicole Elliott, executive marketing manager; Candice Harman, cover designer; and the copyeditor, Michelle Ponce. Furthermore, I am grateful to the reviewers who took the time to read the chapters, raise questions, and provide constructive comments and suggestions to assist me in clarifying and strengthening the manuscript. The following are the reviewers for third edition:

Erin M. Fekete, University of Indianapolis

Brooke Magnus, Boston College

Loretta Neal McGregor; Arkansas State University

Michael T. Moore, Adelphi University

Megan M. Ringel, St. Edward's University

Bettina Viereck, University of Hartford

Janice Weaver, Community College of Allegheny County

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ABOUT THE AUTHOR

Howard T. Tokunaga is professor of Psychology at San Jose State University, where he serves as coordinator of the MS Program in Industrial/Organizational (I/O) Psychology and teaches undergraduate and graduate courses in statistics, research methods, and I/O psychology. He received his bachelor's degree in psychology at U.C. Santa Cruz and his PhD in psychology at U.C. Berkeley. In addition to his teaching, he has consulted with a number of public sector and private sector organizations on a wide variety of management and human resource issues. He is author of *Moving from IBM SPSS to R and RStudio: A Statistics Companion* , and coauthor (with G. Keppel) of *Introduction to Design and Analysis* : *A Student's Handbook* . In his spare time, he enjoys the outdoors, museums, live theater, and being as physically, socially, and politically active as possible.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CHAPTER 1 INTRODUCTION TO STATISTICS

Learning Objectives

1.1 Define the concept of statistics.

1.2 Describe different reasons for learning statistics.

1.3 Describe the main steps in the research process and the tasks and concepts within each step.

In introducing this book to you, we assume you're a college student taking what is perhaps your first course in statistics to fulfill a requirement for your major or a general education requirement. If so, you may be asking yourself two questions:

What is statistics?

Why learn statistics?

The ultimate goal of this book is to help you begin to answer these two questions.

1.1 What Is Statistics?

You may not know it, but you encounter a variety of "statistics" in your day-to-day activities: the typical cost of going to college, the yearly income of the average college graduate, the average price of a car, and so on. So what exactly is "statistics"?

The Merriam-Webster dictionary defines **statistics** as *a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data* . When most people think about statistics, they focus on the "analysis" aspect of this definition—that is to say, they focus on numbers that result from analyzing data. However, statistics not only is concerned about how data are *analyzed* but also how data are *collected* and how the results of statistical analyses are *interpreted* and *communicated* . The main purpose of this book is to introduce, describe, and illustrate the role of statistics within the larger research process.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1.2 Why Learn Statistics?

There are many reasons why you should learn statistics. First, not only do you currently encounter statistics in your daily activities, but *throughout your life you are affected by the results of research and statistical analyses* . Which college you attend is based in part on test scores developed by researchers. You may have to take a personality or intelligence test to get a job. The choices of drugs and medicines available to you are based on medical research. Learning about statistics will help you become a more informed and aware consumer of research and statistical analyses that affect many aspects of your life.

A second reason for learning statistics is that *you may be required to interpret the results of statistical analyses* . Many college courses require students to read academic research journal articles. Evaluating published research is complicated by the fact that different people studying the same topic may come up with diverse or even opposing conclusions. Understanding statistics and their role in the research process will help you decide whether conclusions drawn in research articles are appropriate and justified.

Another reason for learning statistics is that *you might be required to do your own research* . College courses sometimes have students design and conduct mini-research studies; undergraduate majors might encourage students to do senior honors theses; graduate programs often require masters' theses and doctoral dissertations. Learning to collect and analyze data will help you address your own questions in an objective, systematic manner.

A final reason for learning statistics is that it may *help you in your future career* . The website Careercast.com conducts an annual survey in which they evaluate jobs on four dimensions: environmental factors, income, employment outlook, and stress. Recently, their #1 rated job was "data scientist," and the #2 rated job was "statistician." They noted, "A statistician's skill set can be used to break down and analyze large quantities of data. The demand for these skills spans a variety of industries, including marketing, banking, government, sports, retail, and even healthcare."

It's generally a good idea for students to maintain a healthy level of curiosity or even skepticism in regard to their education. However, we find that when it comes to learning statistics, students' feelings may be characterized as fear or anxiety. We understand these feelings, and we hope the benefits of learning statistics will become clear to you and help overcome any concerns you have.

1.3 Introduction to the Stages of the Research Process

Much of scientific research involves asking questions. Throughout this book, we'll examine a broad range of questions regarding human attitudes and behavior that contemporary researchers have asked and attempted to answer. The following are research questions we'll address in this chapter to introduce the stages of the research process:

> Does appealing to one's fear of missing out (FOMO) affect the intention to invest in cryptocurrencies?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Do pro-cannabis messages on social media lead to cannabis use?

Do parents differ in their awareness of their children's online activities?

Do nonbinary college students have the same experiences as cis men or cis women?

How might you try to answer questions such as these? You could base your answers on your personal beliefs, or you could adopt the answers given to you by others. But rather than relying on subjective beliefs and feelings, researchers in many academic disciplines test their ideas using the scientific method. The **scientific method** is a method of investigation that uses the objective and systematic collection and analysis of empirical data to test theories and hypotheses.

At its simplest, this book will portray the scientific method as consisting of five main steps or phases:

developing a research hypothesis to be tested,

collecting data,

analyzing the data,

drawing a conclusion regarding the research hypothesis, and

communicating the findings of the study.

Accomplishing each of these five steps requires completing a number of tasks, as shown in Figure 1.1 . Because this sequence of steps serves as the model for the research process we'll discuss throughout this book, each step is briefly introduced in the following sections. It's important to understand that the research process depicted in Figure 1.1 represents an ideal way of doing research. The "real" way, as you may discover in your own efforts, is often anything but a smooth ride but rather is filled with false starts, dead ends, and wrong turns.

Developing a Research Hypothesis to Be Tested

The initial stage—and the first step—of the research process is to develop a research hypothesis to be tested. A **research hypothesis** may be defined as a statement regarding an expected or predicted relationship between variables. A **variable** is a property or characteristic of an object, event, or person that can take on different values. One example of a variable is "U.S. state," a variable with 50 possible values (Alabama, Arkansas, etc.).

Research hypotheses are usually developed through the completion of several tasks:

identifying a question or issue to be examined,

reviewing and evaluating relevant theories and research, and

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

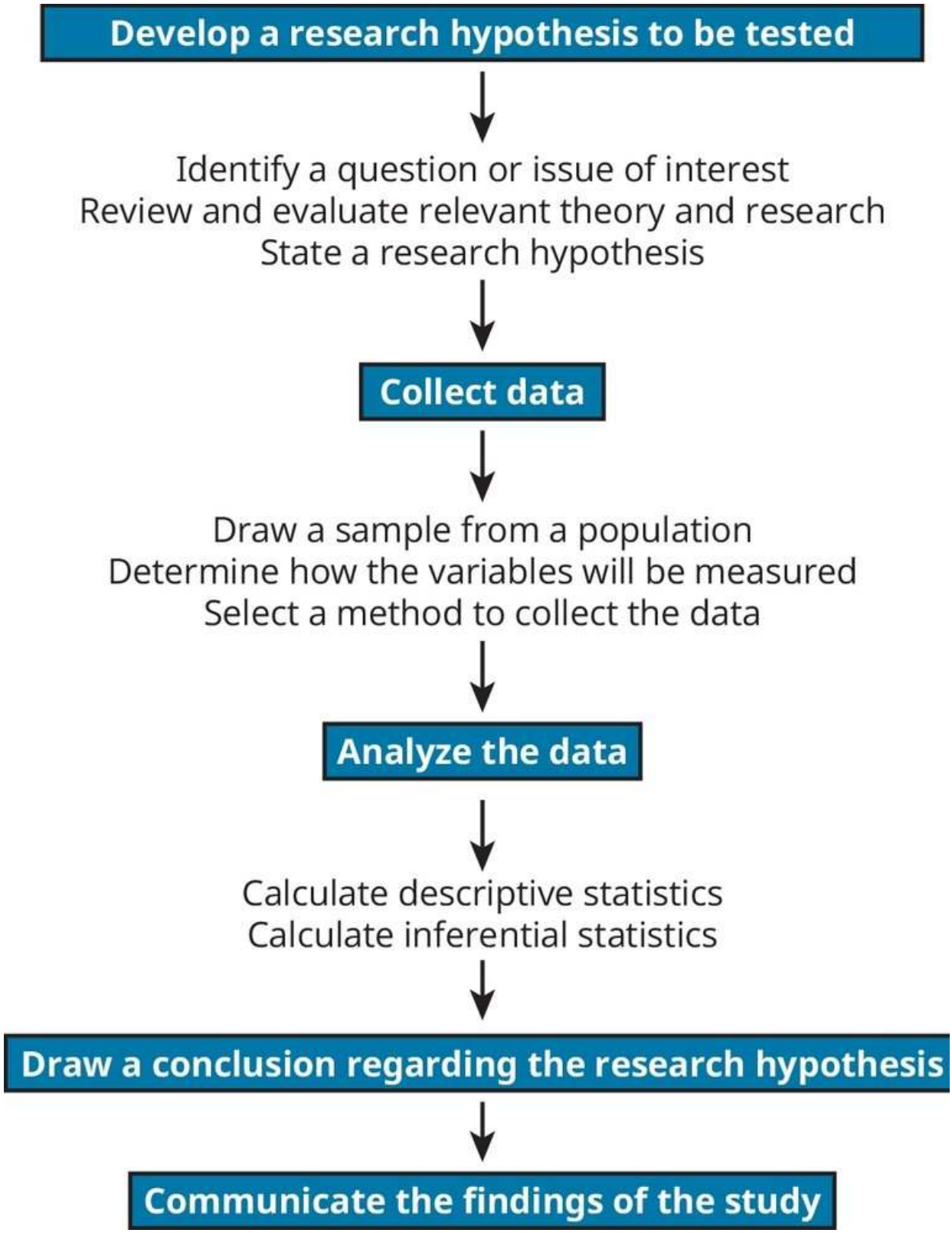
stating a research hypothesis.

Identifying a Question or Issue to Be Examined

Most research starts with a question posed by the researcher that often comes from the researcher's own ideas and daily observations. Although this may not seem terribly scientific, there's an advantage in using one's own experience as a starting point: People are generally much more motivated to explore issues that concern them personally. In teaching statistics, we advise students developing their own research projects to study something of interest to them. Conducting research can be tedious, difficult, and frustrating. At various points, you may ask yourself, "Why am I doing this?" Being able to provide a satisfactory answer to this question will help you overcome obstacles you encounter along the way.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

▶ Long Description: Figure 1.1 Steps in the Research Process Within the Scientific Method

Figure 1.1 Steps in the Research Process Within the Scientific Method

Reviewing and Evaluating Relevant Theories and Research

Beyond your own curiosity, research questions often arise from studying the theories, ideas, and research of others. A **theory** is a set of propositions used to describe or explain a phenomenon. The purpose of a theory is to summarize and explain specific facts using a logically consistent framework. Placing a question within a theoretical framework provides guidance and structure to research.

Reviewing and evaluating existing theories and research helps you decide whether it's worth the time and energy required to conduct a study. By seeing what others have done, you may conclude that the current way of thinking on a topic is incomplete or mistaken. By doing this review, you're able to ensure that the research you conduct adds to and improves upon an existing body of knowledge.

Stating a Research Hypothesis: Independent and Dependent Variables

Understanding and evaluating an existing literature not only helps develop a question of interest, but it may also lead to a predicted answer to that question. Within the scientific method, this answer is stated as a research hypothesis, defined earlier as a statement regarding an expected or predicted relationship between variables. Table 1.1 lists the research questions and research hypotheses for the studies listed at the beginning of this section. For example, the first research hypothesis states that "consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos."

One characteristic of research hypotheses such as those listed in Table 1.1 is that they identify the variables that are the focus of their research. As mentioned earlier, a variable is a property or characteristic with different values. Variables can be classified in several ways. In specifying a research hypothesis, researchers often speak in terms of "independent" and "dependent" variables. An **independent variable** may be defined as a variable manipulated or controlled by the researcher. A **dependent variable** , on the other hand, is a variable measured by the researcher that is expected to change or vary as a function of the independent variable. The independent variable is seen as the "cause" or "treatment," whereas the dependent variable is the "effect" or "outcome."

Researchers are interested in examining the effect of the independent variable on the dependent variable. Consider the first research hypothesis provided in Table 1.1 : "Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos." Here, the independent variable is the type of appeal consumers are exposed to: FOMO or non-FOMO. The dependent variable is the intention to invest in cryptos. In this study, the effect of the independent variable on the dependent variable is that

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

differences in the intention to invest in cryptos (the dependent variable) may “depend” upon which type of appeal (FOMO or non-FOMO) a consumer receives. Table 1.2 lists the independent and dependent variables for each of the research hypotheses in Table 1.1 .

In addition to identifying variables, a second characteristic of research hypotheses is that they specify the nature and direction of the relationship between variables. For example, the first research hypothesis in Table 1.1 states that “Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *higher* intention to invest in cryptos.” The word *higher* indicates the nature and direction of the relationship between the independent variable (type of appeal) and the dependent variable (intention to invest in cryptos). The direction of the relationship would not have been stated if the hypothesis had included the less specific phrase “will have a *different* intention,” which simply indicates that the two groups of consumers are not expected to be the same. Table 1.3 provides directional and nondirectional research hypotheses for the research studies from Table 1.1 .

Table 1.1 Examples of Research Questions and Hypotheses

Research Question	Research Hypothesis
Does appealing to one’s fear of missing out (FOMO) affect the intention to invest in cryptocurrencies? (Friederich et at., 2024)	Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.
Do pro-cannabis messages on social media lead to cannabis use? (Willoughby et al., 2024)	Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.
Do parents differ in their awareness of their children’s online activities? (Symons et al., 2017)	Mothers have more accurate knowledge of their children’s online activities than do fathers.
Do nonbinary college students have the same experiences as cis men or cis women? (Marx et al., 2024)	Nonbinary college students experience higher levels of victimization than do cis men or cis women.

The ability to form a directional research hypothesis is dependent on the state of the literature on the question of interest. If little or perhaps conflicting research has been conducted, researchers may not be able to form a directional hypothesis. For example, one study examining the relationship between exercise deprivation (not allowing people to exercise) and depression. reported, “We did not have a directional hypothesis, and when participants asked what we expected to find in this study, we replied: ‘We really are not sure since the results of earlier work on exercise deprivation are mixed’” (Mondin et al., 1996, p. 1200).

Table 1.2 Research Hypotheses and Their Independent and Dependent Variables

Research Hypothesis	Independent Variable	Dependent Variable
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher	Type of appeal	Intention to invest in cryptos

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Independent Variable	Dependent Variable
intention to invest in cryptos.		
Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.	Exposure to pro-cannabis messages	Intention to use cannabis
Mothers have more accurate knowledge of their children's online activities than do fathers.	Parent	Knowledge of children's online activities
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	Gender identity	Amount of victimization

Collecting Data

Once a research hypothesis has been formulated, researchers are ready to proceed to the second stage in the research process: collecting data. This step is seen as being composed of three tasks:

drawing a sample from a population,

determining how the variables will be measured, and

selecting a method by which to collect the data.

Table 1.3 Directional and Nondirectional Research Hypotheses

Directional Research Hypothesis	Nondirectional Research Hypothesis
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *higher* intention to invest in cryptos.	Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *different* intention to invest in cryptos.
Exposure to pro-cannabis messages on social media is related to *increased* intentions to use cannabis.	Exposure to pro-cannabis messages on social media is *related to* intentions to use cannabis.
Mothers have *more* accurate knowledge of their children's online activities than do fathers.	Mothers have *different* knowledge of their children's online activities than do fathers.
Nonbinary college students experience *higher* levels of victimization than do cis men or cis women.	Nonbinary college students experience *different* levels of victimization than do cis men or cis women.

Drawing a Sample From a Population

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The first step in collecting data is to identify the group of participants to which the research hypothesis applies. The group to which the results of a study may be applied or generalized is called a population. A **population** is the total number of possible units or elements that could potentially be included in a study. Although people often think the word "population" refers to everyone in the world, researchers typically define their populations more narrowly. For example, the population for the first research hypothesis in Table 1.1 could be "consumers," "U.S. consumers," "Georgia consumers," or "University of Georgia consumers." Researchers typically try to define their populations as broadly as possibly (e.g., "consumers" rather than "University of Georgia consumers") to maximize the applications or implications of their research.

Learning Check 1

Reviewing What You've Learned So Far



1. Review questions
    a. What are the five steps involved in the research process?
    b. Why is it useful to review and evaluate theories and research before conducting a study?
    c. What are the two main characteristics of research hypotheses?
    d. What is the difference between an independent variable and a dependent variable?
2. The following are several research hypotheses from published studies. For each research hypothesis, identify the independent and dependent variable.
    a. "College students will rate instructors who dress formally (i.e., business suit and tie) as having more expertise than instructors who dress casually (i.e., slacks and shirt)" (Sebastian & Bristow, 2008, p. 197).
    b. "It was expected tha. . . . greater amounts of television viewin. . . . would predict greate. . . . posttraumatic stress symptoms" (McLeish & Del Ben, 2008, p. 417).
    c. "Do borrowers who did not complete college have greater financial distress compared with . . . non-borrowers who did not complete college [or] borrowers who earned a degree?" (Jabbari et al., 2023, p. 6).
    d. "We expected that drivers would believe that texting while driving is more detrimental to their driving performance than browsing social media" (Hashash et al., 2019, p. 70).

It's extremely difficult to collect data from all members of a population. Imagine, for example, the time and money needed to collect information from every consumer in the United States! For this reason, researchers typically draw conclusions about populations based on information collected from a sample of the population. A **sample** is a subset or portion of a

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

population. Table 1.4 describes the samples used in the four studies introduced in Table 1.1 . As you can see from Table 1.4 , samples vary greatly in terms of their targeted population and size (the number of participants). The extent to which the results of a study can be generalized to the target population depends on the extent to which the sample is representative of the population.

Determining How Variables Will Be Measured: Levels of Measurement

The research hypotheses described in Table 1.1 involve variables such as type of appeal, intentions to use cannabis, and gender identity. To conduct a research study, the researcher must decide how to measure the variables stated in the research hypothesis. **Measurement** is the assignment of numbers or categories to objects or events according to rules.

For example, to measure the variable “height,” a researcher might use “number of inches from the ground in bare feet” as a form of measurement. To measure “success in college,” a student’s grade point average (GPA) may be obtained from school transcripts. “Self-esteem” might be measured by having people complete a questionnaire and be categorized as having either “low” or “high” self-esteem. As these examples demonstrate, different types of variables have different forms of measurement. Consequently, researchers have identified four distinct levels for measuring variables: nominal, ordinal, interval, and ratio.

The values of variables measured at the **nominal level of measurement** differ in category or type. The word *nominal* implies having to do with “names,” such that we use first and last names to distinguish between people. Gender identity is an example of a nominal variable, in that it consists of categories or types (such as cis male and cis female) rather than numeric values. In the first research hypothesis in Table 1.1 , the independent variable of “type of appeal” is measured at the nominal level with two categories: FOMO and non-FOMO.

Variables measured at the **ordinal level of measurement** have values that can be placed in an order relative to the other values. Rankings (such as finishing first, second, or third in a race), size (small, medium, large, or extra-large), and ratings (below average, average, or above average) are examples of ordinal variables. The values of ordinal variables indicate that one value represents relatively more or less of the variable than other values, but they don’t specify the precise difference between values. For example, although you can say that a “medium” drink is smaller than a “large” drink which is smaller than an “extra-large” drink, you can’t specify the exact difference in size between the three drinks.

Table 1.4 Research Hypotheses and Study Samples

Research Hypothesis	Study Sample
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.	“We recruited 150 participants from Amazon MTurk” (Friederich et at., 2024, p. 105).
Exposure to pro-cannabis messages on social media is	“Teens (N = 350) . . . needed to be between 13 and 17 years old and from Washington state” (Willoughby et

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Study Sample
related to increased intentions to use cannabis.	al., 2024, p. 185).
Mothers have more accurate knowledge of their children's online activities than do fathers.	"triadic data including reports from children aged 13 to 18, mothers and fathers . . . recruited in Flanders, the northern, Dutch speaking region of Belgium . . . a total of 357 valid triads was achieved" (Symons et al., 2017, p. 405).
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	"undergraduate college students aged 18–30 who were enrolled at 4-year institutions. This yielded a sample of 46,510 students" (Marx et al., 2024, p. 334).

The values of variables measured at the **interval level of measurement** are numbers equally spaced along a continuum. One example of an interval variable is the Fahrenheit scale of temperature. Here, a difference of 5 degrees has the same meaning anywhere along the scale; for example, the difference between 45°F and 50°F is the same as the difference between 65°F and 70°F. Many variables studied in the behavioral sciences (e.g., personality characteristics or attitudes) are considered to be measured at the interval level of measurement.

Variables at the **ratio level of measurement** are identical to interval variables, with one exception: Ratio scales possess what is known as a true zero point, for which the value of zero (0) represents the complete absence of the variable. Variables that describe a physical dimension (such as height, weight, distance, and time duration) typically have a true zero point. In the last research hypothesis in Table 1.1 , the variable "number of times victimized in last 12 months" is a ratio variable because it has a true zero point, where zero would indicate the complete absence of victimizations.

One advantage of ratio measurement versus interval is that ratio variables allow for a greater number of comparisons among values. Consider, for example, the value 6 for the ratio variable "inches." Not only is the difference between 4 and 6 inches the same as the difference between 6 and 8 inches (thereby involving addition and subtraction, which also can be used with interval variables), 6 inches is twice as long as 3 inches and half as long as 12 inches (involving multiplication and division). Multiplication and division comparisons cannot be made with interval variables. You cannot, for example, say that a temperature of 90°F is three times as much temperature as 30°F.

The boldfaced and italicized text in Table 1.5 illustrates how the four studies described in the earlier tables measured their independent and dependent variables. As you can see, the variables in these studies have been measured in a variety of ways at different levels of measurement.

Table 1.5 Research Hypotheses and Measurement of Variables

Research Hypothesis	Independent Variable	Dependent Variable

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Independent Variable	Dependent Variable
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.	Type of appeal **FOMO, non-FOMO (nominal)**	Intention to invest in cryptos **Likelihood of investing** **1-7 scale (1 = Highly unlikely to 7 = Highly likely) (ordinal)**
Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.	Exposure to pro-cannabis messages on social media **How often saw pro-cannabis messages (ratio)**	Intentions to use cannabis **Likelihood of using cannabis in the future (ratio)**
Mothers have more accurate knowledge of their children's online activities than do fathers.	Type of parent **Mother, Father (nominal)**	Accuracy of online activities **Accurate, somewhat accurate, inaccurate (nominal)**
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	Gender identity **Nonbinary, cis man, cis woman** **(nominal)**	Number of times victimized in last 12 months **(ratio)**

What difference does the level of measurement make?

A variable's level of measurement has important implications for researchers in that it influences how research hypotheses are stated and how data are analyzed. For example, imagine you study the variable "success in college." You could measure this variable at the *nominal* level of measurement by classifying students into one of two groups: successful or unsuccessful. Researchers typically use nominal variables when they are interested in studying differences between groups, such as, "Do successful and unsuccessful students differ in their study habits?"

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

You could instead measure success in college as an *ordinal* variable by ranking students from top to bottom, which enables a researcher to study relative differences with less concern for the precise size of these differences. For example, using ranked data, you could ask, “How similar are younger and older faculty members’ rankings of their students?”

To study success in college at the *interval* level, you could rate each student on a scale from 1 (low) to 100 (high). You could then use these ratings to address a research question such as, “Is there a relationship between faculty ratings of their students and students’ ratings of themselves?”

Finally, to measure success in college using the *ratio* level of measurement, you could use students’ salary after graduation (in dollars) as an indicator of success. Salary is a ratio variable because it contains a true zero point. In practice, ratio variables can be used to ask many of the same questions as those involving interval variables. For example, you could ask, “Is there a relationship between faculty members’ ratings of students and students’ salaries after graduation?”

As you see, how researchers measure their variables influences how they state their research questions and hypotheses. The level of measurement has another important implication: It helps determine the statistics researchers use to analyze their data. Certain statistics can be applied to variables measured at the interval or ratio levels of measurement, whereas others are appropriate for nominal or ordinal variables. As we move from chapter to chapter of this book, the level of measurement of the variables being analyzed will be noted.

Selecting a Method to Collect the Data: Experimental and Nonexperimental Research Methods

In addition to drawing a sample from a population and determining how the variables will be measured, the third step in collecting data is to determine the type of research method to use to collect the data on these variables. Research methods may be classified into two main types:

experimental research methods, and

nonexperimental research methods.

Experimental research methods.

Experimental research methods are methods designed to test causal relationships between variables—more specifically, whether changes in independent variables produce or cause changes in dependent variables. To make inferences about cause-effect relationships, researchers conducting an experiment must first eliminate other possible causes or explanations for changes in the dependent variable besides the independent variable. If it can be claimed that a variable other than the independent variable created the changes in the dependent variable, the research results are said to be confounded. A **confounding variable** is a variable related to an independent variable that provides an alternative explanation for the relationship between the independent and dependent variables.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

To illustrate the concept of confounding variables, consider the following question: Is there a relationship between a mother's ethnicity and the birth weight of her baby? Although research has shown differences in the birth weights of babies of different ethnicities, one study noted that some of this research did not eliminate the possibility that these differences may be partly due to differences between ethnic groups on factors associated with low birth weights, such as the mother's average age at the time of pregnancy and behaviors such as smoking and drinking (Ma, 2008). These other factors are considered confounding variables in that they provide alternative explanations for the relationship between ethnicity and babies' birth weights. Confounding variables make it difficult for researchers to draw firm conclusions regarding the relationship between the independent and dependent variable.

Learning Check 2

Reviewing What You've Learned So Far



1. Review questions
    a. What is the difference between a population and a sample?
    b. Within the research process, what is the relationship between populations and samples?
    c. What are the four levels of measurement? How do they differ?
2. For each of the following variables, name the scale of measurement (nominal, ordinal, interval, or ratio).
    a. Type of school attended (public, private)
    b. Probability of graduating college in 4 years (0% to 100%)
    c. Rating of television (good, better, best)
    d. Number of computers in home

Researchers minimize the influence of confounding variables in two main ways. First, researchers exercise experimental control by making the research setting (i.e., characteristics of the research participants, the instruments or measures administered, the methods used to collect data) the same for all participants. In the birth weight study, for example, the research design might control for the effect of parental smoking on birth weight by only including nonsmokers in the study sample, thereby excluding those who smoke.

Another way researchers exert control over the research setting is by including a condition known as a **control group** , a group of participants in an experiment not exposed to the independent variable. For example, if you conducted an experiment examining the effects of caffeine on one's health, one group (the experimental group) might be instructed to drink coffee while a second group (the control group) drinks decaffeinated coffee. By comparing the results of the two groups, researchers can then assess the impact of caffeinated coffee on the health of those who drink it.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Because researchers cannot possibly identify and control for all potential confounding variables, a second strategy used to minimize the influence of confounding variables is **random assignment** , which is assigning participants to each category of an independent variable in such a way that each participant has an equal chance of being assigned to each category. For example, a researcher could randomly assign a participant to the experimental or control condition by flipping a coin: heads for experimental, tails for control. The purpose of random assignment is to equalize or neutralize the effects of confounding variables by distributing them equally over the levels of the independent variable.

Nonexperimental (correlational) research methods.

Experimental research designs are one of the best tools researchers have for making causal inferences. However, the ability to make causal inferences requires a great deal of control that may not always be possible or desirable. For this reason, researchers often employ **nonexperimental research methods** , which are research methods designed to measure relationships between variables without having the ability to infer cause-effect relationships. These research methods are sometimes referred to as *correlational* research methods. It is important to understand that correlational research methods are different from correlational statistics we'll introduce in Chapter 13. Correlational research methods pertain to how data are *collected* ; correlational statistics pertain to how data are *analyzed* . Some of the most common types of nonexperimental research designs include quasi-experiments, survey research, observational research, and archival research.

Quasi-experimental research compares naturally formed or preexisting groups rather than randomly assigning participants to conditions. One example of this type of research was a study examining the effects of the COVID-19 pandemic on three groups of students: those living on campus, those living off campus but within commuting distance to school, and those living far away in another city, state, or country (Bisconer & McGill, 2023). This study is a quasi-experiment because the students were not randomly assigned to one of the three living conditions before collecting data.

Survey research methods obtain information directly from a group of people regarding their opinions, beliefs, or behavior. The goal of survey research, which can include using questionnaires and interviews, is to obtain information from a sample to represent or estimate the views of a larger population. One example of survey research examined college students' mental health by having them fill out a survey in which they reported their spirituality (the desire for truth, meaning, and significance of one's life) and their self-acceptance (knowledge and acceptance of one's self and limitations) (Nonis et al., 2024). Survey research is nonexperimental in that the variables being measured are not manipulated by the researcher.

Observational research is the systematic and objective observation of naturally occurring behavior or events. Observational research is often used to study phenomena that the researcher cannot or should not deliberately manipulate. For example, researchers studying interpersonal conflict among children would never force children to fight each other. One study resolved this issue by videotaping groups of adolescents discussing topics such as competition and peer pressure and then counting the number of times the adolescents teased or mocked each other (Connolly et al., 2015).

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Rather than observing or measuring behavior directly, **archival research** is the use of archives (records or documents of the activities of individuals, groups, or organizations) to examine research questions or hypotheses. One example of an archival research study studied the question of whether men are perceived to be more creative than women (Proudfoot et al., 2015). In this study, the 100 most viewed talks on TED.com were identified, and the researchers compared viewers' posted ratings of the talks given by men versus women.

Labeling variables in nonexperimental research.

One issue that often confuses students when learning about nonexperimental research is the practice of referring to the variables as the "independent variable" and the "dependent variable." For example, in the study described previously that involved biological sex and TED.com ratings, biological sex may be called the "independent variable." However, because biological sex isn't manipulated by the researcher, it would be inappropriate to say that biological sex "causes" differences in TED.com ratings. To avoid confusion, researchers sometimes refer to the independent variable and the dependent variable in a nonexperimental study as the **predictor variable** and **criterion variable**, respectively.

Analyzing the Data

Once data collection has been completed, the next step in the research process involves data analysis. This part of the research process addresses the primary topic of this book: statistics. Because this is the first chapter of this book, we won't describe analyzing data in detail. Instead, we introduce the notion that there are two main purposes of analyzing data: (1) to organize, summarize, and describe data that has been collected and (2) to test and draw conclusions about ideas and hypotheses. These two purposes are met by calculating two main types of statistics: descriptive statistics and inferential statistics.

Calculating Descriptive Statistics

Descriptive statistics are statistics used to summarize and describe a set of data for a variable. For example, you've probably heard of statistics used to describe or summarize aspects of our society, such as crime statistics and unemployment statistics. Using a research-related example, Caitlin Abar, a researcher at Brown University, conducted a study examining increases in students' alcohol-related behaviors after entering college (Abar, 2012). To measure students' level of alcohol use, she created a variable called "typical weekend drinking," which "was measured as the sum of drinks consumed on a typical Friday and Saturday within the past 30 days" (p. 22). One way to summarize students' responses to the typical weekend drinking variable (0, 1, 2, etc.) would be to calculate the mean, which is the mathematical average of a set of scores. The mean, described in Chapter 3, is an example of a descriptive statistic. The first part of this book describes a variety of descriptive statistics used by researchers to summarize data they've collected.

Calculating Inferential Statistics

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Besides summarizing and describing data, a second purpose of statistics is to test hypotheses. **Inferential statistics** are statistical procedures used to test hypotheses and draw conclusions from data collected during research studies. Using inferential statistics, researchers make inferences about what's believed to exist in populations based on data gathered from samples of the population.

As an example of an inferential statistic, let's return to the example of students' alcohol-related behaviors (Abar, 2012). This study was interested in seeing whether these behaviors were related to aspects of these students' relationships with their parents. One aspect was "alcohol communications," defined as "how often they discussed alcohol related topics with their parents" (p. 22). To test a hypothesis regarding the relationship between students' alcohol use and their alcohol communications, we could use an inferential statistic known as the Pearson correlation coefficient, which will be discussed in Chapter 13. The chapters in the last half of this book discuss a broad variety of inferential statistics, along with detailed instructions for analyzing and drawing conclusions from statistical data and presenting research findings.

Drawing a Conclusion Regarding the Research Hypothesis

Once statistical analyses have been completed, the next step is to interpret the results of the analyses as they relate to the study's research hypothesis. More specifically, do the findings of the analyses support or not support the research hypothesis? The word *support* in the previous sentence is very important. Students and researchers are sometimes tempted to say their findings "prove" their hypotheses are either "true" or "false." However, as we'll discuss in Chapter 5, because researchers typically don't collect data from the entire population, they can't know with 100% certainty whether their hypothesis has been proven to be true or false. A *huge* part of this book is learning how to draw appropriate conclusions from statistical analyses.

Communicating the Findings of the Study

Conducting research requires a variety of different skills: conceptual skills to develop research hypotheses, methodological skills to collect data, and mathematical skills to analyze these data. Another integral part of the research process is the communication skills needed to inform others about a study. Researchers must be not only skilled scientists but also effective writers.

Learning Check 3

Reviewing What You've Learned So Far



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1. Review questions
    a. What are the main differences between experimental and nonexperimental research methods?
    b. What are confounding variables? What do researchers do to minimize the effects of confounding variables?
    c. What are the main types of nonexperimental research methods?
    d. What are the relative strengths and weaknesses of experimental and nonexperimental research?
    e. What are the main purposes of descriptive and inferential statistics?

For many years, the American Psychological Association (APA), the professional association for psychologists, has provided researchers guidance on how to communicate the results of their research. In 1929, the APA published a seven-page article entitled, “Instructions in Regard to Preparation of Manuscript.” In contrast, the seventh edition of the *Publication Manual of the American Psychological Association* , published in 2020, consists of 428 pages! This increase in length highlights the challenge of communicating research effectively and appropriately.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1.4 Plan of the Book

The primary purpose of this chapter was to introduce statistics and place it within the larger research process. The remainder of this book discusses different statistics used by researchers to examine various questions of interest. Although it's critical for you to learn to correctly calculate statistics, it is equally important to understand the role of statistics within the research process and appreciate issues related to the use (and sometimes misuse) of statistics.

The first half of the book (Chapters 2 through 6) introduces conceptual and mathematical issues that are the foundation of statistical analyses. Chapters 2, 3, and 4 discuss how data may be examined, described, summarized, and presented in numeric, visual, and graphic form. Descriptive statistics are introduced and discussed in these chapters. Chapters 5 and 6 introduce characteristics of the inferential statistics used to test research hypotheses. The chapters in the second half of the book cover a number of different inferential statistics used with different types of variables.

In most cases, the chapters in this book share a uniform format and structure and are centered on the research process, using examples of actual published research studies that address questions with which you may be familiar. For each example, we'll clearly state the research hypothesis, briefly describe how the data in the study were collected, illustrate the calculation and interpretation of descriptive and inferential statistics, discuss the extent to which the statistical analyses support the research hypothesis, and show how to communicate one's research to a broader audience.

1.5 Looking Ahead

We began this chapter by defining statistics and providing reasons why learning about them may be of value to you. Next, we placed statistics within the stages of the larger research process, a process that will be the foundation of this textbook. As we've explained, the research process is centered on a research hypothesis, which is a predicted relationship between variables. Once a hypothesis has been stated, the next step is to collect data to be analyzed. However, before conducting statistical analyses on a set of data, there is preliminary work that must be completed. The next chapter will focus on methods used by researchers to examine data.

1.6 Summary

Statistics may be defined as a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data. As such, statistics not only is concerned about how data are analyzed but also recognizes the importance of understanding how data are collected and how the results of analyses are interpreted and communicated. There are a variety of reasons why students should learn statistics: Statistics are encountered in a wide variety of daily activities, students may be asked or required to read and interpret the results of statistical analyses in their courses, students may use statistics in conducting their own research, and statistics may help one's career.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Researchers conduct research using the **scientific method** of inquiry, a method of investigation that uses the objective and systematic collection and analysis of empirical data to test theories and hypotheses. The research process used within the scientific method of inquiry consists of five main steps: developing a research hypothesis to be tested, collecting data, analyzing the data, drawing a conclusion regarding the research hypothesis, and communicating the findings to the study.

A **research hypothesis** is a statement regarding an expected or predicted relationship between variables. Developing a research hypothesis involves identifying a question or issue to be examined, reviewing and evaluating relevant theories and research, and stating the research hypothesis to be tested in the study. A **theory** is a set of propositions used to describe or explain a phenomenon. A research hypothesis contains **variables** , which are properties or characteristics of some object, event, or person that can take on different values. More specifically, a research hypothesis states the nature and direction of a proposed relationship between an **independent variable** (a variable manipulated or controlled by the researcher) and a **dependent variable** (a variable measured by the researcher that is expected to change or vary as a function of the independent variable).

Collecting data involves drawing a sample from a population, determining how the variables will be measured, and selecting a method to collect the data. A **population** is the total number of possible units or elements that could potentially be included in a study; a **sample** is a subset or portion of a population. Variables can be measured at one of four levels of measurement: **nominal** (values differing in category or type), **ordinal** (values placed in an order relative to the other values), **interval** (values equally spaced along a numeric continuum), or **ratio** (values equally spaced along a numeric continuum with a true zero point).

There are two main types of methods used to collect data: **experimental research methods** , designed to test whether changes in independent variables produce or cause changes in dependent variables, and **nonexperimental (correlational) research methods** , designed to examine the relationship between variables without the ability to infer cause-effect relationships.

In experimental research, researchers are concerned about possible **confounding variables** , which are variables related to independent variables that provide an alternative explanation for the relationship between independent and dependent variables. To minimize the impact of confounding variables, a research study may exercise experimental control over various aspects of the situation, including the use of a **control group** , which is a group of participants not exposed to the independent variable the research is designed to test, or **random assignment** , which involves assigning participants to each category of an independent variable in such a way that each participant has an equal chance of being assigned to each category.

Examples of nonexperimental research methods are **survey research** , in which information is directly obtained from a group of people regarding their opinions, beliefs, or behavior; **observational research** , which involves the systematic and objective observation of naturally occurring events; and **archival research** , which uses archives (records or documents of the activities of individuals, groups, or organizations) to examine research questions or hypotheses. In nonexperimental research, the independent variable and

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

dependent variable may be referred to as the **predictor variable** and **criterion variable**, respectively.

Once data have been collected, the next step is to analyze them using two main types of statistics: **descriptive statistics**, which summarize and describe a set of data, and **inferential statistics**, which are statistical techniques used to test hypotheses and draw conclusions from data collected during research studies.

Once the statistical analyses have been completed, the results of the analyses are interpreted regarding whether they support or not support the study's research hypothesis. The final step in the research process is to communicate the study to a broader audience.

1.7 Important Terms

archival research

confounding variable

control group

criterion variable

dependent variable

descriptive statistics

experimental research methods

independent variable

inferential statistics

interval level of measurement

level of measurement (nominal, ordinal, interval, ratio)

measurement

nonexperimental research methods

observational research

population

predictor variable

quasi-experimental research

random assignment

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

research hypothesis

sample

scientific method

statistics

survey research

theory

variable

1.8 Exercises

1. Identify the independent and dependent variable for each of the following research questions and hypotheses from published articles.
    a. “We hypothesize that finance majors will embody narcissistic traits more than other business majors” (Wood et al., 2021, p. 924).
    b. “Men are expected to employ physical characteristics (particularly those that are directly related to sex) more often than women in selecting dating candidates” (Hetsroni, 2000, p. 91).
    c. “It is hypothesized that perceptions of emotional pain and support needs will be higher for the community of the Black victim than those of the officers’” (Kahn et al., 2024).
    d. “Gen Z students are more likely to believe that ‘Group projects bring out the worst in people’ than Millennials” (Schlee et al., 2020, p. 145).
    e. “We hypothesized to find . . . higher stress and anxiety post COVID in the Portuguese group in contrast with the British ” (Penalba-Sanchez et al., 2022, p. 322).
    f. “We hypothesized that consumers would think that a ‘sale price’ presentation would generate a greater monetary savings than an ‘everyday low price’ presentation” (Tom & Ruiz, 1997, p. 403).
2. The following are a number of research questions and hypotheses from actual published articles. For each hypothesis, identify the independent and dependent variable.
    a. “The use of color in a Yellow Pages advertisement will increase the perception of quality of the products for a particular business when compared with noncolor advertisements” (Lohse & Rosen, 2001, p. 75).
    b. “We hypothesized that parents who use more frequent corporal and verbal punishmen. . . . will report more problem behaviors in their children” (Brenner & Fox, 1998, p. 252).
    c. “It was hypothesized that adolescents with anorexia nervosa woul. . . . be more respectful when compared with peers with bulimia nervosa” (Pryor & Wiederman, 1998, p. 292).
    d. “The purpose of the present research was to assess brand name recognition as a function of humor in advertisements. . . . It was predicted that participants would recognize product brand names that had been presented with humorous

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

advertisements more often than brand names presented with nonhumorous advertisements" (Berg & Lippman, 2001, p. 197).

e. "The purpose of this study was to explore the relation between time spent in daycare and the quality of exploratory behaviors in 9-month-old infant. . . . it was hypothesized tha. . . . infants who spent greater amounts of time in center-based care would demonstrate more advanced exploratory behaviors than infants who did not spend as much time in center-based care" (Schuetze et al., 1999, p. 269).

f. "It was hypothesized that students would score higher on test items in which the narrative contains topics and elements that resonate with their daily experiences versus test items comprised of material that is unfamiliar" (Erdodi, 2012, p. 172).

3. Identify the independent and dependent variable for each of the following research questions and hypotheses from published articles.
    a. "It was hypothesized that the number of selfies posted . . . would be positively correlated with dimensions of narcissism" (Barry et al., 2017, p. 51).
    b. "We hypothesized that preschoolers' self-reported preferences for aggressive media . . . would be positively related to parent reports of children's aggressive behaviors" (Jamnik & DiLalla, 2018, p. 159).
    c. "We hypothesized that helicopter parenting would lead to higher levels of depression" (Darlow et al., 2017, p. 3).
    d. "We predicted that weight-related shame would be associated with greater binge eating" (Craven & Fekete, 2019, p. 45).
    e. "We hypothesize that men with strong gender identities will be more sensitive to manipulated threats to their masculinity compared with men with gender identities that are less central to their identity" (Harrison & Michelson, 2019, p. 65).
    f. "After exposure to fake news story with user comments identifying the news as fake, people are less likely to share the fake news than after exposure to the fake news story with user comments supportive of the content" (Colliander, 2019, p. 204).
4. Name the scale of measurement (nominal, ordinal, interval, ratio) for each of the following variables:
    a. The amount of time needed to react to a sound
    b. Gender
    c. Score on the Scholastic Aptitude Test (SAT)
    d. Political orientation (not at all conservative, conservative, very conservative)
    e. Political affiliation (Democrat, Republican, Independent)
5. Name the scale of measurement (nominal, ordinal, interval, ratio) for each of the following variables:
    a. One's age (in years)
    b. Size of soft drink (small, medium, large, extra large)
    c. Voting behavior (in favor vs. against)
    d. IQ score
    e. Parent (mother vs. father)
6. A faculty member wishes to assess the relationship between students' scores on the SAT and their performance in college.
    a. What is a possible research hypothesis in this situation?
    b. What are the independent and dependent variables?
    c. How could you measure the variable "performance in college" at each of the four levels of measurement?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

7. A researcher hypothesizes that drivers who use cellular phones will get into a greater number of traffic accidents than drivers who do not use these phones. For this example,
    a. What are the independent and dependent variables?
    b. How could you measure the dependent variable?
    c. How could you conduct this study using an experimental research method? How could you conduct this study using a nonexperimental method?

Answers to Learning Checks

Learning Check 1

1.

    a. IV: Instructor's dress; DV: Expertise

    b. IV: Amount of television viewing; DV: Posttraumatic stress symptoms

    c. IV: Student group; DV: Financial distress

    d. IV: Activity while driving; DV: Driving performance

Learning Check 2

2.

    a. Nominal

    b. Ratio

    c. Ordinal

    d. Ratio

Answers to Odd-Numbered Exercises

1.

    a. IV: Type of business major; DV: Narcissistic traits

    b. IV: Biological sex; DV: Employment of physical characteristics

    c. IV: Community; DV: Perceptions of emotional pain and support needs

    d. IV: Generation; DV: Beliefs about group projects

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

e. IV: Country; DV: Stress and anxiety

f. IV: Price presentation; DV: Perception of savings

3.

a. IV: Number of selfies posted; DV: Narcissism

b. IV: Preference for aggressive media; DV: Parent reports of children's aggressive behaviors

c. IV: Helicopter parenting; DV: Depression

d. IV: Weight-related shame; DV: Binge eating

e. IV: Strength of gender identity; DV: Sensitivity to threat to masculinity

f. IV: Type of user comments to news story; DV: Sharing of news story

5.

a. Ratio

b. Ordinal

c. Nominal

d. Interval

e. Nominal

7.

a. IV: Cellular phone use; DV: Number of traffic accidents

b. Example: Determine the number of traffic accidents each person has experienced in the past year.

c. Example of experimental: Use a driving simulation with the "driver" talking on a phone and measure the number of potential driving mistakes or accidents.

d. Example of nonexperimental: Take a survey of the number of accidents people have been in and whether or not they were using a phone during the accident.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CHAPTER 2 EXAMINING DATA: TABLES AND FIGURES

Learning Objectives

2.1 Describe the main purpose and the research hypothesis of the “lottery” study.

2.2 Describe different reasons for examining data.

2.3 Describe the steps used to create a frequency distribution table.

2.4 Describe the steps used to create a grouped frequency distribution table.

2.5 Explain the differences between bar charts, pie charts, histograms, and frequency polygons.

2.6 Describe the three aspects and types of distributions of data.

Chapter 1 described a process by which research may be conducted, which begins by stating a research hypothesis. The next step involves collecting data that will ultimately be used in statistical analyses designed to test the research hypothesis. However, researchers don’t conduct analyses until they’ve first examined the data to make sure appropriate conclusions can be drawn from the analyses. This chapter describes *why* researchers examine data and *how* data may be examined using tables, charts, and graphs. As with many of the chapters in this book, this discussion will be guided by the findings from a published research study.

2.1 An Example From the Research: Winning the Lottery

A friend of yours flips a coin and asks you to guess whether it’s “heads” or “tails.” You guess “tails,” but the answer is “heads.” Your friend flips the coin a second time, and again you predict “tails.” But again the answer is “heads.” On the third coin flip, you guess “tails” once more—but, once again, you’re wrong. As your friend prepares to flip the coin a fourth time, you tell yourself there’s little chance the coin could be “heads” again. But the reality is that the fourth coin flip is equally likely to be “heads” or “tails” because the outcome of the first three coin flips has absolutely no impact on the fourth flip. In other words, the outcome of a coin flip is a random event that can’t be controlled or predetermined. But do you really believe this?

The impact of random events can take on greater significance than the outcome of a coin flip. For example, government-run lotteries provide the opportunity to win huge amounts of money through the random selection of numbers. In 2001, Dr. Karen Hardoon, a researcher at McGill University in Canada, studied the thought processes people use when purchasing lottery tickets (Hardoon et al., 2001). As previous research had found gamblers believe they can control random events such as the rolling of dice or the spinning of a roulette wheel, Dr. Hardoon and her colleagues were interested in

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

determining “whether individuals with gambling problems perceive the purchase of lottery tickets in a similar manner as non-problem gamblers” (p. 752). On the basis of earlier findings, Dr. Hardoon hypothesized that problem gamblers are less likely to believe lottery outcomes are random than are nonproblem gamblers.

To test their research hypothesis, Dr. Hardoon and her associates located a group of gamblers and asked whether they had ever experienced problems as a result of their gambling. Each participant was classified into one of two groups: problem gamblers and nonproblem gamblers.

The participants in the study, which we’ll refer to as the lottery study, were shown four lottery tickets and asked, “If you were to buy a ticket to play in the lottery, which one would you select?” Each ticket contained six numbers selected from among the numbers 1 through 49; the four tickets were given a particular label. The *sequence* ticket had numbers in consecutive order (30-31-32-33-34-35); numbers on the *pattern* ticket increased by 5s (5-10-15-20-25-30); the *nonequilibrated* numbers (35-37-40-43-44-49) were clustered at the upper end of the 49 numbers, and the *random* ticket (7-8-23-34-36-42) shared none of the characteristics of the other three tickets. We’ll call the variable in this study “lottery ticket,” a variable with four categories: sequence, pattern, nonequilibrated, and random.

The researchers recorded which of the four tickets each participant chose. Table 2.1 lists the data for the lottery ticket variable for the nonproblem gamblers (the data for the problem gamblers are provided in the exercises at the end of this chapter). Looking at this table, we see that the nonproblem gamblers differed in their lottery ticket choices; however, it’s difficult to describe these differences given the unorganized nature of the collected data. Consequently, before conducting statistical analyses on their data, researchers typically organize and examine them. In the following sections, we will discuss why and how data are organized and examined.

2.2 Why Examine Data?

There are a variety of reasons why researchers examine data they’ve collected before analyzing the data. These reasons include the following:

To gain an initial sense of the data

To identify outliers

To evaluate research methodology

To determine whether data meet statistical criteria and assumptions

Each of these reasons is discussed in the following sections.

Table 2.1 The Lottery Ticket Choices of 22 Nonproblem Gamblers

Participant	Lottery Ticket
1	**Random**
2	**Random**
3	**Nonequilibrated**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Participant	Lottery Ticket
4	**Pattern**
5	**Random**
6	**Sequence**
7	**Random**
8	**Random**
9	**Pattern**
10	**Pattern**
11	**Random**
12	**Pattern**
13	**Random**
14	**Nonequilibrated**
15	**Random**
16	**Pattern**
17	**Nonequilibrated**
18	**Random**
19	**Random**
20	**Pattern**
21	**Random**
22	**Random**

Gaining an Initial Sense of the Data

Because researchers spend a great amount of time and energy designing their research studies, they're eager to analyze the data they've collected. Students taking courses in statistics, on the other hand, face a different challenge: They're given sets of data with which they have no prior experience. Although both researchers and students may be motivated to analyze their data as quickly as possible, it's important to first examine the data in order to gain an initial sense of them.

As an example, imagine you conduct a study regarding the fuel efficiency of automobiles and collect data on the miles per gallon (MPG) of different types of cars. You might expect the scores in your data to range from 10 to 50 MPG, with the majority between 25 and 35 MPG. Examining data helps researchers gain an initial sense of whether the data are in line with their expectations. Examining data also provides a way of confirming the results of later statistical analyses. We've found that students' errors in their calculations could have been avoided had they first looked at the set of data. For example, examining data will help you avoid making statements such as, "The average MPG in my sample was 127 MPG."

Identifying Outliers

Researchers also examine their data to identify **outliers** , defined as rare, extreme scores that lie outside of the range of the majority of scores in a set of data. Imagine, for example, that a group of 15 students is asked to take a 20-item test of mathematical calculations; each student's score is the number of test items answered correctly. In scoring the tests, a researcher finds that 14 students

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

correctly answered between 8 and 17 questions. However, one student provided the wrong answer to every question for a score of 0. In this case, the score of 0 would be considered an outlier.

If not identified, outliers may distort conclusions researchers draw about their sample and their data, particularly when the sample is relatively small. In the previous example, the single score of 0 might lead an observer to conclude the class as a whole performed more poorly on the test than they actually did.

Evaluating Research Methodology

Examining data is also useful in assessing the effectiveness of the methods used to collect the data. For the 20-item math test, for example, it might be reasonable to expect the majority of scores to fall in the middle of the 0 to 20 range. But what if every student answered at least 17 of the 20 questions correctly? You might conclude the test was too easy and that more difficult questions are needed to accurately assess students' math skills. By comparing the collected data for a variable with the range of possible values, researchers are able to evaluate and modify their measurement tools.

Determining Whether Data Meet Statistical Criteria and Assumptions

Researchers also examine their data to see whether they meet certain statistical criteria and assumptions. Many of the statistics in this book are based on assumptions regarding the shape and nature of a set of data that has been collected. For example, these procedures assume that in most data sets, the majority of scores will fall in the middle of the range of possible values, with a small amount of scores at the highest and lowest possible values (you may have heard this referred to as "the bell curve"). In the example of the 20-item math test, it might be assumed that most scores will be between 7 and 13, with smaller number of scores less than 7 or greater than 13. When assumptions such as this are not met, it's possible to draw inappropriate conclusions from statistical analyses that are conducted.

2.3 Examining Data Using Tables

One part of examining data that have been collected for a variable involves answering a series of questions:

> What are the values of the variable?
>
> How many participants in the sample have each value of the variable?
>
> What percentage of the sample has each value of the variable?

As it may be difficult to answer questions such as these just by looking at the data in Table 2.1 , researchers typically organize their data by creating a table known as a **frequency distribution table** , a table that summarizes the number and percentage of participants for the different values of a variable. This section discusses the steps involved in creating frequency distribution tables using these questions.

What Are the Values of the Variable?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The first step in organizing the data for a variable is to identify all of the possible values for the variable. Table 2.2 begins the construction of a frequency distribution table by identifying the four values for the lottery ticket variable: sequence, pattern, nonequilibrated, and random.

How Many Participants in the Sample Have Each Value of the Variable?

Once the values of the variable have been identified, the next step in organizing a set of data is to determine the **frequency** of each value, which is the number of participants in a sample corresponding to a value of a variable. These frequencies help indicate the nature and shape of the distribution of scores for a variable in a sample.

Table 2.2 Identifying the Values of the Lottery Ticket Variable

Number	Lottery Ticket
1	Sequence
2	Pattern
3	Nonequilibrated
4	Random

Table 2.3 Using the Tick Mark Method to Determine the Frequencies of the Lottery Ticket Variable

Lottery Ticket	Tick Marks	*f*
Sequence	|	1
Pattern	||||— |	6
Nonequilibrated	|||	3
Random	||||— ||||— ||	12
Total	||||— ||||— ||||—||||—||	22

One way to determine the frequencies for a variable is to sort the data into the different values and then count the number of participants having each value using a “tick mark method,” where a tick mark is placed next to a value of the variable each time that value appears in the sample. As shown in Table 2.3 , the tick marks are then counted to determine the frequency for each value. For example, the number of tick marks and therefore the frequency (f) for the sequence ticket is 1 because it was picked by one participant (Participant 6 in Table 2.1). Similarly, the tick marks for each of the other three types of tickets can be recorded and counted until all of the data in the sample are accounted for.

What Percentage of the Sample Has Each Value of the Variable?

Although it’s important to determine the frequency for each value of a variable, these frequencies by themselves are not always particularly informative. For example, how you interpret “There are 35 Democrats in my sample” is different when the sample consists of 40 voters rather than 400 voters.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Because the interpretation of frequencies depends on the size of the sample, it's useful to calculate the *percentage* of the sample having each value of a variable. These percentages can be calculated using the following formula:

$$\% = \frac{f}{total\,number\,of\,scores} * 100$$

where *f* is the frequency for a value of a variable.

Looking at Table 2.3 , 1 of the 22 nonproblem gamblers chose the sequence ticket. Therefore, the percentage of the sample choosing this ticket is

$$\begin{aligned} \% &= \frac{f}{total\,number\,of\,scores} * 100 \\ &= \frac{1}{22} * 100 = .05 * 100 \\ &= 5\% \end{aligned}$$

After calculating the percentage for each of the four tickets, Table 2.4 provides the frequency distribution table for the lottery ticket variable.

Once a frequency distribution table has been constructed for a variable, researchers can begin to draw conclusions about their sample. For the lottery ticket variable, the random ticket had the highest frequency and percentage (*f* = 12, 54%) compared to the pattern (*f* = 6, 27%), nonequilibrated (*f* = 3, 14%), and sequence (*f* = 1, 5%) tickets. As a result, the researchers in the lottery study made the following observation:

"The results of the present study indicated that, for the entire sample . . . the greatest percentage of tickets selected by the entire sampl. . . . were random tickets." (Hardoon et al., 2001, p. 760)

Table 2.4 Frequency Distribution Table for the Lottery Ticket Variable

Lottery Ticket	***f***	**%**
Sequence	1	**5%**
Pattern	6	**27%**
Nonequilibrated	3	**14%**
Random	12	**54%**
Total	22	**100%**

Learning Check 1

Reviewing What You've Learned So Far

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



1. Review questions
    a. What are some reasons for examining data before conducting statistical analyses?
    b. Why are outliers of concern to researchers?
    c. What questions do you address about a set of data for a variable in constructing frequency distribution tables?
    d. Why is it useful to calculate percentages in addition to frequencies for each value of a variable?
2. For each of the following situations, create a frequency distribution table.
    a. A college counselor asks a group of 116 seniors what their plans are after graduating from college; 71 say they are going to “work,” 34 are planning on going to “graduate school,” and 11 say they are “not sure” what their plans are.
    b. A high school instructor teaching a course finds that of the students she teaches, 69 are Freshmen, 18 are Sophomores, 12 are Juniors, and 9 are Seniors.
    c. A pollster stops 12 people and asks if they are either “in favor” or “against” a local proposition in Table 2.5 :

Table 2.5 Example of People’s Position

Person	Position
1	**In favor**
2	**In favor**
3	**Against**
4	**In favor**
5	**Against**
6	**Against**
7	**In favor**
8	**Against**
9	**In favor**
10	**Against**
11	**In favor**
12	**In favor**

    d. The company that makes M&Ms candy (www.mms.com) conducted a survey asking people which new color they would like to have added: purple, aqua, or pink. Table 2.6 are the votes of a hypothetical sample:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Table 2.6
Hypothetical Sample of New M&M Color

Person	Color
1	**Pink**
2	**Purple**
3	**Pink**
4	**Aqua**
5	**Purple**
6	**Purple**
7	**Purple**
8	**Aqua**
9	**Purple**
10	**Pink**
11	**Aqua**
12	**Pink**
13	**Pink**
14	**Purple**
15	**Aqua**
16	**Pink**
17	**Purple**
18	**Purple**
19	**Aqua**
20	**Purple**
21	**Purple**
22	**Pink**
23	**Pink**
24	**Purple**

2.4 Grouped Frequency Distribution Tables

The frequency distribution table in Table 2.4 was sufficient for presenting the frequencies for the lottery ticket variable. However, other situations may involve variables with a large number of possible values. Consider, for example, the variable grade point average (GPA). The values for GPA typically range from 0.00 to 4.00, with hundreds of possible values in between. Creating a frequency distribution table for GPA could result in an extremely large table, with many values having frequencies of zero.

To examine a variable with a large number of values, it is useful to construct a **grouped frequency distribution table** , a table that groups the values of a variable measured at the interval or ratio level

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

of measurement into a small number of intervals and then provides the frequency and percentage for each interval. To illustrate how grouped frequency distribution tables are constructed, Table 2.7 (a) provides a sample of 42 hypothetical GPAs. In Table 2.7 (b), each of these 42 GPAs has been placed into one of eight intervals that comprise a grouped frequency distribution table.

To illustrate how to read and interpret a grouped frequency distribution table, the lowest interval in Table 2.7 (b) consists of students with GPAs less than (<) 2.00. The frequency in this interval (*f* = 1) consists of one student: Student 29 (GPA = 1.96). The next interval consists of students with GPAs ranging from 2.00 to 2.29 and contains three students (4, 22, and 38). This grouping continues until all of the GPAs in the sample have been accounted for.

To create a grouped frequency distribution table, the real limits of each interval should be identified. The **real limits** of an interval are the values of the variable that are halfway between the top of one interval and the bottom of the next interval. For example, because of how values of GPA are rounded, the lowest value of GPA that would lead someone to be placed in the interval 2.30–2.59 is 2.295, which is halfway between the top of the 2.00–2.29 interval and the bottom of the 2.30–2.59 interval. The **real lower limit** is the smallest value of a variable that would be grouped into a particular interval. In this case, 2.295 represents the real lower limit for the 2.30–2.59 interval. On the other hand, the **real upper limit** is the largest value of a variable that would be grouped into a particular interval. For the 2.30–2.59 interval, the real upper limit is 2.594.

Once the grouped frequency table for the GPA example has been created (Table 2.7(b)), what preliminary conclusions might be drawn about these students? First, because the 2.90–3.19 interval has the greatest frequency (*f* = 10), you might conclude that the typical student in this sample has a GPA close to 3.00. Second, moving in both directions away from the 2.90–3.19 interval, the frequencies of the different intervals grow progressively smaller, with fewer and fewer students having either very high or very low GPAs. Table 2.8 provides a list of guidelines for creating grouped frequency distribution tables.

Cumulative Percentages

In addition to determining the percentage of the sample that has each individual value or interval for a variable, it is sometimes useful to combine these percentages. For the GPA example, you may be interested in knowing, “What percentage of the sample had GPAs less than 2.60?” To answer this question, the percentages in the “%” column of the frequency distribution table must be combined. This type of percentage is referred to as a **cumulative percentage** , defined as the percentage of a sample at or below a particular value of a variable.

Table 2.7a Grade Point Average (GPA), 42 Students: Individual GPAs

Student	GPA
1	**3.05**
2	**2.83**
3	**3.26**
4	**2.19**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Student	GPA
5	**3.52**
6	**3.34**
7	**3.02**
8	**2.97**
9	**3.71**
10	**2.50**
11	**2.77**
12	**3.10**
13	**3.06**
14	**3.83**
15	**2.39**
16	**3.16**
17	**3.92**
18	**2.37**
19	**3.65**
20	**3.70**
21	**3.00**
22	**2.26**
23	**2.65**
24	**3.40**
25	**2.87**
26	**2.92**
27	**2.89**
28	**3.41**
29	**1.96**
30	**3.77**
31	**3.20**
32	**2.86**
33	**2.94**
34	**2.71**
35	**3.43**
36	**3.95**
37	**3.32**
38	**2.28**
39	**3.08**
40	**3.25**
41	**2.40**
42	**3.22**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Table 2.7b Grade Point Average (GPA), 42 Students: Grouped Frequency Distribution Table

GPA	*f*	%
3.80+	3	7%
3.50−3.79	5	12%
3.20−3.49	9	21%
2.90−3.19	10	24%
2.60−2.89	7	17%
2.30−2.59	4	10%
2.00−2.29	3	7%
<2.00	1	2%
Total	42	100%

Table 2.8 Guidelines for Creating Grouped Frequency Distribution Tables

Guideline	Guideline Description
1. *Variables are grouped into approximately 10 intervals.*	The exact number of intervals depends on the data in each sample. For example, data that fall within a small range of values may be accurately represented with a relatively small number of intervals, whereas data extending across a wide range may require a larger number of intervals.
2. *The number of intervals should accurately represent the data.*	The intervals should represent the nature of the data as accurately as possible. For example, there should not be so large a number of intervals that many intervals have frequencies of zero (e.g., 3.00−3.05) or so few intervals that a large majority of the sample falls into only one or two intervals (e.g., 3.00−3.99).
3. *Intervals should be of equal size.*	To make the intervals comparable, they should be of equal size or width. For example, you would not want one interval of 2.01−2.50 (a width of .50) and another of 2.51−2.70 (a width of .20). One exception to this guideline is the lowest or highest interval (e.g., less than 2.00 or greater than 3.70), which is typically wider than the other intervals because it contains a relatively small frequency.
4. *Intervals should not overlap.*	The intervals should be distinct from each other, such that each score may be placed in only one interval. For example, intervals of 2.30− *2.60* and *2.60*− 2.90 are a problem because the GPA 2.60 is in both intervals leading to confusion regarding within which interval to place someone with a GPA of 2.60.

Table 2.9 includes the cumulative percentages for the GPA example. As an example of a cumulative percentage, the cumulative percentage for the 2.30–2.59 interval (19%) has been calculated by combining the percentages for three intervals: < 2.00, 2.00–2.29, and 2.30–2.59: 2% + 7% + 10% = 19%. As a result, we could conclude that 19% of this sample had GPAs less than 2.60. The accumulating of percentages continues until all (100%) of the scores in the sample are combined. As

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

shown in Table 2.9 , cumulative percentages are typically located in the final column of a frequency distribution table and labeled “Cum %.”

Table 2.9 Frequency Distribution Table With Cumulative Percentages for the GPA Variable

GPA	*f*	%	Cum %
3.80+	3	7%	**100%**
3.50−3.79	5	12%	**93%**
3.20−3.49	9	21%	**81%**
2.90−3.19	10	24%	**60%**
2.60−2.89	7	17%	**36%**
2.30−2.59	4	10%	**19%**
2.00−2.29	3	7%	**9%**
< 2.00	1	2%	**2%**
Total	42	100%	

2.5 Examining Data Using Figures

In addition to organizing data into frequency distribution tables, researchers also examine data visually using figures such as charts or graphs. There are four types of figures often used to visually portray data for variables: bar charts, pie charts, histograms, and frequency polygons. Which type of figure a researcher may use depends on the variable’s level of measurement (nominal, ordinal, interval, ratio). Data for variables measured at the nominal or ordinal level of measurement are typically displayed using bar charts or pie charts. The values of interval and ratio variables are graphed using histograms and frequency polygons. Each of these types of visual illustration is discussed in the following sections.

Displaying Nominal and Ordinal Variables: Bar Charts and Pie Charts

The values of variables measured at the *nominal* level of measurement differ in category or type; one example of a nominal variable is the lottery ticket variable discussed earlier, which consisted of four categories: sequence, pattern, nonequilibrated, and random. The values of variables measured at the *ordinal* level of measurement can be placed in an order relative to the other values; age (young, middle age, elderly) is one example of an ordinal variable.

One way to display nominal and ordinal variables is to use a bar chart. A **bar chart** is a figure that uses bars to represent the frequency or percentage of a sample corresponding to each value of a variable, with the different bars not touching each other. Figure 2.1 shows a bar chart for the data from the lottery ticket variable. In this figure, each bar corresponds to one of the four types of lottery tickets; the height of each bar corresponds to the frequency (*f*) of nonproblem gamblers who selected that type of ticket.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

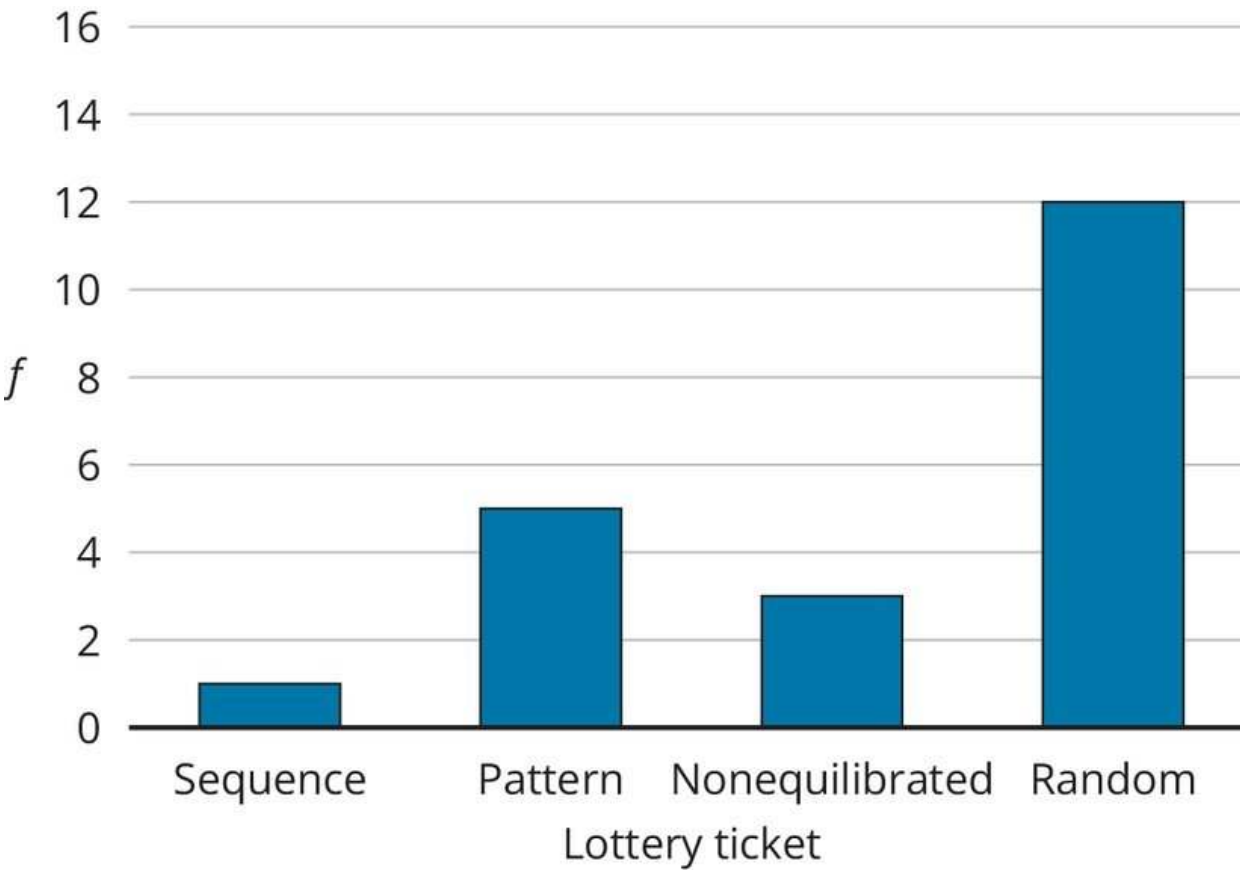


Figure 2.1 Bar Chart Showing the Frequencies of the Lottery Ticket Variable for the Nonproblem Gamblers

When creating a bar chart, the horizontal (X) axis contains the name of the variable and the values of the variable; the vertical (Y) axis contains the frequency (*f*) or percentage (%) for each value. Notice that the bars in the bar chart in Figure 2.1 do not touch each other; the gap between the bars indicates that the values of the variable represent distinct groups or categories that can't be placed along a numeric continuum.

A second type of figure used to display nominal or ordinal variables is a pie chart. A **pie chart** is a figure that uses a circle divided into proportions (slices) that represent the percentage of the sample for each value of a variable. Pie charts are created by dividing the 360 degrees of a circle into proportions corresponding to the percentage of each value.

Figure 2.2 shows a pie chart for the data for the lottery ticket variable. To illustrate how pie charts are constructed, the frequency distribution table in Table 2.4 shows that 5% of nonproblem gamblers chose the sequence ticket. Given that a pie chart consists of 360 degrees, 5% of the 360 degrees are needed to represent the sequence ticket. As 5% * 360 = 18, 18 degrees of the pie is assigned to the sequence ticket. Similarly, 97 degrees of the pie chart (27% of 360) are used to represent the pattern ticket. This process continues until all 360 degrees of the pie chart are accounted for.

The decision whether to use a bar chart or pie chart to illustrate a variable may depend on the number of values for the variable. For example, a pie chart for a variable that has 12 values may

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

result in many of the 12 “slices” being very thin and small, making them hard to see, meaning that a bar chart may be a better choice to illustrate this variable.

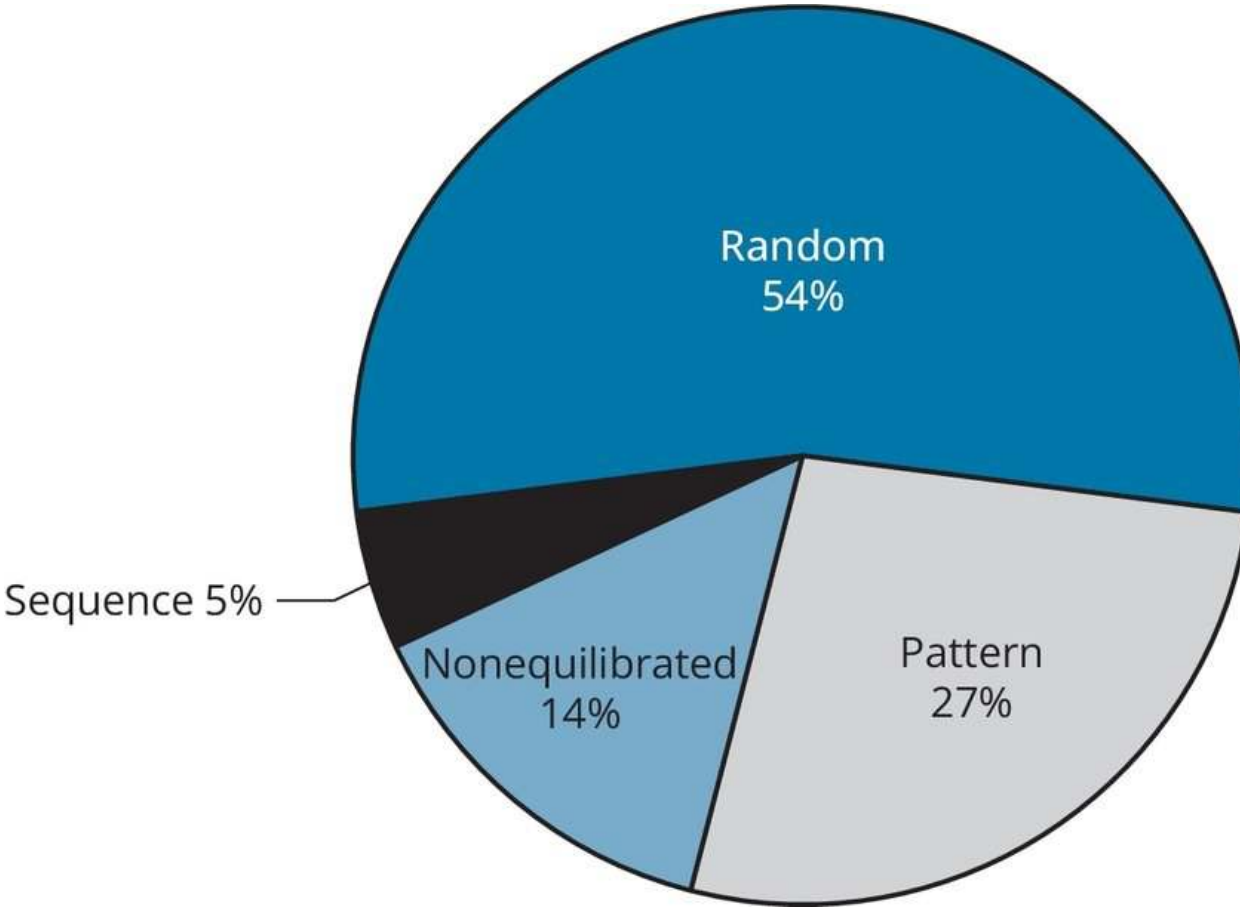


Figure 2.2 Pie Chart Showing the Percentages of the Lottery Ticket Variable for the Non−Problem Gamblers

Displaying Interval and Ratio Variables: Histograms and Frequency Polygons

Interval and ratio variables have values equally spaced along a numeric continuum and are identical to each other with one exception: Ratio scales possess a true zero point, for which the value of zero (0) represents the complete absence of the variable. The values of interval and ratio variables may be illustrated using histograms and frequency polygons, the choice of which depends on the number of values for the variable.

A **histogram** is a figure in which bars are used to represent the frequency of each value of a variable, with the different bars touching each other. The name and values of the variable are located on the X-axis, and the frequency or percentage for each value is on the Y-axis. Histograms are similar to bar charts, with one important difference: The bars in a histogram touch each other, indicating that the values of the variable are connected to each other along a numeric continuum.

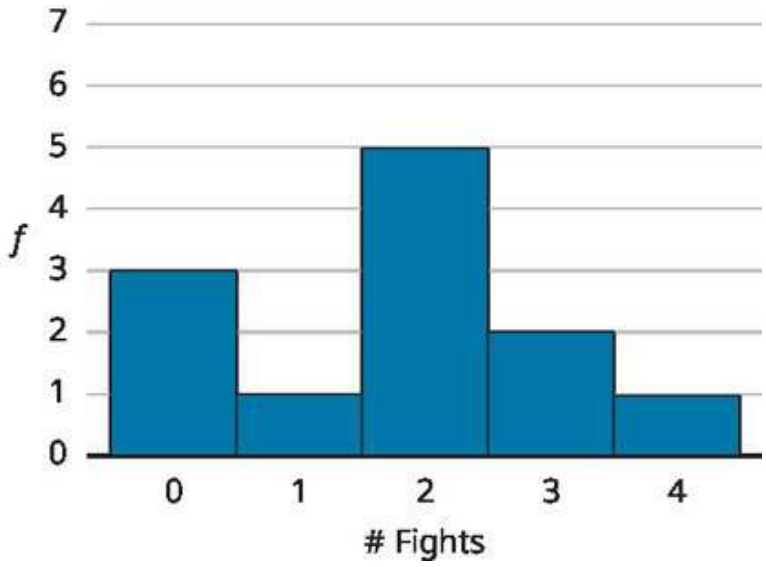
join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

As an example, imagine a teacher is interested in studying aggressive behavior in children and measures this variable by counting the number of fights each of her students is involved in over a 3-day period. “Number of fights” is a ratio variable rather than an interval variable because the value of 0 represents the complete absence of fights. Figure 2.3 provides a frequency distribution table and histogram for this variable.

A histogram is used when an interval or ratio variable consists of a relatively small number of values. But what if, for example, the number of fights could be anywhere from 0 to 20 rather than 0 to 5? In order to have the histogram fit on one page, the 21 bars would need to be very thin and crowded together, making the histogram confusing and difficult to read. In situations such as this, the values of the variable may be more clearly illustrated using a figure known as a **frequency polygon**, which is a line graph that uses data points to represent the frequency of each value of a variable, with straight lines connecting the data points. In a frequency polygon, the X-axis represents the name and values of the variable, while the Y-axis represents the frequency for each value.

# Fights	*f*
4	1
3	2
2	5
1	1
0	3
Total	12



▶ Long Description: Figure 2.3 Frequency Distribution Table and Histogram for the Number of Fights Variable

Figure 2.3 Frequency Distribution Table and Histogram for the Number of Fights Variable

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

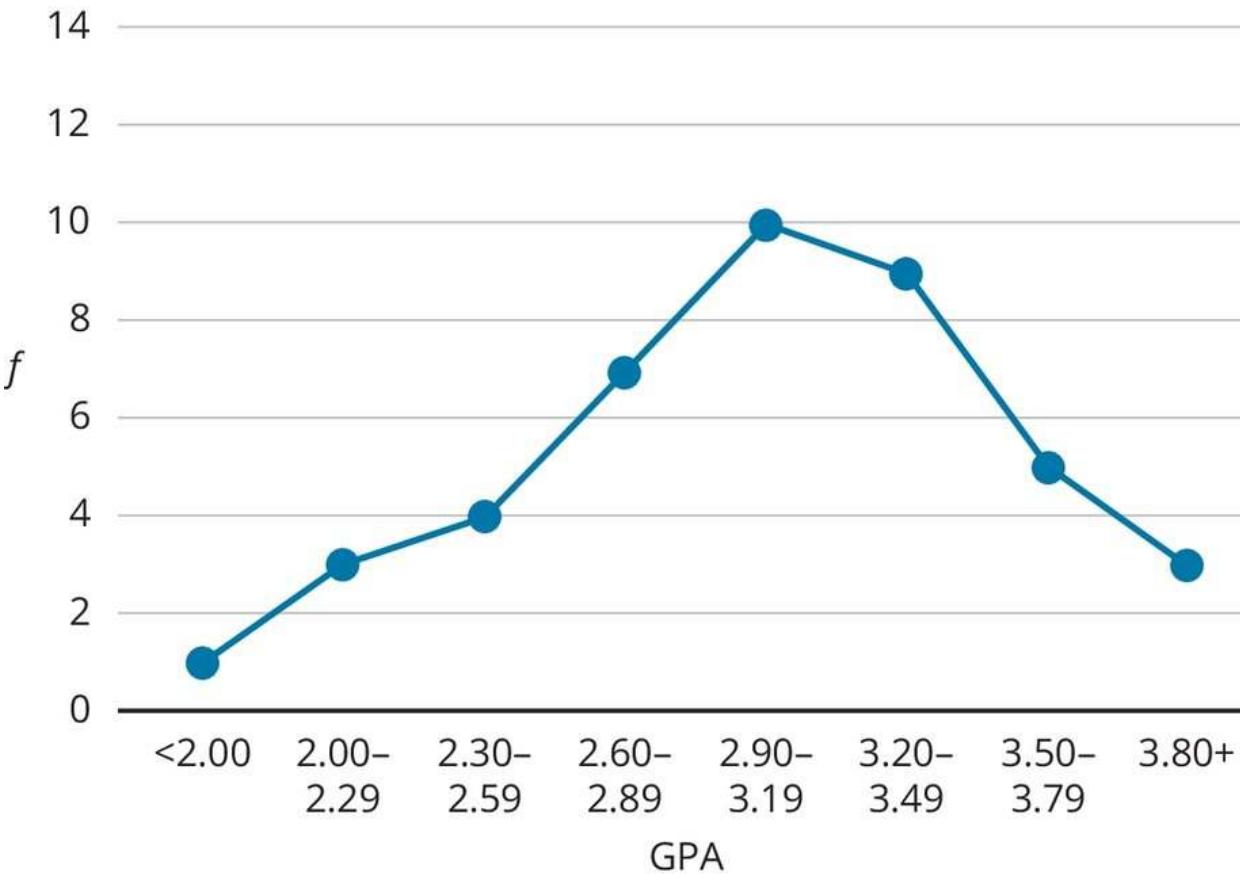


Figure 2.4 Frequency Polygon Showing the Frequency of Students in Each Interval of the Grade Point Average (GPA) Variable

Figure 2.4 shows a frequency polygon based on the grouped frequency distribution table of the earlier example of students' GPA (Table 2.7(b)). In this frequency polygon, the data points representing the frequencies of the eight GPA intervals are connected using straight, sharp lines. These lines indicate that the values of the variable can be connected together along a numeric continuum.

Although a frequency polygon like the one in Figure 2.4 may look jagged as it moves from one data point to the next, it provides a general sense of the shape of the distribution of scores for a variable. For example, looking at Figure 2.4 , you might conclude that the majority of GPAs in this sample are in the middle of the 1.00 to 4.00 range, with relatively low frequencies of students with either low or high GPAs.

Learning Check 2

Reviewing What You've Learned So Far

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



1. Review questions
    a. What is the difference between a frequency distribution table and a grouped frequency distribution table?
    b. For what levels of measurement would you create a pie chart, bar chart, histogram, or a frequency polygon?
    c. What is the difference between a bar chart and a histogram?
    d. Why would it be inappropriate to create a histogram or frequency polygon for a nominal or ordinal variable?
    e. Under what circumstances might you use a frequency polygon rather than a histogram to graph an interval or ratio variable?
2. For each of the following variables, identify whether you could use a bar chart, pie chart, histogram, and/or a frequency polygon to visually display the data.
    a. Miles per gallon (MPG)
    b. Living situation (on-campus, off-campus)
    c. Number of brothers and sisters
    d. Marital status (single, married, divorced)

Drawing Inappropriate Conclusions From Figures

A well-constructed figure can greatly facilitate the understanding of complex information. However, a poorly designed figure may not only be confusing, it can lead to inappropriate or misleading conclusions. Box 2.1 provides guidelines for constructing figures based on recommendations from the American Psychological Association (APA).

A simple example may demonstrate how altering the features of a figure can lead to different interpretations of the same set of data. Imagine a pollster assesses how voters with different political affiliations plan to vote on a ballot measure. She asks a sample of Democrats, Republicans, and Independents the question, “Do you plan on voting in favor of Measure A?” In examining her data, the pollster finds 56% of Democrats, 47% of Republicans, and 52% of Independents intend to vote “yes” on the measure.

Figure 2.5 graphs these three percentages using two bar charts; bar charts are used to display the data because the variable of interest (political affiliation) is a nominal variable (Democrat, Republican, or Independent). The bars don’t touch each other because the three political affiliations are distinct from each other and can’t be placed along a numeric continuum.

Although the two bar charts are similar, there is one important difference in how they are constructed. Following the guidelines in Box 2.1, the bar chart in Figure 2.5 (a) starts the Y-axis with the value 0% and ends with 100%. However, the bar chart in Figure 2.5 (b) starts with the value 45% and ends with 65%. By limiting the range of values of the percentages, the bar chart in Figure 2.5 (b) exaggerates differences in the heights of the three bars. As a result, someone looking at Figure 2.5 (b) may interpret the differences between the three political affiliations as being larger than someone looking at Figure 2.5 (a).

The saying “a picture is worth a thousand words” is particularly relevant for the presentation of statistical data. Because figures can provide a vivid representation of data, it’s the responsibility of researchers to portray their data in an accurate manner and to avoid creating images that may mislead or confuse.

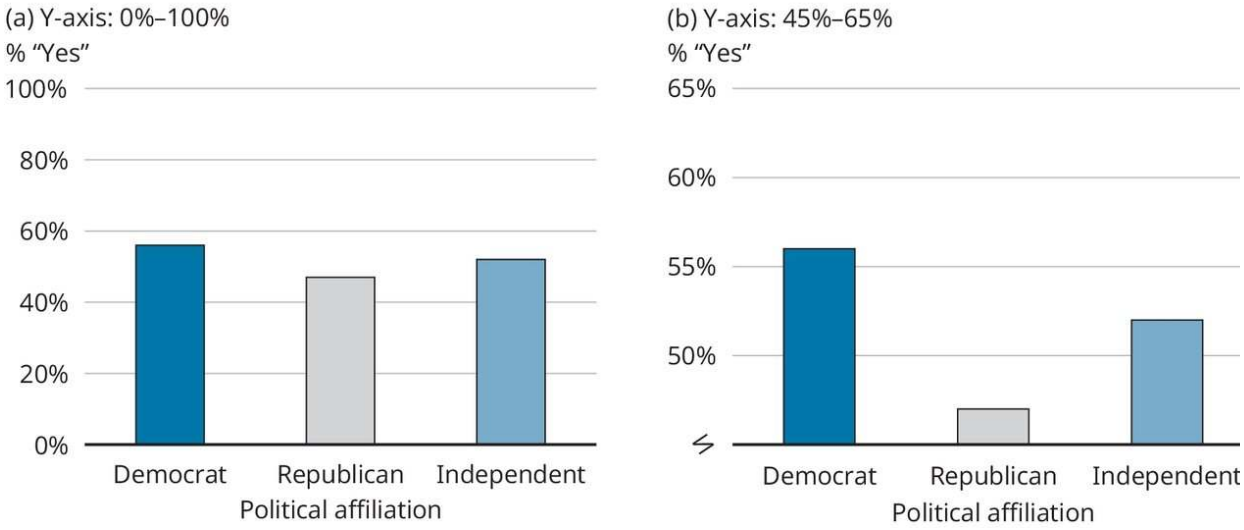
join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Box 2.1 Some APA Guidelines for Creating Figures

1. The figure should be formatted in a portrait rather than landscape page orientation (the reader should not turn the page sideways to read the table).
2. The figure is numbered using Arabic numbers (Figure 1, not Figure *I* or Figure *A*).
3. The figure includes a title that describes the contents of the figure. For example, it is not enough to name a figure simply " Figure 2.1 " or " Figure 2.1 . Bar chart."
4. Figures have a rectangular shape, with the height of the vertical (Y) axis two thirds to three fourths the length of the horizontal (X) axis.
5. The horizontal (X) axis contains the values of the variable; the vertical (Y) axis contains the frequency (*f*) or percentage (%) of the sample having each value of the variable.
6. The *X* - and *Y* -axes include descriptive labels written in both upper- and lowercase letters (as opposed to all capital letters).
7. The labels for the *X* - and *Y* -axes are short enough to fit on one line along the axes.
8. The label of the Y-axis is often placed parallel to the Y-axis (the letters face sideways rather than vertically) but can sometimes be seen on top of the y-axis for accessibility.
9. The Y-axis starts with the value zero (0) and is divided into equally spaced values or intervals.
10. The Y-axis is divided into 6 to 10 values (just enough values to accommodate bar length). The bars should neither be flattened along the bottom of the Y-axis nor bumped up against the top of the Y-axis.

Political affiliation	% "Yes"
Democrat	56%
Republican	47%
Independent	52%



▶ Long Description: Figure 2.5 Different Displays of the Same Data

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Figure 2.5 Different Displays of the Same Data

2.6 Examining Data: Describing Distributions

We've discussed how researchers construct tables and figures to organize their data. Looking at a table or figure enables researchers to describe the distribution of values or scores for a variable. Describing a distribution involves addressing questions such as, "Which values of the variable occur most often?," "How do the frequencies of the values of the variable change in relation to the most common values?," and "What is the amount of differences among the different values?"

Questions such as these pertain to three aspects of distributions: modality, symmetry, and variability. These three aspects of distributions are introduced in the following sections; they'll also be discussed in greater detail in the next two chapters of this book.

Modality

The first feature of a distribution of data for a variable is its modality. The **modality** of a distribution refers to the value or values of the variable that have the highest frequency or occur most often in a set of data. Using the lottery study as an example, identifying the modality of the lottery ticket variable involves answering the question, "Which of the four lottery tickets was chosen the most often?"

The modality of a distribution can be identified by looking at a frequency distribution table; the modality is the value(s) of the variable with the highest frequency (*f*) or percentage (%). The modality may also be determined by examining a figure and locating the tallest bar in a bar chart or histogram, the highest point in a frequency polygon, or the biggest "slice" in a pie chart. For the lottery ticket variable, the modality is the random ticket.

In identifying the modality of a distribution, it may be that one value or a limited number of values adjacent to one another clearly have the highest frequency. In such cases, the distribution is referred to as a unimodal distribution. A **unimodal distribution** is a distribution where one value of a variable occurs with the greatest frequency. Looking at the frequency distribution table (Table 2.4) or figure (Figure 2.1) for the lottery ticket variable, this distribution would be considered unimodal as one value, the random ticket, clearly occurred more often than the other three ticket choices.

In contrast to a unimodal distribution, it's possible that two distinct values may both have relatively high frequencies; if so, the distribution is referred to as a **bimodal distribution** , which is a distribution where two values clearly distinct from each other have the greatest frequency. A bimodal distribution sometimes resembles the two humps on a camel's back. As an example of a bimodal distribution, in April 2024 the Pew Research Center (www.pewresearch.org) asked a sample of 7,000 voters whether they felt Donald Trump had the mental fitness needed to be president, with five responses ranging from "extremely" to "not at all." The majority of voters chose one of the two extreme choices, either "extremely" (23%) or "not at all" (35%), with only 14% choosing the middle response of "somewhat."

And, in addition to unimodal and bimodal distributions, it's possible for a distribution to have more than two values with relatively equally high frequencies; this type of distribution is known as a **multimodal distribution** —a distribution where more than two values have the greatest frequency. In examining people's attitudes toward gun control, in May 2023, the news network CNN asked 1,227 people the question, "Do you think allowing gun owners to carry their guns in public places makes those places safer, less safe, or makes no difference in how safe a public place is?" They found people differed greatly in their responses, with 32% saying "makes public places safer," 36% saying

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

"makes public places less safe," and 32% saying "makes no difference," Figure 2.6 presents an example of a unimodal, bimodal, and multimodal distribution.

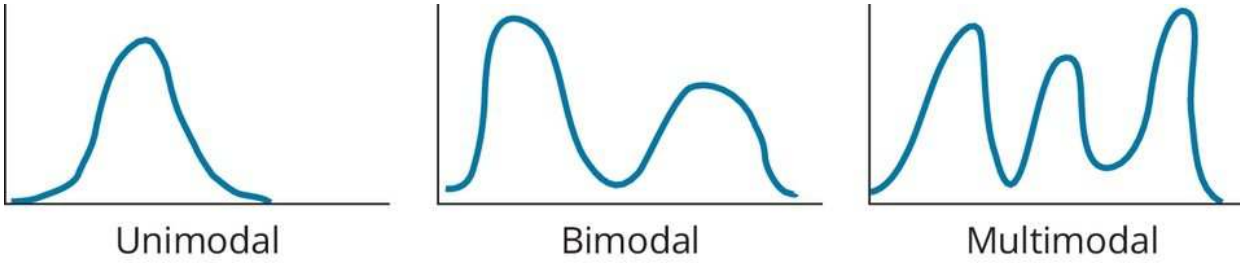


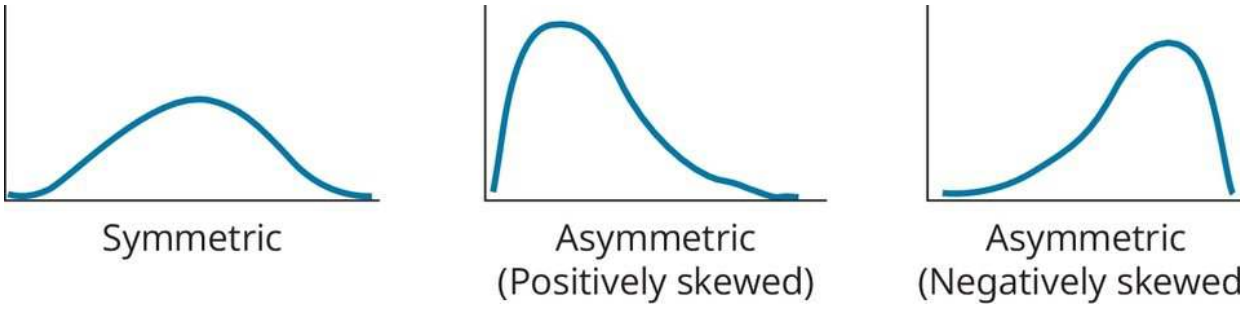
Figure 2.6 Distributions With Different Modalities (Unimodal, Bimodal, and Multimodal)

Symmetry

The second feature of a distribution of data is its symmetry. The **symmetry** of a distribution refers to how the frequencies of values of a variable change in relation to the most common or frequently occurring values. Describing the symmetry of a distribution means answering the question, "What is the shape of the distribution?"

The symmetry of a distribution can be determined by looking at the frequencies in a frequency distribution table or the bars or points in a figure. By doing so, a distribution may be described as being either symmetric or asymmetric. A **symmetric distribution** is one in which the frequencies change in a similar manner moving away in both directions from the most frequently occurring values. An example of a symmetric distribution is illustrated in Figure 2.7 . One way to determine whether a distribution is symmetric is to draw the distribution on a piece of paper and then fold the distribution vertically down the middle so that the two halves are on top of each other. A distribution is symmetric when the two halves are similar in shape such that they resemble mirror images of each other.

However, a distribution may also be asymmetric. An **asymmetric distribution** (also known as a *skewed* distribution) is one in which the frequencies change in a different manner moving away in both directions from the most frequently occurring values. In asymmetric distributions, the highest frequencies are located at one end of the distribution rather than in the middle, such that if you fold an asymmetric distribution down the middle, the shapes of the two halves do not look the same.



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Figure 2.7 Distributions With Different Symmetries (Symmetric, Positively Skewed, and Negatively Skewed)

In an asymmetric distribution, the frequencies take on the shape of a long tail as the values move toward the end of the distribution; scores in the tail of the distribution are often considered to be outliers. For example, in the 2022 Current Population Survey, the U.S. Census Bureau (www.census.gov) estimated that most Americans (63%) had annual incomes less than $50,000. However, a very small percentage (3%) were greater than $200,000; these higher incomes would be considered outliers.

Depending on the location of its tail, asymmetric distributions are referred to either as positively skewed or negatively skewed. A **positively skewed distribution** is one in which the higher frequencies are at the lower end of the distribution, with the tail on the upper (right) end of the distribution; a **negatively skewed distribution** is one in which the higher frequencies are at the upper end of the distribution, with the tail on the lower (left) end of the distribution. Figure 2.7 provides an illustration of a symmetric distribution and two asymmetric distributions (one positively skewed and one negatively skewed). Annual income, illustrated by the U.S. Census Bureau example mentioned earlier, is an example of a positively skewed distribution.

Variability

The third way in which distributions may be described is in terms of the amount or nature of differences among the different values for a variable. The **variability** of a distribution refers to the amount of differences in a distribution of data for a variable. The variability in a distribution is described by answering the question, “To what extent are the scores in the distribution different from each other?”

The amount of variability in a distribution can be estimated by looking at a frequency distribution table or figure. Imagine, for example, an instructor gives an 11-item quiz to her class and finds almost all of the students answer six, seven, or eight of the items correctly. In this situation, the distribution would resemble the shape of a tall mountain: thin and tall, with many of the scores residing in a few values of the variable. A distribution with a small amount of variability such as this one is referred to as a **peaked distribution** , defined as a distribution where much of the data are in a small number of values of a variable.

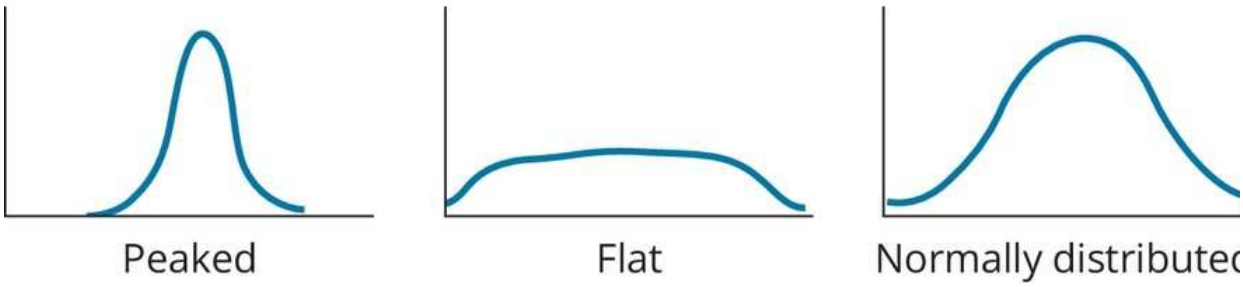


Figure 2.8 Distributions With Different Variability (Peaked, Flat, and Normally Distributed)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

On the other hand, what would the distribution look like if she had found her students' scores were distributed relatively evenly across the entire range of possible scores? In this situation, the distribution would not be tall and thin like a mountain but rather would be low and wide like a plateau. This type of distribution is referred to as a **flat distribution** , which is a distribution where the data are spread evenly across the values of a variable. The CNN gun control example described earlier is an example of a flat distribution.

In discussing modality, symmetry, and variability, distributions may have more than one mode, may be asymmetric (skewed), and may be peaked or flat. But what if a distribution has none of these qualities? A distribution that is unimodal, symmetric, and neither peaked nor flat may be described as "bell shaped." If you look at a bell, you'll notice that it is the highest in the middle, and as you move away from the top of a bell in both the left and right directions, the shape of the bell slopes downward in the same manner. A bell-shaped distribution for a variable is given a special name: a normally distributed variable.

A **normally distributed variable** is a distribution for a variable that is considered unimodal, symmetric, and neither peaked nor flat. Note that the word *normal* is a statistical concept and not a value judgment. In other words, asymmetric, peaked, or flat distributions are not considered "abnormal." The concept of normal distributions will be discussed in greater detail in Chapter 5. Figure 2.8 provides an illustration of three distributions with different amounts of variability: peaked, flat, and normally distributed.

Learning Check 3

Reviewing What You've Learned So Far



1. Review questions
    a. What are the three aspects used to describe distributions of variables?
    b. What is the difference between unimodal, bimodal, and multimodal distributions?
    c. What is the difference between symmetric and asymmetric (skewed) distributions?
    d. What is the difference between positively and negatively skewed distributions?
    e. What is the difference between peaked and flat distributions?
    f. What are the characteristics of a normally distributed variable?
2. For each of the following sets of data, determine whether the distribution of scores is unimodal, bimodal, or multimodal.
    a. 4, 2, 11, 6, 10, 2, 6, 13, 2, 10, 6, 10, 1
    b. 15, 17, 13, 18, 17, 14, 20, 18, 17, 14, 17
    c. 20, 10, 50, 10, 80, 30, 70, 10, 60, 10, 80, 0, 30, 80, 10, 100, 80
3. For each of the following sets of data, determine whether the distribution of scores is symmetric or asymmetric (skewed).
    a. 4, 5, 1, 4, 2, 4, 3, 5, 3, 4, 4, 3
    b. 220, 230, 250, 230, 200, 240, 210, 230, 270, 220, 240, 230, 240, 230
    c. 3, 1, 9, 2, 3, 6, 2, 7, 3, 5, 2, 6, 3, 2, 2
4. For each of the following sets of data, determine whether the distribution of scores is peaked or flat.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

a. 16, 10, 19, 12, 13, 15, 14, 17, 15, 20, 11, 13, 19, 16, 15
b. 12, 15, 12, 11, 13, 11, 12, 12, 14, 10, 12, 11, 16, 12, 13, 12, 11
c. 14, 12, 16, 14, 13, 15, 12, 14, 14, 15, 14, 14, 13, 14

2.7 Looking Ahead

This chapter has discussed why and how researchers examine data. Examining data by creating tables and figures requires a number of different skills, including organizational skills (to prepare a set of data for statistical analysis), communication skills (to present data in tables or figures), and conceptual skills (to describe a distribution). The next two chapters focus on mathematical skills related to the statistical analysis of data, describing how researchers describe distributions numerically rather than visually. Given that this is a statistics textbook, we will focus on calculating statistics to describe and analyze data. However, and we cannot say this strongly enough, it is *extremely* important to examine a set of data before conducting statistical analyses as this will prevent errors both in your calculations of statistics as well as your interpretation of the results of these calculations.

2.8 Summary

Before conducting statistical analyses, researchers typically examine their data to gain an initial sense of the data, detect data coding or data entry errors, identify ***outliers*** (rare, extreme scores that lie outside of the range of the majority of scores in a set of data), evaluate research methodology, and determine whether the data meet statistical criteria and assumptions.

One common way of examining variables involves creating tables known as ***frequency distribution tables*** (tables that summarize the number and percentage of participants that express the different values for a variable) and ***grouped frequency distribution tables*** (tables that group the values of a variable measured at the interval or ratio level of measurement into a small number of intervals and then provide the frequency and percentage within each interval). To create a grouped frequency table, the ***real limits*** of each interval (the values of the variable that fall halfway between the top of one interval and the bottom of the next interval) need to be identified.

Data for a variable may also be examined by creating figures such as bar charts, pie charts, histograms, and frequency polygons. ***Bar charts*** and ***pie charts*** are used for variables measured at the nominal or ordinal level of measurement; a bar chart uses bars to represent the frequency or percentage of a sample corresponding to each value of a variable, with the different bars not touching each other; a pie chart is a circle divided into proportions that represent the percentage of the sample corresponding to each value of a variable. ***Histograms*** and ***frequency polygons*** are used for variables measured at the interval or ratio level of measurement; a histogram is a figure in which bars are used to represent the frequency of each value of a variable, with the different bars touching each other; a frequency polygon is a line graph that uses data points to represent the frequency of each value of a variable, with lines connecting the data points.

From examining their data, researchers can describe the distribution of values or scores for a variable. There are three main aspects of distributions: modality, symmetry, and variability.

The ***modality*** of a distribution is the value(s) of the variable that have the highest frequencies or occur most often in a set of data. ***Unimodal distributions*** are distributions where one value occurs with the greatest frequency, ***bimodal distributions*** have two values with the greatest frequency, and ***multimodal distributions*** have more than two values with the greatest frequencies.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The ***symmetry*** of a distribution refers to how the frequencies of values of a variable change in relation to the most common or frequently occurring values of the variable. In ***symmetric distributions*** , the frequencies change in a similar manner, moving away in both directions from the most common values of the variable. In ***asymmetric (skewed) distributions*** , the frequencies change in a different manner, moving away in both directions from the most common values of the variable.

The ***variability*** of a distribution refers to the amount of differences in a distribution of data. In ***peaked distributions*** , much of the data are in a few values of a variable; in ***flat distributions*** , the data are spread evenly across the values of a variable. A ***normally distributed variable*** has a symmetric distribution that is neither peaked nor flat.

2.9 Important Terms

asymmetric (skewed) distribution

bar chart

bimodal distribution

cumulative percentage

flat distribution

frequency (*f*)

frequency distribution table

frequency polygon

grouped frequency distribution table

histogram

modality

multimodal distribution

negatively skewed distribution

normally distributed variable

outlier

peaked distribution

pie chart

positively skewed distribution

real limits

real lower limit

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

real upper limit

symmetric distribution

symmetry

unimodal distribution

variability

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

2.10 Formula Introduced in This Chapter

Percentage (%)

$$\% = \frac{f}{total\ number\ of\ scores} * 100$$

2.11 Using IBM SPSS Software *

Creating Data Files: Defining Variables and Entering Data

* IBM SPSS software. SPSS is a registered trademark of International Business Machines Corporation.

1. Open a new data file (See Figure 2.9).

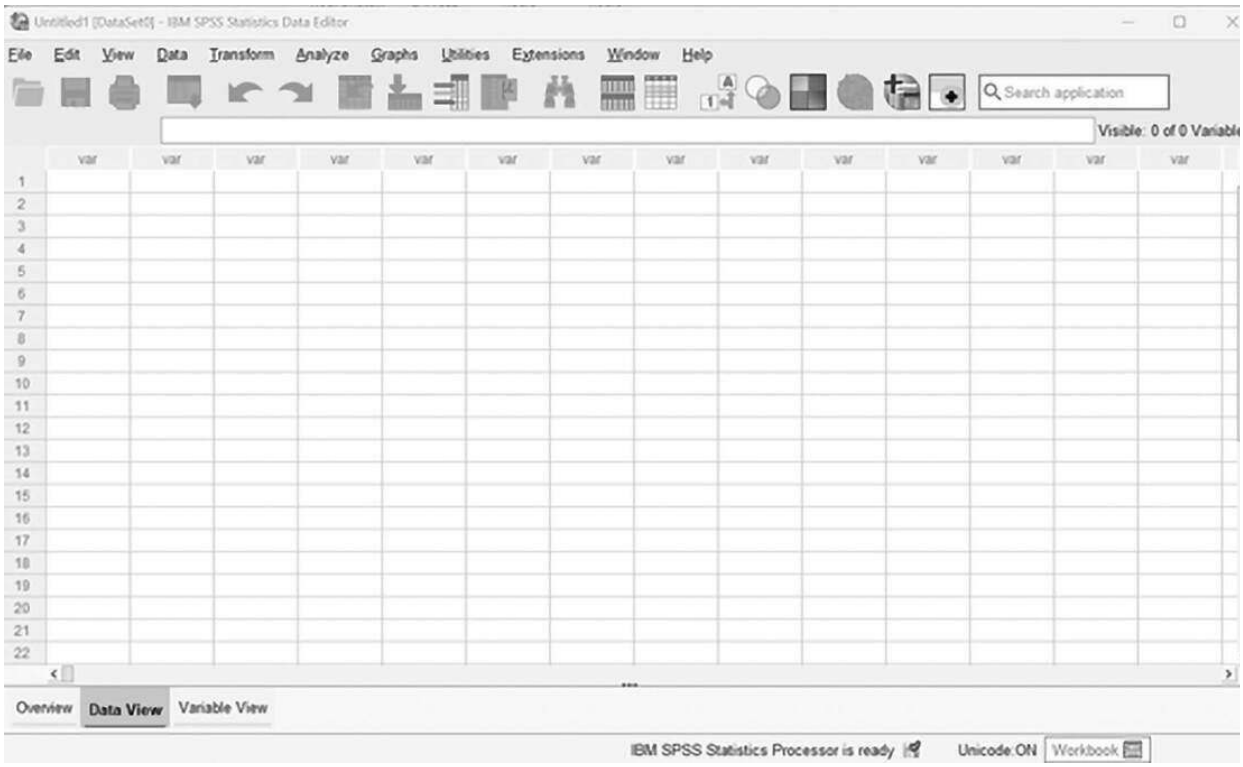


Figure 2.9 Data View

2. Define and name variable(s) (See Figure 2.10).

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

How? (1) Click **Variable View** tab, (2) click below **Name** , and (3) type name of variable in the cell.

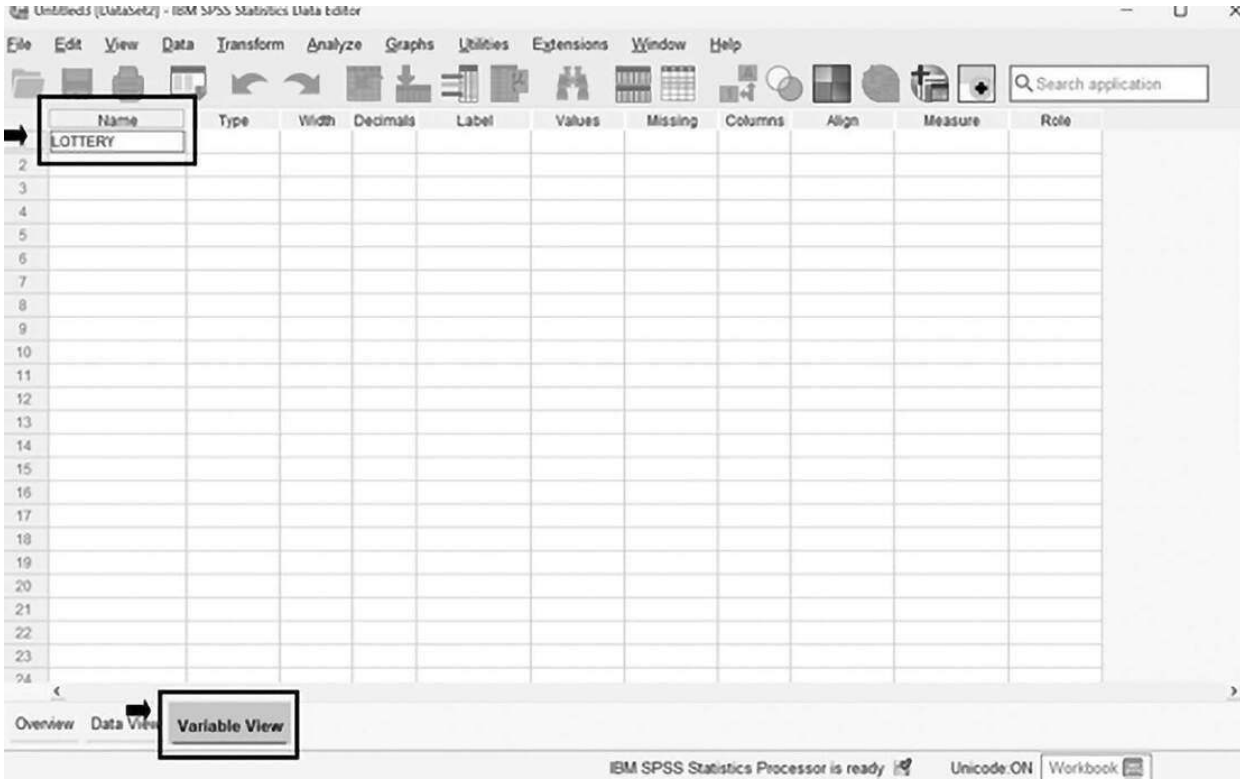


Figure 2.10 Variable View

3. Determine number of decimals for your variable(s) (See Figure 2.11).

How? Click desired number of decimal places in the **Decimals** box.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

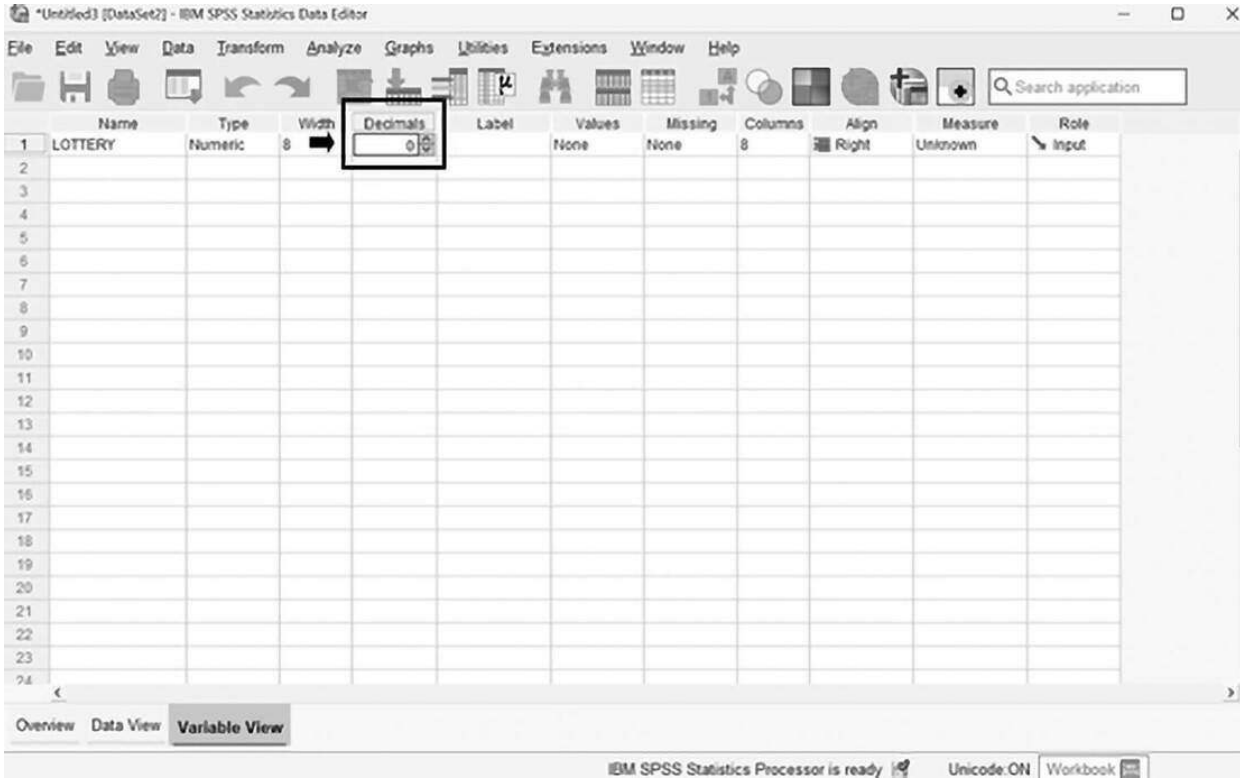


Figure 2.11 Decimals Box

4. Provide labels for your variable(s) (See Figure 2.12).

How? (1) Click below **Label** , and (2) type the label for the variable in the cell.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

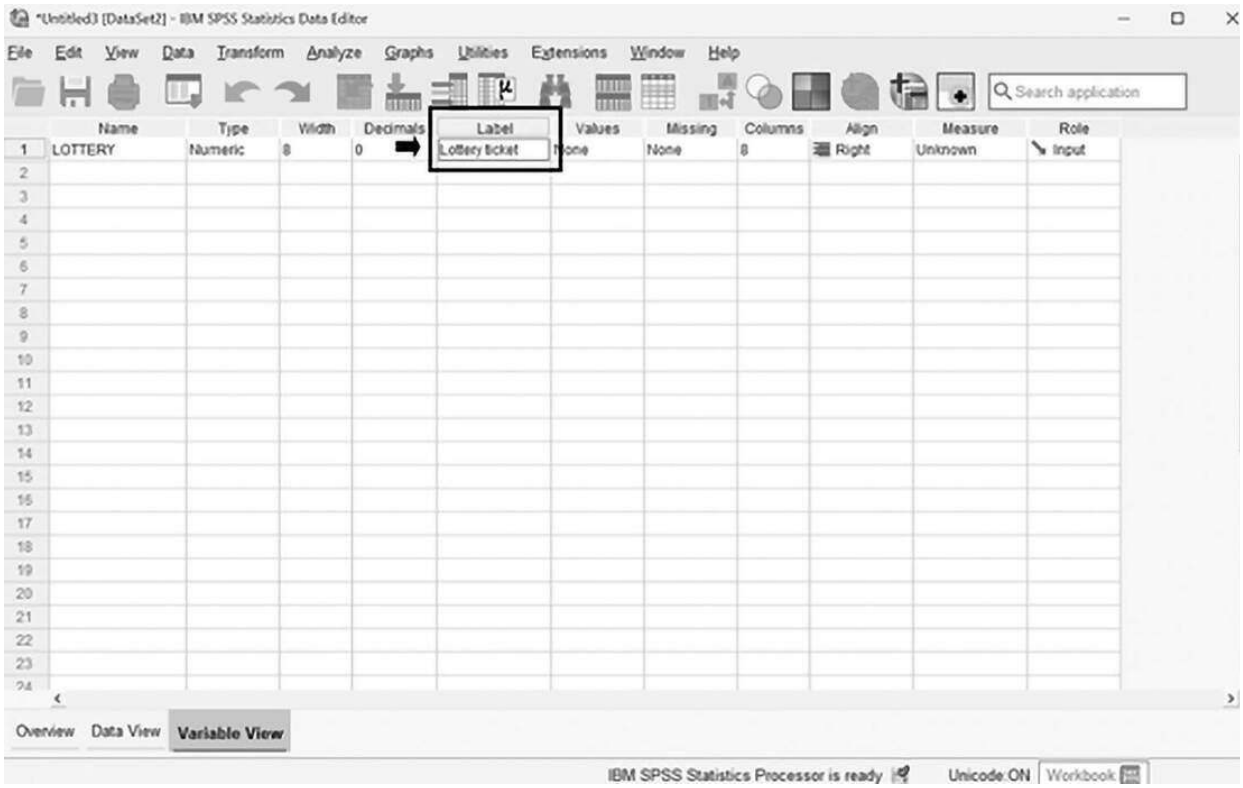


Figure 2.12 Label Box

5. Provide labels for values of nominal (categorical) variable(s) (See Figure 2.13).

How? (1) Click below **Values** , and (2) type labels for coded values of the variable.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

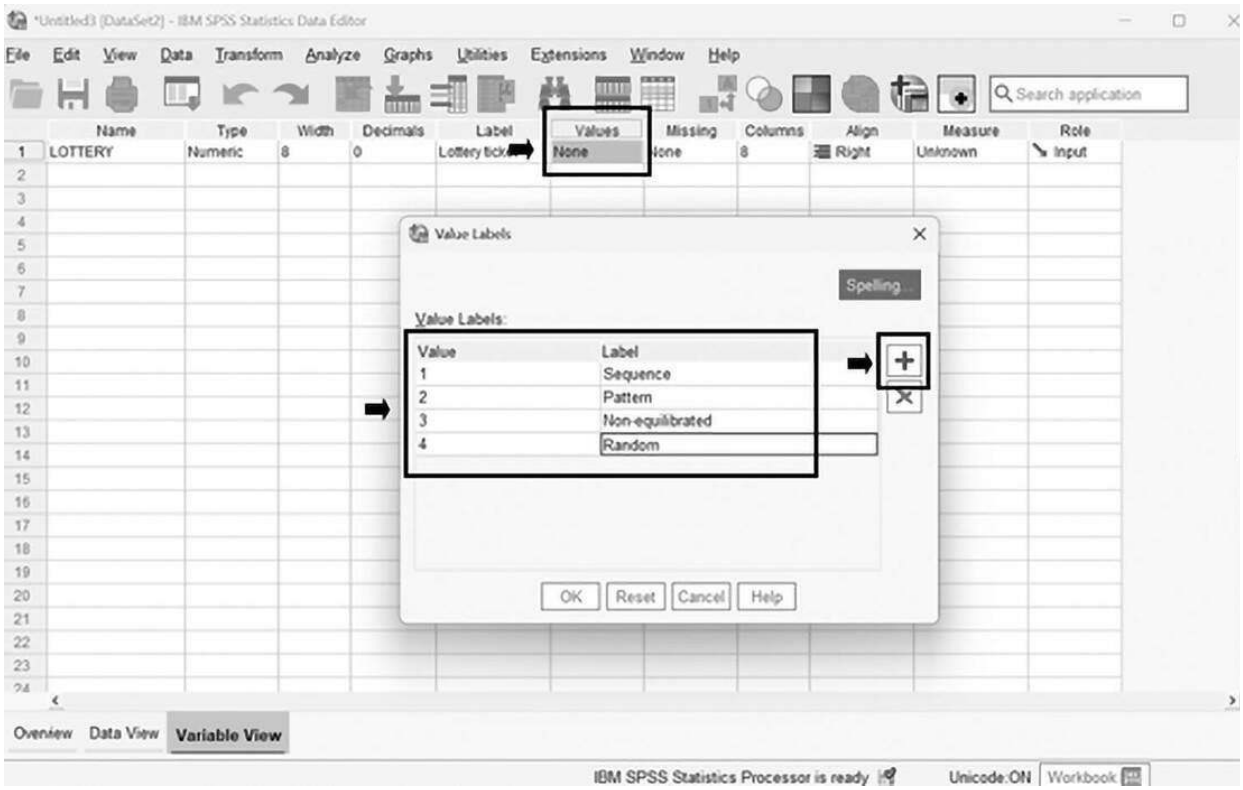


Figure 2.13 Nominal Variables

6. Enter data (See Figure 2.14).

How? (1) Click **Data View** tab and (2) enter data for the variable in the cells.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

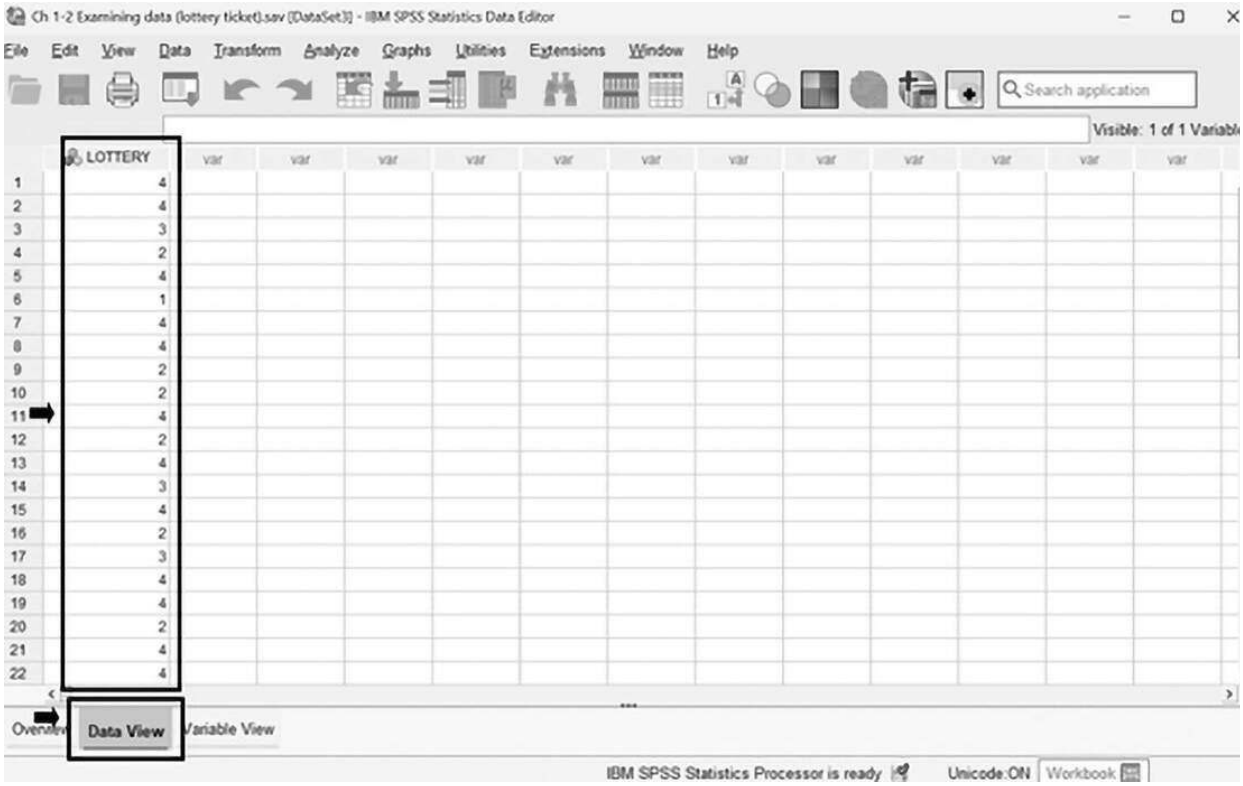


Figure 2.14 Enter Data View

Examining Data: Frequency Distribution Tables and Figures: The Lottery Study (2.1)

1. Select the frequency distribution procedure within SPSS (See Figure 2.15).

   How? (1) Click **Analyze** menu, (2) click **Descriptive Statistics** , and (3) click **Frequencies** .

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

To Jennifer , my kids Meagan and Eika , and my parents, Katsumi and Grayce Tokunaga.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

As a global academic publisher, Sage is driven by the belief that research and education are critical in shaping society. Our mission is building bridges to knowledge—supporting the development of ideas into scholarship that is certified, taught, and applied in the real world.

Sage's founder, Sara Miller McCune, transferred control of the company to an independent trust, which guarantees our independence indefinitely. This enables us to support an equitable academic future over the long term by building lasting relationships, championing diverse perspectives, and co-creating social and behavioral science resources that transform teaching and learning.

▶ Long Description:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

FUNDAMENTAL STATISTICS FOR THE SOCIAL AND BEHAVIORAL SCIENCES

Third Edition

Howard T. Tokunaga

San Jose State University



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

eISBN: 978-1-0718-6160-8

Copyright © 2027 by Sage.

All rights reserved. Except as permitted by U.S. copyright law, no part of this work may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without permission in writing from the publisher.

All third party trademarks referenced or depicted herein are included solely for the purpose of illustration and are the property of their respective owners. Reference to these trademarks in no way indicates any relationship with, or endorsement by, the trademark owner.



FOR INFORMATION:

2455 Teller Road, Thousand Oaks, California 91320, E-mail: orders@sagepub.com

3rd floor, HYLO, 103-105 Bunhill Row, London, EC1Y 8LZ, United Kingdom

10th Floor, Emaar Capital Tower 2, MG Road Sikanderpur, Sector 26, Gurugram, Haryana – 122002, India

18 Cross Street #10-10/11/12, China Square Central, Singapore 048423

Library of Congress Control Number: 2026943277

ISBN: 978-1-0718-6144-8

Printed in the United States of America

This book is printed on acid-free paper.

26 27 28 29 30 10 9 8 7 6 5 4 3 2 1

Acquisitions Editor: Anna Villarruel

Content Development Editor: Kenzie Offley

Editorial Assistant: Daria Terry

Production Editor: Syeda Aina Rahat Ali

Copy Editor: Michelle Ponce

Typesetter: Lumina Datamatics

Cover Designer: Candice Harman

Marketing Manager: Victoria Velasquez

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CONTENTS

Preface
Acknowledgments
About the Author
Chapter 1 Introduction to Statistics
Chapter 2 Examining Data: Tables and Figures
Chapter 3 Measures of Central Tendency
Chapter 4 Measures of Variability
Chapter 5 Normal Distributions
Chapter 6 Probability and Introduction to Hypothesis Testing
Chapter 7 Testing One Sample Mean
Chapter 8 Estimating the Mean of a Population
Chapter 9 Testing the Difference Between Two Means
Chapter 10 Errors in Hypothesis Testing, Statistical Power, and Effect Size
Chapter 11 One-Way Analysis of Variance (ANOVA)
Chapter 12 Two-Way Analysis of Variance (ANOVA)
Chapter 13 Correlation
Chapter 14 Linear Regression and Multiple Correlation
Chapter 15 Chi-Square
Appendix A: Tables
Appendix B: Review of Basic Mathematics
Glossary
References
Index

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

DETAILED CONTENTS

Preface

Acknowledgments

About the Author

Chapter 1 Introduction to Statistics

- 1.1 What Is Statistics?
- 1.2 Why Learn Statistics?
- 1.3 Introduction to the Stages of the Research Process
  - Developing a Research Hypothesis to Be Tested
    - Identifying a Question or Issue to Be Examined
    - Reviewing and Evaluating Relevant Theories and Research
    - Stating a Research Hypothesis: Independent and Dependent Variables
  - Collecting Data
    - Drawing a Sample From a Population
    - Determining How Variables Will Be Measured: Levels of Measurement
    - Selecting a Method to Collect the Data: Experimental and Nonexperimental Research Methods
  - Analyzing the Data
    - Calculating Descriptive Statistics
    - Calculating Inferential Statistics
  - Drawing a Conclusion Regarding the Research Hypothesis
  - Communicating the Findings of the Study
- 1.4 Plan of the Book
- 1.5 Looking Ahead
- 1.6 Summary
- 1.7 Important Terms
- 1.8 Exercises

Answers to Learning Checks

Learning Check 1

Learning Check 2

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Odd-Numbered Exercises

Chapter 2 Examining Data: Tables and Figures

2.1 An Example From the Research: Winning the Lottery

2.2 Why Examine Data?

Gaining an Initial Sense of the Data

Identifying Outliers

Evaluating Research Methodology

Determining Whether Data Meet Statistical Criteria and Assumptions

2.3 Examining Data Using Tables

What Are the Values of the Variable?

How Many Participants in the Sample Have Each Value of the Variable?

What Percentage of the Sample Has Each Value of the Variable?

2.4 Grouped Frequency Distribution Tables

Cumulative Percentages

2.5 Examining Data Using Figures

Displaying Nominal and Ordinal Variables: Bar Charts and Pie Charts

Displaying Interval and Ratio Variables: Histograms and Frequency Polygons

Drawing Inappropriate Conclusions From Figures

2.6 Examining Data: Describing Distributions

Modality

Symmetry

Variability

2.7 Looking Ahead

2.8 Summary

2.9 Important Terms

2.10 Formula Introduced in This Chapter

Percentage (%)

2.11 Using IBM SPSS Software *

Creating Data Files: Defining Variables and Entering Data

Examining Data: Frequency Distribution Tables and Figures: The Lottery Study (2.1)

2.12 Exercises

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 3 Measures of Central Tendency
3.1 An Example From the Research: The 10% Myth
3.2 Understanding Central Tendency
3.3 The Mode
Determining the Mode
3.4 The Median
Calculating the Median: Odd Number of Scores
Calculating the Median: Even Number of Scores
Calculating the Median When Multiple Scores Have the Median Value
3.5 The Mean
Calculating the Mean
Reporting Calculated Numbers
Rounding Calculated Numbers
Calculating the Mean From Frequency Distribution Tables
The Mean as a Balancing Point in a Distribution
The Sample Mean vs. the Population Mean: Statistics vs. Parameters
3.6 Comparison of the Mode, Median, and Mean
Strengths and Weaknesses of the Mode
Strengths and Weaknesses of the Median
Strengths and Weaknesses of the Mean
Choosing a Particular Measure of Central Tendency
3.7 Measures of Central Tendency: Drawing Conclusions
3.8 Looking Ahead
3.9 Summary
3.10 Important Terms
3.11 Formulas Introduced in This Chapter
Median (Odd Number of Scores)
Median (Even Number of Scores)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Mean
Mean (Frequency Distribution Table)
Mean (Balancing Point in a Distribution)
3.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 4 Measures of Variability
4.1 An Example From the Research: How Many Sometimes" in an Always"?
4.2 Understanding Variability
4.3 The Range
Strengths and Weaknesses of the Range
4.4 The Interquartile Range
Strengths and Weaknesses of the Interquartile Range
4.5 The Variance (s^2)
Definitional Formula for the Variance
Computational Formula for the Variance
Why Not Use the Absolute Value of the Deviation in Calculating the Variance?
Why Divide by N − 1 Rather Than N in Calculating the Variance?
4.6 The Standard Deviation (s)
Definitional Formula for the Standard Deviation
Computational Formula for the Standard Deviation
Eyeball Estimating" the Mean and Standard Deviation
4.7 Measures of Variability for Populations
The Population Variance (σ^2)
The Population Standard Deviation (σ)
Measures of Variability for Samples vs. Populations
4.8 Measures of Variability for Nominal Variables
The Index of Qualitative Variation
4.9 Measures of Variability: Drawing Conclusions

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

4.10 Looking Ahead
4.11 Summary
4.12 Important Terms
4.13 Formulas Introduced in This Chapter
Range
Interquartile Range
Variance (s^2) (Definitional Formula)
Variance (s^2) (Computational Formula)
Standard Deviation (s) (Definitional Formula)
Standard Deviation (s) (Computational Formula)
Population Variance (σ^2) (Definitional Formula)
Population Standard Deviation (σ) (Definitional Formula)
Index of Qualitative Variation (IQV)
4.14 Using SPSS
Calculating Measures of Central Tendency and Variability: The Frequency Rating Study (4.1)
4.15 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 5 Normal Distributions
5.1 Normal Distributions
What Are Normal Distributions?
Why Are Normal Distributions Important?
Is There Only One Normal Distribution?
5.2 The Standard Normal Distribution
What Is the Standard Normal Distribution?
Interpreting z-Scores: Distance From the Mean
Interpreting z-Scores: Relative Position Within the Distribution
Finding the Area Between the Mean and a z-Score

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Finding the Area Less or Greater Than a z-Score
Finding the Area Between Two z-Scores
5.3 Applying z-Scores to Normal Distributions
Transforming Normal Distributions Into the Standard Normal Distribution
Understanding the z-Score Formula
Example: Transforming SAT Scores Into z-Scores
Interpreting z-Scores: Distance From the Mean
Interpreting z-Scores: Relative Position Within Distribution
Finding the Area Between the Mean and a Score
Finding the Area Less or Greater Than a Score
Finding the Area Between Two Scores
Transforming Scores Into z-Scores: Drawing Conclusions
5.4 Standardizing Frequency Distributions
Example: Measuring Politicians' Behavior
Standardizing a Frequency Distribution
Interpreting Standardized Scores
Using Standardized Scores to Compare or Combine Variables
Characteristics of Standardized Distributions
When Can You Apply the Normal Curve Table to a Standardized Distribution?
5.5 Looking Ahead
5.6 Summary
5.7 Important Terms
5.8 Formulas Introduced in This Chapter
z-Score
Standardized Score
5.9 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 6 Probability and Introduction to Hypothesis Testing

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

6.1 A Brief Introduction to Probability
- What Is Probability?
- Why Is Probability Important to Researchers?
- Applying Probability to Normal Distributions
- Applying Probability to Binomial Distributions

6.2 Example: Making Heads or Tails of the Super Bowl

6.3 Introduction to Hypothesis Testing
- State the Null and Alternative Hypotheses (H_0 and H_1)
  - The Null Hypothesis (H_0)
  - The Alternative Hypothesis (H_1)
- Make a Decision About the Null Hypothesis
  - Defining a Low" Probability of a Statistic: Setting Alpha (α)
  - Identifying the Values of the Statistic With a Low Probability
  - Stating a Decision Rule
  - Calculating a Value of a Statistic
  - Making the Decision Whether to Reject the Null Hypothesis
- Draw a Conclusion From the Analysis
- Relate the Result of the Analysis to the Research Hypothesis
- Summary of Steps in Hypothesis Testing

6.4 Issues Related to Hypothesis Testing: An Introduction
- The Issue of Proof" in Hypothesis Testing
- Errors in Decision Making
- Factors Influencing the Decision About the Null Hypothesis
  - Sample Size
  - Alpha (α)
  - Directionality of the Alternative Hypothesis

6.5 Looking Ahead

6.6 Summary

6.7 Important Terms

6.8 Formula Introduced in This Chapter

Probability

6.9 Exercises

Answers to Learning Checks

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 7 Testing One Sample Mean
7.1 An Example From the Research: Do You Read Me?
7.2 The Sampling Distribution of the Mean
Characteristics of the Sampling Distribution of the Mean
7.3 Inferential Statistics: Testing One Sample Mean (σ Known)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: z-Test for One Mean
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the z-Test for One Mean
Summary
7.4 A Second Example From the Research: Unique Invulnerability
7.5 Introduction to the t-Distribution
Characteristics of the Student t-Distribution
7.6 Inferential Statistics: Testing One Sample Mean (σ Not Known)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: t-Test for One Mean
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the t-Test for One Mean

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Summary
7.7 Factors Affecting the Decision About the Null Hypothesis
Sample Size
Alpha (α)
Directionality of the Alternative Hypothesis
7.8 Looking Ahead
7.9 Summary
7.10 Important Terms
7.11 Formulas Introduced in This Chapter
z-Test for One Mean
Population Standard Error of the Mean $\left(\mathbf{i}t\sigma_{\mathbf{i}t\overline{X}}\right)$
Degrees of Freedom (df) for One Sample
t-Test for One Mean
Standard Error of the Mean $\left(\mathbf{i}ts_{\mathbf{i}t\overline{X}}\right)$
7.12 Using SPSS
Testing One Sample Mean (σ Not Known): The Unique Invulnerability Study (7.6)
7.13 Exercises
Answers to Learning Checks
Learning Check 2
Learning Check 3
Learning Check 5
Learning Check 6
Answers to Odd-Numbered Exercises
Chapter 8 Estimating the Mean of a Population
8.1 An Example From the Research: Salary Survey
8.2 Introduction to the Confidence Interval for the Mean
Using the Sampling Distribution of the Mean to Create the Confidence Interval for the Mean
8.3 The Confidence Interval for the Mean (σ Not Known)
State the Desired Level of Confidence
Calculate the Confidence Interval and Confidence Limits
Calculate the Standard Error of the Mean $\left(s_{\overline{X}}\right)$

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Calculate the Degrees of Freedom (df) and Identify the Critical Value of t

Calculate the Confidence Interval and Confidence Limits

Draw a Conclusion About the Confidence Interval

Probability of an Interval or the Probability of a Population Mean?

Summary

8.4 The Confidence Interval for the Mean (σ Known)

State the Desired Level of Confidence

Calculate the Confidence Interval and Confidence Limits

Calculate the Population Standard Error of the Mean $\left(\sigma_{\overline{X}}\right)$

Identify the Critical Value of z

Calculate the Confidence Interval and Confidence Limits

Draw a Conclusion About the Confidence Interval

Summary

8.5 Factors Affecting the Width of the Confidence Interval for the Mean

Sample Size

Estimating the Sample Size Needed for a Desired Confidence Interval

Level of Confidence

Choosing Between Different Levels of Confidence: Probability Versus Precision

8.6 Interval Estimation and Hypothesis Testing

Concerns About Hypothesis Testing

The Interpretation of Statistical Significance and Nonsignificance

The Influence of Sample Size on the Decision About the Null Hypothesis

Interval Estimation as an Alternative to Hypothesis Testing

8.7 Looking Ahead

8.8 Summary

8.9 Important Terms

8.10 Formulas Introduced in This Chapter

Confidence Interval for the Mean (σ Not Known)

Confidence Interval for the Mean (σ Known)

Estimated Sample Size for the Confidence Interval for the Mean

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

8.11 Using SPSS
Calculating the confidence interval for the mean (The salary survey study (8.1))
8.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 9 Testing the Difference Between Two Means
9.1 An Example From the Research: You Can Just Wait
9.2 The Sampling Distribution of the Difference
Characteristics of the Sampling Distribution of the Difference
9.3 Inferential Statistics: Testing the Difference Between Two Sample Means
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: t-Test for Independent Means
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the t-Test for Independent Means
Summary
9.4 Inferential Statistics: Testing the Difference Between Two Sample Means (Unequal Sample Sizes)
An Example From the Research: Kids' Motor Skills and Fitness
Inferential Statistics: Testing the Difference Between Two Sample Means (Unequal Sample Sizes)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Relate the Result of the Analysis to the Research Hypothesis

Assumptions of the t-Test for Independent Means (Unequal Sample Sizes)

Summary

9.5 Inferential Statistics: Testing the Difference Between Paired Means

Research Situations Appropriate for Within-Subjects Research Designs

An Example From the Research: Web-Based Family Interventions

Inferential Statistics: Testing the Difference Between Paired Means

Calculate the Difference Between the Paired Scores

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Draw a Conclusion From the Analysis

Relate the Result of the Analysis to the Research Hypothesis

Assumptions of the t-Test for Dependent Means

Summary

9.6 Looking Ahead

9.7 Summary

9.8 Important Terms

9.9 Formulas Introduced in This Chapter

Degrees of Freedom (df), Difference Between Two Sample Means

t-Test for Independent Means

Standard Error of the Difference $\left(s_{\overline{X}_1-\overline{X}_2}\right)$

Standard Error of the Difference (Unequal Sample Sizes)

Degrees of Freedom (df), Difference Between Paired Means

t-Test for Dependent Means

Standard Error of Difference Scores (s D)

9.10 Using SPSS

Testing the Difference Between Two Sample Means: The Parking Lot Study (9.1)

Testing the Difference Between Paired Means: The Web-Based Intervention Study (9.5)

9.11 Exercises

Answers to Learning Checks

Learning Check 1

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Answers to Odd-Numbered Exercises
Chapter 10 Errors in Hypothesis Testing, Statistical Power, and Effect Size
10.1 Hypothesis Testing vs. Criminal Trials
10.2 An Example From the Research: Truth or Consequences
Correct Decisions and Mistakes in Criminal Trials
10.3 Two Errors in Hypothesis Testing: Type I and Type II Error
Type I Error: The Risk in Rejecting the Null Hypothesis
What Is the Probability of Making a Type I Error?
What Is the Probability of Not Making a Type I Error?
Why Is Type I Error a Concern?
Type II Error: The Risk in Not Rejecting the Null Hypothesis
What Is the Probability of Making a Type II Error?
What Is the Probability of Not Making a Type II Error?
Why Is Type II Error a Concern?
Why Does Type II Error Occur?
Type I and Type II Error: Summary
10.4 Controlling Type I and Type II Error
Controlling Type I Error
Concerns About Controlling Type I Error
Controlling Type II Error
Increasing Sample Size
Raising Alpha (α)
Using a Directional Alternative Hypothesis (H1)
Increasing Between-Group Variability
Decreasing Within-Group Variability
Controlling Type I and Type II Error: Summary
10.5 Measures of Effect Size
Interpreting Inferential Statistics: The Influence of Sample Size
The Inappropriateness of “Highly Significant”

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Effect Size as Variance Accounted for

Measure of Effect Size for the Difference Between Two Sample Means: r^2

Interpreting Measures of Effect Size

Presenting Measures of Effect Size

Reasons for Calculating Measures of Effect Size

A Second Measure of Effect Size for the Difference Between Two Sample Means: Cohen's d

10.6 Looking Ahead

10.7 Summary

10.8 Important Terms

10.9 Formulas Introduced in This Chapter

Probability of Type I Error

Probability of Type II error

r Squared (r^2)

Cohen's d

10.10 Exercises

Answers to Learning Checks

Learning Check 1

Learning Check 2

Learning Check 3

Learning Check 4

Answers to Odd-Numbered Exercises

Chapter 11 One-Way Analysis of Variance (ANOVA)

11.1 An Example From the Research: It's Your Move

11.2 Introduction to Analysis of Variance (ANOVA)

Review: Testing the Difference Between the Means of Two Samples

Understanding Between-Group and Within-Group Variability

Relating Between-Group and Within-Group Variability

Introduction to Analysis of Variance (ANOVA) and the F-Ratio

Characteristics of the F-Ratio Distribution

11.3 Inferential Statistics: One-Way Analysis of Variance (ANOVA)

State the Null and Alternative Hypotheses (H_0 and H_1)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Value, and State a Decision Rule
Calculate a Statistic: F-Ratio for the One-Way ANOVA
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (R^2)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Summary
11.4 A Second Example: The Parking Lot Study Revisited
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Value, and State a Decision Rule
Calculate a Statistic: F-Ratio for the One-Way ANOVA
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (R^2)
Draw a Conclusion From the Analysis
The Relationship Between the t-Test and the F-Ratio
11.5 Analytical Comparisons Within the One-Way ANOVA
Planned Versus Unplanned Comparisons
Between-Group and Within-Group Variance in Analytical Comparisons
Conducting Planned Comparisons
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Conducting Unplanned Comparisons
Concerns Regarding Unplanned Comparisons
Methods of Controlling Familywise Error

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Controlling Familywise Error: The Tukey Test

11.6 Looking Ahead

11.7 Summary

11.8 Important Terms

11.9 Formulas Introduced in This Chapter

Degrees of Freedom for Between-Group Variance, One-Way ANOVA (df_{BG})

Degrees of Freedom for Within-Group Variance, One-Way ANOVA (df_{WG})

Between-Group Variance, One-Way ANOVA (MS_{BG})

Total Sample Mean, One-Way ANOVA $\left(\overline{X}_T\right)$

Within-Group Variance, One-Way ANOVA (MS_{WG})

F-Ratio, One-Way ANOVA (F)

Measure of Effect Size, One-Way ANOVA (R^2)

Relationship Between t-Test and F-Ratio for the One-Way ANOVA

F-Ratio, Analytical Comparison (F_{comp})

Between-Group Variance, Analytical Comparison $\left(MS_{\mathrm{BG}_{\mathrm{comp}}}\right)$

Measure of Effect Size, Analytical Comparison $\left(R^2_{\mathrm{comp}}\right)$

Critical Value, Tukey Test for Unplanned Comparisons (F_T)

11.10 Using SPSS

One-Way Analysis of Variance (ANOVA): The Chess Study (11.1)

Analytical Comparisons: Planned Comparisons—Explanation vs. Prediction (11.5)

11.11 Exercises

Answers to Learning Checks

Learning Check 2

Learning Check 3

Learning Check 4

Answers to Odd-Numbered Exercises

Chapter 12 Two-Way Analysis of Variance (ANOVA)

12.1 An Example From the Research: Vote—or Else!

12.2 Introduction to Factorial Research Designs

Testing Interaction Effects

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

An Example of Interaction Effect

An Example of No Interaction Effect

Testing Main Effects

The Relationship Between Main Effects and Interaction Effects

Factorial Designs vs. Single-Factor Designs: Advantages and Disadvantages

12.3 The Two-Factor (A × B) Research Design

Notational System for Components of the Two-Factor Research Design

Analyzing the Two-Factor Research Design: Descriptive Statistics

Notational System for Means in the Two-Factor Research Design

Examining the Descriptive Statistics

12.4 Introduction to Analysis of Variance (ANOVA) for the Two-Factor Research Design

F-Ratios in the Two-Factor Research Design

12.5 Inferential Statistics: Two-Way Analysis of Variance (ANOVA)

State the Null and Alternative Hypotheses (H_0 and H_1)

Main Effect of Factor A

Main Effect of Factor B

A × B Interaction Effect

Make Decisions About the Null Hypotheses

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Values, and State Decision Rules

Calculate Statistics: F-Ratios for the Two-Way ANOVA

Make Decisions Whether to Reject the Null Hypotheses

Determine the Levels of Significance

Calculate Measures of Effect Size (R^2)

Draw Conclusions From the Analyses

Relate the Results of the Analyses to the Research Hypothesis

12.6 Investigating a Significant A × B Interaction Effect: Analysis of Simple Effects

12.7 Looking Ahead

12.8 Summary

12.9 Important Terms

12.10 Formulas Introduced in This Chapter

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Degrees of Freedom for Main Effect of Factor A, Two-Way ANOVA (df_A)
Degrees of Freedom for Main Effect of Factor B, Two-Way ANOVA (df_B)
Degrees of Freedom for A × B Interaction Effect, Two-Way ANOVA ($df_{A\times B}$)
Degrees of Freedom for Within-Group Variance, Two-Way ANOVA (dfWG)
Between-Group Variance for Main Effect of Factor A, Two-Way ANOVA (MS_A)
Between-Group Variance for Main Effect of Factor B, Two-Way ANOVA (MS_B)
Sums of Squares for Between-Group Variance, Two-Way ANOVA (SS_{BG})
Between-Group Variance for A × B Interaction, Two-Way ANOVA ($MS_{A\times B}$)
Within-Group Variance, Two-Way ANOVA (MS_{WG})
F-Ratio for Main Effect of Factor A, Two-Way ANOVA (F_A)
F-Ratio for Main Effect of Factor B, Two-Way ANOVA (F_B)
F-Ratio for A × B Interaction, Two-Way ANOVA ($F_{A\times B}$)
Measure of Effect Size for Main Effect of Factor A, Two-Way ANOVA(R^2_A)
Measure of Effect Size for Main Effect of Factor B, Two-Way ANOVA(R^2_B)
Measure of Effect Size for A × B Interaction, Two-Way ANOVA(R^2_{AB})
12.11 Using SPSS
Two-Way Analysis of Variance (ANOVA): The Voting Message Study (12.1)
12.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 13 Correlation
13.1 An Example From the Research: It's Good for You!
13.2 Introduction to the Concept of Correlation
Describing the Relationship Between Variables
Nature of the Relationship
Direction of the Relationship
Strength of the Relationship

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Measuring the Relationship Between Variables: Correlational Statistics
Introduction to the Pearson Correlation Coefficient (r)
Calculating Correlational Statistics: The Role of Variance and Covariance
13.3 Inferential Statistics: Pearson Correlation Coefficient
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: Pearson Correlation Coefficient (r)
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Calculate a Measure of Effect Size (r^2)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the Pearson Correlation
A Computational Formula for the Pearson r
Represent the Covariance Between X and Y (SPXY)
Represent the Variance of Variable X (SSX)
Represent the Variance of Variable Y (SSY)
Calculate the Pearson Correlation Coefficient (r)
13.4 Correlating Two Sets of Ranks: The Spearman Rank-Order Correlation
Introduction to Ranked Variables
An Example From the Research: Do Students and Teachers Think Alike?
Inferential Statistics: The Spearman Rank-Order Correlation Coefficient
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
13.5 Correlational Statistics vs. Correlational Research
Does ANOVA = Causality?
“Correlation Is Not Causality”—What Does That Mean?
13.6 Looking Ahead

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

13.7 Summary
13.8 Important Terms
13.9 Formulas Introduced in This Chapter
Degrees of Freedom for Pearson Correlation (df)
Sum of Products (Definitional Formula) (SPXY)
Sum of Squares for the X Variable (Definitional Formula) (SSX)
Sum of Squares for the Y Variable (Definitional Formula) (SSY)
Pearson Correlation (Definitional Formula) (r)
Measure of Effect Size for the Pearson Correlation Coefficient (r^2)
Sum of Products (Computational Formula) (SPXY)
Sum of Squares for the X Variable (Computational Formula) (SSX)
Sum of Squares for the Y Variable (Computational Formula) (SSY)
Pearson Correlation (Computational Formula) (r)
Spearman Rank-Order Correlation (rs)
Measure of Effect Size for the Spearman Rank-Order Correlation $\left(r_s^2\right)$
13.10 Using SPSS
Pearson Correlation: The Picky Eater Study (13.1)
Spearman Rank-Order Correlation: The Business Student Study (13.4)
13.11 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Answers to Odd-Numbered Exercises
Chapter 14 Linear Regression and Multiple Correlation
14.1 Predicting One Variable From Another: Linear Regression
The Linear Regression Equation
Calculating the Linear Regression Equation
Calculate the Slope of the Equation (b)
Calculate the Y-Intercept of the Equation (a)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Report the Linear Regression Equation

Drawing the Linear Regression Equation

A Second Example: Put Down Your Pencils

Calculating the Linear Regression Equation

Drawing the Linear Regression Equation

14.2 Correlation With Two or More Predictors: Introduction to Multiple Correlation and Regression

The Limitations of Using Only One Predictor

Introduction to Multiple Correlation

Multiple Correlation With Two Predictors: An Example

Measuring the Relationship Between Two Predictors and a Criterion: The Multiple Correlation (R)

Testing the Multiple Correlation for Statistical Significance

Predicting a Criterion From Two Predictors: The Multiple Regression Equation

Conclusion

14.3 Looking Ahead

14.4 Summary

14.5 Important Terms

14.6 Formulas Introduced in This Chapter

Linear Regression Equation

Estimate of Slope (b)

Estimate of Y-Intercept (a)

Multiple Correlation Coefficient (R)

F-Ratio for the Multiple Correlation Coefficient (F)

Multiple Regression Equation

Regression Coefficient

Estimate of Y-Intercept (a) (Multiple Regression)

14.7 Using SPSS

Linear Regression: The Picky Eater Study (14.1)

Multiple Regression: The College Admission Example (14.2)

14.8 Exercises

Answers to Learning Checks

Learning Check 1

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 2

Answers to Odd-Numbered Exercises

Chapter 15 Chi-Square

15.1 An Example From the Research (One Categorical Variable): Are You My Type?

15.2 Introduction to the Chi-Square Statistic

Observed and Expected Frequencies: The Basis of the Chi-Square Statistic

Characteristics of the Distribution of Chi-Square Statistics

15.3 Inferential Statistic: Chi-Square Goodness-of-Fit Test

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Value, and State a Decision Rule

Calculate a Statistic: Chi-Square (χ^2)

Make a Decision Whether to Reject the Null Hypothesis

Determine the Level of Significance

Calculate a Measure of Effect Size (Cramér's φ)

Draw a Conclusion From the Analysis

Relate the Result of the Analysis to the Research Hypothesis

Assumptions Underlying the Chi-Square Statistic

Conducting the Chi-Square Goodness-of-Fit Test With Unequal Hypothesized Proportions

Calculate Expected Frequencies (f_e)

Calculate the Chi-Square Statistic (χ^2)

15.4 An Example From the Research (Two Categorical Variables): Seeing Red

15.5 Inferential Statistic: Chi-Square Test of Independence

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Calculate the Degrees of Freedom (df)

Set Alpha (α), Identify the Critical Value, and State a Decision Rule

Calculate a Statistic: Chi-square (χ^2)

Make a Decision Whether to Reject the Null Hypothesis

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Determine the Level of Significance
Calculate a Measure of Effect Size (Cramér's φ)
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
15.6 Parametric and Nonparametric Statistical Tests
Parametric vs. Nonparametric Statistical Tests
Reasons for Using Nonparametric Statistical Tests
Examples of Nonparametric Statistical Tests
15.7 Looking Ahead
15.8 Summary
15.9 Important Terms
15.10 Formulas Introduced in This Chapter
Degrees of Freedom for the Chi-Square Goodness-of-Fit Test
Expected Frequency for the Chi-Square Goodness-of-Fit Test
Chi-Square Statistic (χ^2)
Cramér's φ for the Chi-Square Goodness-of-Fit Test
Degrees of Freedom for the Chi-Square Test of Independence
Expected Frequency for the Chi-Square Test of Independence
Cramér's φ for the Chi-Square Test of Independence
15.11 Using SPSS
Chi-Square Goodness-of-Fit Test—Equal Hypothesized Proportions: The Personality Study (15.1)
Chi-Square Goodness-of-Fit Test—Unequal Hypothesized Proportions: The Personality Study (15.3)
Chi-Square Test of Independence: The Color Preference Study (15.4)
Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Learning Check 6

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Odd-Numbered Exercises

Appendix A: Tables

Appendix B: Review of Basic Mathematics

I. Symbols

II. Terms

III. Addition, Subtraction, Multiplication, Division

A. All Positive Numbers

B. Adding All Negative Numbers

C. Adding a Combination of Positive and Negative Numbers

D. Subtracting Negative Numbers

E. Multiplying Two Negative Numbers

F. Multiplying a Combination of Positive and Negative Numbers

G. Dividing Two Negative Numbers

H. Dividing One Positive Number and One Negative Number

IV. Order of Operations

V. Fractions, Decimals, Percentages

A. Fraction

B. Decimal

C. Percentage

VI. Solving Simple Equations With One Unknown (X)

A. X Has a Value Subtracted From It (Solve for X Using Addition)

B. X Has a Value Added to It (Solve for X Using Subtraction)

C. X Is Multiplied by a Value (Solve for X Using Division)

D. X Is Divided by a Value (Solve for X Using Multiplication)

VII. Solving More Complex Equations With One Unknown (X)

A. X Is Multiplied by a Value and Then Has a Value Either Added to It or Subtracted From It (Solve for X Using Either Subtraction or Addition and Then Division)

B. X Is Divided by a Value and Then Has a Value Either Added to It or Subtracted From It (Solve for X Using Either Addition or Subtraction and Then Multiplication)

C. X Has a Value Added to It or Subtracted From It and Is Then Multiplied by a Value (Solve for X Using Division and Then Either Subtraction or Addition)

D. X Has a Value Either Added to It or Subtracted From It and Is Then Divided by a Value (Solve for X Using Multiplication and Then Either Subtraction or

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Addition)
Glossary
References
Index

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

PREFACE

A Few (?) Words About Me

When I entered college, I thought my path would take me to law school; however, taking an Intro to Psych class made me realize I enjoy the challenge of trying to understand the human mind. The college I attended, U.C. Santa Cruz, was a fairly unconventional university, but somehow in the midst of a sea of humanistic psychologists I became attracted to the empirical and methodological aspects of psychology by taking classes from instructors such as David Harrington and Dane Archer, who showed me that statistics could appeal to students if taught with a gentle, guiding approach that addresses questions relevant to students' lives. After graduating from college, I was able to get a job as a research assistant for a human resource consulting firm. Despite my lack of work experience, I was hired primarily as a result of having taken statistics and research methods courses, which taught me that learning statistics has benefits both inside and outside of the classroom.

Several years later, I started grad school at U.C. Berkeley, where two events critical to this book took place. First, serving as a teaching assistant, I found I really enjoyed helping students, particularly in statistics and research methods classes often approached with fear and suspicion. Second, I took graduate classes from Geoff Keppel, who had developed his own unique system for analyzing experimental research designs. His lectures and books showed me that statistics can be taught in a systematic way that highlights similarities rather than differences between different research situations. Geoff managed to transform something as daunting-sounding as a "3 x 2 x 4 research design" into the mathematical equivalent of playing with wooden toy alphabet blocks labeled "A," "B," and "C."

After leaving Cal, I was hired by San Jose State, where I teach an introductory statistics course aimed at students with a wide range of backgrounds, ability, and motivation. When I went to select a textbook for the course, for the first time I looked carefully at the wide range of offerings. What I found striking (and still find striking) was that the majority of books focused on providing formulas and very small sets of data designed to demonstrate how to correctly calculate the correct numbers from these formulas. Little emphasis, however, was given to what these numbers meant or implied. Given my own experiences learning and using statistics, I thought a book was needed that focusses on how statistics may be used to answer questions and test ideas within the larger research process.

Goal of this Book

This textbook, aimed primarily at students majoring in the behavioral and social sciences, includes many of the same topics and chapters as the majority of existing introductory statistics books. However, what I believe distinguishes this book is its thematically-based approach that places statistics within the research process. The primary purpose of this

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

book is to not just teach students how to calculate statistics but how to interpret the results of statistical analyses in light of a study's research hypothesis, and to communicate one's results and interpretations to a broader audience. Hopefully, this book will not only help students understand the purpose and use of statistics, but it will also give them a greater appreciation of how research studies are conceived, conducted, and communicated.

The book is intended for students with a limited knowledge of mathematics, statistics, and research design and methodology. Some common obstacles faced by students are their math anxiety, fear of conducting arithmetic calculations, and an inability to see the relevance of the material to their daily lives. To address these concerns, this books attempts to judiciously use mathematical terminology, write in a conversational manner, and utilize actual research situations rather than contrived sets of numbers.

Organization of the Book

The 15 chapters of this book may be placed into three general categories. Chapters 1 to 4 are designed to introduce students to the research process and how data that have been collected may be organized, presented, and summarized:

Chapter 1 (Introduction to Statistics) defines the concept of statistics, explains why statistics are important to learn, and describes the steps in the research process that are the framework for the rest of the book.

Chapter 2 (Examining Data: Tables and Figures) discusses the importance of examining data that have been collected and describes how tables and figures may be created to represent data.

Chapter 3 (Measures of Central Tendency) presents measures of central tendency (the mode, median, and the mean) and compares them in terms of their relative strengths and weaknesses.

Chapter 4 (Measures of Variability) presents measures of variability (range, interquartile range, variance, standard deviation, index of qualitative variation) and the concept of parameters (population variance and standard deviation).

Chapters 5 through 10 discuss the process of conducting statistical analyses to test research questions and hypotheses, as well as issues and controversies regarding this process:

Chapter 5 (Normal Distributions) introduces theoretical distributions, starting with normal distributions, before moving on to the standard normal distribution (normal curve).

Chapter 6 (Probability and Introduction to Hypothesis Testing) presents the structure for later chapters by introducing, defining, and illustrating steps in hypothesis testing.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Chapter 7 (Testing One Sample Mean) tests one sample mean using the *z* -test and *t* -test and introduces sampling distributions, the central limit theorem, and the Student *t* distribution.

Chapter 8 (Estimating the Mean of a Population) introduces interval estimation by developing the confidence interval for the mean and discusses interval estimation as an alternative to null hypothesis significance testing (NHST).

Chapter 9 (Testing the Difference Between Two Means) covers the *t* -test for independent means for both equal and unequal sample size and introduces within-subjects research designs using the *t* -test for dependent means.

Chapter 10 (Errors in Hypothesis Testing, Statistical Power, and Effect Size) discusses Type I and Type II error and introduces measures of effect size as a way of drawing appropriate conclusions about an analysis.

Chapters 11 to 15 discuss different statistical procedures used in research situations that vary in the number of independent variables as well as how the variables have been measured:

Chapter 11 (One-way Analysis of Variance (ANOVA)) discusses the one-way ANOVA, the relationship between the *t* -statistic and the F-ratio, and planned and post-hoc comparisons.

Chapter 12 (Two-way Analysis of Variance (ANOVA)) introduces factorial research designs, the calculation of F-ratios for main effects and interaction effects, and illustrates the use of simple effects analyses to test research hypotheses.

Chapter 13 (Correlation) defines and illustrates the concept of correlation, discusses the Pearson and Spearman correlation coefficients, and makes the distinction between correlational statistics and correlational research methods.

Chapter 14 (Linear Regression and Multiple Correlation) introduces the concept of regression, discusses linear regression, and works through an example of multiple regression with two predictor variables.

Chapter 15 (Chi-Square) describes the analysis of categorical variables using the chi-square goodness-of-fit test and the chi-square test of independence before ending the book with a brief introduction to nonparametric statistics.

In most cases, the chapters in this book share a uniform format and structure, centered on the research process, based on examples of actual published research studies addressing topics relevant to students' lives. For each example, the research hypothesis is clearly identified; after a brief description of how the data in the study were collected, the calculation and interpretation of both descriptive and inferential statistics are presented,

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ending with the extent to which the results of the statistical analyses support the research hypothesis and how these findings may be communicated to a broader audience.

Pedagogical Features

The chapters in this textbook contain several reoccurring features designed to enhance students' comprehension of the material. These include the following:

Periodic learning checks within chapters. Short learning checks, in the form of conceptual review questions and computational exercises, are provided throughout each chapter.

Explicit presentation of formulas and calculations. This book provides both definitional formulas to highlight their relation to the concept being measured and computational formulas to facilitate calculations. In applying these formulas, all of the steps in the calculation of statistics such as the *t* -test and F-ratio are included to help students in completing exercise and homework assignments.

Illustration of the presentation and communication of analyses. Every chapter includes a discussion and illustration of how statistical analyses may be reported as well how they may be communicated to a larger audience in both written and visual form.

Highlighted definition of terms. Every term and concept introduced in this book is boldfaced, immediately followed by an explicit definition.

End of chapter summaries and exercises. The end of each chapter contains a summary of critical points, a list of terms and formulas introduced in the chapter, and a large number of exercises, many of which refer to published research studies.

Calculation of statistics using SPSS. Many chapters include end-of-chapter content that provides an annotated presentation of how statistics may be calculated using the SPSS software. Screen shots of data files, menu selections, and output files are included and annotated to locate and identify relevant aspects of the analysis.

What's New to the 3rd Edition

I'd like to start by expressing my most heartfelt appreciation to the instructors and reviewers of the first two editions of this book as their evaluation, insights, criticisms, and suggestions were absolutely vital to the changes made in this third edition. I'd like to particularly thank the instructors who've adopted the book who were kind enough to spontaneously send me their thoughts and suggestions. It is my sincere hope that all of you see the incorporation of your input in this edition.

There were two main goals in this edition. The first goal was to make the writing as brief and concise as possible, manifested in the goal of reducing the length of each chapter by 5% to

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

10%. This goal was achieved by simplifying the language and shortening or eliminating what I, the book's adopters, and the book's reviewers felt was unnecessary detail or discussion. The second goal was to increase and update the exercises at the end of many of the chapters, including the use of recent published research as the basis of these exercises. Another goal was to update the demonstration of the use of SPSS to analyze data using the most recent version (Version 31 at the time of this book's publication).

More specifically, in terms of content, the following changes were made to these chapters:

Chapter 1 (Introduction to Statistics): The published research studies used to illustrate concepts in the text, learning checks, and exercises were added to and updated using topics relevant to students (e.g., fear of missing out (FOMO), cryptocurrency, legalization of cannabis, Gen Z, COVID, fake news).

Chapter 6 (Probability and Introduction to Hypothesis Testing): The Super Bowl example used to introduce and illustrate the process of hypothesis testing was updated.

Chapter 8 (Estimating the Mean of a Population): The main research example, which involves data from an actual national salary survey, has been updated throughout this chapter with more current salary data.

As with the first two editions, any and all questions, concerns, suggestions, and complaints you have about the book are most welcome and can be sent to me at howard.tokunaga@sjsu.edu .

A Few Tips for Students

To you, the college student about to read this book as part of taking a statistics course: "Welcome!" and "Great job!" We welcome you because you're about to embark on a semester-long journey that we hope enhances your skills and widens your perspective; we congratulate you because it's a journey not everyone is willing to take. We say to our students that by taking just one statistics course they are ahead of 95% of the population in their knowledge of statistics.

We also tell our students that our primary goal for them is not to become a statistics genius but just to become comfortable with the process of analyzing and interpreting data. In that regard, we agree with a 2012 *Forbes* magazine article entitled *The Six Classes That Will Make Any College Grad Employable* that noted "a good statistics class will help the student become comfortable pushing numbers around and avoiding false interpretations."

At the present moment, we know our encouragement may be of little comfort because you might be anxious about having to learn statistics. Some of you might be anxious about having to learn statistical concepts and formulas; others might be anxious because the research process seems pretty complicated. Talking with students who have taken our courses over the years has helped us assemble the following advice:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Master the Material Presented in the Early Chapters

Chapters 1 through 6 discuss how research is typically conducted, how data are summarized and described, and the process by which researchers conduct statistical analyses to test their ideas. It is absolutely critical for you to have a firm understanding of these chapters as they lay the groundwork for later chapters that discuss a variety of statistical procedures used by researchers to test hypotheses.

What does it mean to master this material? First, *read the chapters both before and after they're discussed in class* . Reading the chapters beforehand helps you identify anything that's unclear and have your questions addressed by your instructor; re-reading the chapters afterward helps confirm your understanding of the material. Next, *be able to define and explain key conceptspresented in these chapters* . These concepts are highlighted in bold-faced letters, and it's important to learn them when they're first introduced because they'll appear throughout the remainder of the book. Next, *do the learning checks within the chapters and the exercises at the end of the chapters* , because they include exercises and review questions you can ask yourself to assess your understanding of the material. Most importantly, *do not miss class during the early part of the semester* ! It's been our experience that students who miss critical lectures at the beginning of the term often have difficulty keeping up as the semester continues.

Review High School Algebra

If you happen to move beyond the introductory statistics course for which you're reading this book, you'll find that statistics are often calculated using computers and statistical software rather than hand calculations. Consequently, some statistics textbooks no longer include mathematical formulas but instead have students use statistical software or Microsoft Excel. However, this book takes a different approach for two main reasons. First, we believe that in order to understand the results of statistical analyses, one must understand their underlying foundation. Learning statistics via computer software can lead to a "brain-dead" approach in which students are at the mercy (rather than control) of their computers. We refer to this as the "I'm only as smart as my printout" method of learning statistics, also known as "Because this is what my computer told me."

Second, this book contains mathematical formulas and calculations for one simple reason: The statistics in this book are not hard to calculate. To assess whether you're prepared to read this book, take the following test:

1. Do you know how to add, subtract, multiply, and divide?
2. Do you know how to calculate the square or square root of a number?
3. Do you understand the order of operations and parentheses within mathematical formulas?

If your answers to these questions are "Yes," congratulations! You possess the ability to conduct *every* mathematical calculation in this book because *none* of the formulas involve geometry, trigonometry, or calculus. However, if you're unsure of your mathematical ability, we suggest you review your high school algebra. Also, there's an appendix at the end of this

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

book that includes the mathematical concepts and operations needed to conduct the statistical analyses in this book. We highly recommend you read this appendix to review and assess your mathematical skills.

Learn to Use a Statistical Calculator

Although learning the statistics in this book doesn't require the use of computers or software, you will need a hand calculator. Calculators with statistical capabilities are typically identified as "statistical" or "scientific" calculators. Although there's a wide selection of brands and models, know that the calculators most appropriate for this book cost only $10 to $15. We do *not* recommend you use or purchase a more expensive calculator! We prefer these simpler calculators because students are able to do their calculations more quickly and with fewer errors. We do suggest you purchase a calculator similar to the one your instructor uses as they will be more familiar with its features and idiosyncrasies.

In selecting a calculator, look closely at the keyboard; you'll need a calculator whose keys have labels such as $\bar{X}$, σ, and ΣX (Chapters 3 and 4 discuss what these symbols represent). When you go calculator shopping, you'll find calculators that have keys with labels such as ΣY and XY. Although calculators with these Y symbols are only slightly more expensive than calculators with just the X symbols, we do *not* recommend their purchase because these features are not needed to perform the calculations in this book. In fact, we've found students with Y calculators have a more difficult time performing simpler calculations.

Ask Questions

Conducting research is the process of asking and answering questions. Accordingly, throughout this book issues are often framed in the form of questions, which we hope encourages you to ask questions as well. The first person you should direct any questions to is yourself. As you work your way through the chapters, get into the habit of asking yourself, "What does this mean?" A good test of your level of comprehension is whether you can explain the material to yourself in a meaningful way.

You should also be sure to work with your instructor to confirm your level of understanding and clarify unanswered questions. Instructors often say to students, "The only bad question is the one you don't ask." Don't be hesitant about asking questions in class. Instructors will tell you they use one student's question to assess an entire class's level of understanding. This is what we call the "pencil test"—when one student asks a question, we look to see how many other pencils get raised, which reflects the number of students who had the same question. Asking questions enhances the learning experience for yourself, your classmates, and your instructor.

Form a Study Group

No matter how much we encourage students to ask questions in class, it seems they hesitate for fear of drawing attention to themselves. As a result, we turn to another well-

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

traveled saying: “Misery loves company.” Forming a study group with other students will help you keep up with reading assignments, confirm your comprehension of the material, clarify any confusion you have, prepare for exams, and provide social support. We do suggest that your group contain at least one person who will raise the group’s questions and concerns to the instructor during class or office hours.

Practice, Practice, Practice

The best way to learn the topics covered by this book is to obtain as much practice as possible. Throughout the chapters are examples in which the calculations have been worked out for you; in the middle and end of each chapter are a number of exercises. The answers to some of the exercises are provided at the back of this book, and your instructor has access to the answers to the other exercises. Do the end-of-chapter exercises even if your instructor does not assign them as homework in order to identify any reoccurring mistakes you make. We find that the vast majority of errors students make are simple computational errors, committed when students do the calculations too quickly.

Also note that the publisher of this book, Sage Publications, has developed a webpage for this book that contains a huge amount of supplemental tools such as videos and quizzes for each chapter. Find it by going to the Sage Publishing website **collegepublishing.sagepub.com** , searching for my name (Tokunaga), and looking for the **Resources** tab.

Enough cheerleading and preaching – let’s get started!

WELCOME!

GREAT JOB!

Additional Teaching Resources

Visit **collegepublishing.sagepub.com** , and navigate to the Resources tab on your book’s page to find the teaching materials designed to accompany this textbook. On this site you will find an array of materials that will save you time and to help you keep students engaged, including the following:

Learning management system cartridges that easily integrate with your course management system so student test results and graded assignments seamlessly flow into your gradebook

Test banks , aligned to Bloom’s Taxonomy, that provide a diverse range of test items, including multiple choice, true/false, and essay questions

Lecture notes that provide an outline and the key concepts in each chapter to aid in lecture preparation

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

PowerPoint ® slides that offer a flexible, accessible, and customizable solution for creating multimedia lectures

Tables and Figures from the book to support lecture preparation and class discussions

Sample course syllabi , including suggested models for structuring your course

Answers to even-numbered questions from the text help facilitate grading

SPSS datasets for use with exercises from the text.

Multimedia content , including original SAGE videos featuring tutorials with author Howard T. Tokunaga that bring concepts to life and appeal to diverse learners

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ACKNOWLEDGMENTS

This book has traveled a long journey, with many people playing both transient and persevering roles in its development. First and foremost, I'd like to thank my kids, Meagan and Will, for providing me perspective and inspiration amid all of the twists and turns we faced while this book was being written. I'm so very blessed to have you as my "burden." Away from home, I'm grateful to my graduate students at San Jose State—working on this book has given me much empathy for their thesis-related travails. However, they should be aware that the completion of this book ends the unspoken agreement that I not ask, "So how's your thesis?" if they don't ask, "So how's your book?" A number of friends have provided most welcome relief after long days staring at a computer screen, with particular thanks going to Ellie, Eileen, Susan, Bonnie, Ann, Lisa, and Terry. I would like to thank Charles Linsmeier and David Shirley for their many helpful comments and advice on earlier versions of this book, Julie West for her help in developing chapter exercises, and Youn-Jeong Choi and Naomi Ran for their careful examination of the chapters and exercises.

I would like to express my appreciation to everyone at SAGE who has played a role in shepherding this book to completion. First and foremost, I must express my gratitude to my editor, Vicki Knight. It has been a joy and an honor to work with someone with such a truly exceptional combination of experience, knowledge, perspective, and humor. In production I was fortunate enough to work with others on the SAGE team who made the process a pleasure: Yvonne McDuffee, editorial assistant; Katie Guarino, assistant editor; Nicole Elliott, executive marketing manager; Candice Harman, cover designer; and the copyeditor, Michelle Ponce. Furthermore, I am grateful to the reviewers who took the time to read the chapters, raise questions, and provide constructive comments and suggestions to assist me in clarifying and strengthening the manuscript. The following are the reviewers for third edition:

Erin M. Fekete, University of Indianapolis

Brooke Magnus, Boston College

Loretta Neal McGregor; Arkansas State University

Michael T. Moore, Adelphi University

Megan M. Ringel, St. Edward's University

Bettina Viereck, University of Hartford

Janice Weaver, Community College of Allegheny County

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

ABOUT THE AUTHOR

Howard T. Tokunaga is professor of Psychology at San Jose State University, where he serves as coordinator of the MS Program in Industrial/Organizational (I/O) Psychology and teaches undergraduate and graduate courses in statistics, research methods, and I/O psychology. He received his bachelor's degree in psychology at U.C. Santa Cruz and his PhD in psychology at U.C. Berkeley. In addition to his teaching, he has consulted with a number of public sector and private sector organizations on a wide variety of management and human resource issues. He is author of *Moving from IBM SPSS to R and RStudio: A Statistics Companion* , and coauthor (with G. Keppel) of *Introduction to Design and Analysis* : *A Student's Handbook* . In his spare time, he enjoys the outdoors, museums, live theater, and being as physically, socially, and politically active as possible.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CHAPTER 1 INTRODUCTION TO STATISTICS

Learning Objectives

1.1 Define the concept of statistics.

1.2 Describe different reasons for learning statistics.

1.3 Describe the main steps in the research process and the tasks and concepts within each step.

In introducing this book to you, we assume you're a college student taking what is perhaps your first course in statistics to fulfill a requirement for your major or a general education requirement. If so, you may be asking yourself two questions:

What is statistics?

Why learn statistics?

The ultimate goal of this book is to help you begin to answer these two questions.

1.1 What Is Statistics?

You may not know it, but you encounter a variety of "statistics" in your day-to-day activities: the typical cost of going to college, the yearly income of the average college graduate, the average price of a car, and so on. So what exactly is "statistics"?

The Merriam-Webster dictionary defines **statistics** as *a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data* . When most people think about statistics, they focus on the "analysis" aspect of this definition—that is to say, they focus on numbers that result from analyzing data. However, statistics not only is concerned about how data are *analyzed* but also how data are *collected* and how the results of statistical analyses are *interpreted* and *communicated* . The main purpose of this book is to introduce, describe, and illustrate the role of statistics within the larger research process.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1.2 Why Learn Statistics?

There are many reasons why you should learn statistics. First, not only do you currently encounter statistics in your daily activities, but *throughout your life you are affected by the results of research and statistical analyses* . Which college you attend is based in part on test scores developed by researchers. You may have to take a personality or intelligence test to get a job. The choices of drugs and medicines available to you are based on medical research. Learning about statistics will help you become a more informed and aware consumer of research and statistical analyses that affect many aspects of your life.

A second reason for learning statistics is that *you may be required to interpret the results of statistical analyses* . Many college courses require students to read academic research journal articles. Evaluating published research is complicated by the fact that different people studying the same topic may come up with diverse or even opposing conclusions. Understanding statistics and their role in the research process will help you decide whether conclusions drawn in research articles are appropriate and justified.

Another reason for learning statistics is that *you might be required to do your own research* . College courses sometimes have students design and conduct mini-research studies; undergraduate majors might encourage students to do senior honors theses; graduate programs often require masters' theses and doctoral dissertations. Learning to collect and analyze data will help you address your own questions in an objective, systematic manner.

A final reason for learning statistics is that it may *help you in your future career* . The website Careercast.com conducts an annual survey in which they evaluate jobs on four dimensions: environmental factors, income, employment outlook, and stress. Recently, their #1 rated job was "data scientist," and the #2 rated job was "statistician." They noted, "A statistician's skill set can be used to break down and analyze large quantities of data. The demand for these skills spans a variety of industries, including marketing, banking, government, sports, retail, and even healthcare."

It's generally a good idea for students to maintain a healthy level of curiosity or even skepticism in regard to their education. However, we find that when it comes to learning statistics, students' feelings may be characterized as fear or anxiety. We understand these feelings, and we hope the benefits of learning statistics will become clear to you and help overcome any concerns you have.

1.3 Introduction to the Stages of the Research Process

Much of scientific research involves asking questions. Throughout this book, we'll examine a broad range of questions regarding human attitudes and behavior that contemporary researchers have asked and attempted to answer. The following are research questions we'll address in this chapter to introduce the stages of the research process:

> Does appealing to one's fear of missing out (FOMO) affect the intention to invest in cryptocurrencies?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Do pro-cannabis messages on social media lead to cannabis use?

Do parents differ in their awareness of their children's online activities?

Do nonbinary college students have the same experiences as cis men or cis women?

How might you try to answer questions such as these? You could base your answers on your personal beliefs, or you could adopt the answers given to you by others. But rather than relying on subjective beliefs and feelings, researchers in many academic disciplines test their ideas using the scientific method. The **scientific method** is a method of investigation that uses the objective and systematic collection and analysis of empirical data to test theories and hypotheses.

At its simplest, this book will portray the scientific method as consisting of five main steps or phases:

developing a research hypothesis to be tested,

collecting data,

analyzing the data,

drawing a conclusion regarding the research hypothesis, and

communicating the findings of the study.

Accomplishing each of these five steps requires completing a number of tasks, as shown in Figure 1.1 . Because this sequence of steps serves as the model for the research process we'll discuss throughout this book, each step is briefly introduced in the following sections. It's important to understand that the research process depicted in Figure 1.1 represents an ideal way of doing research. The "real" way, as you may discover in your own efforts, is often anything but a smooth ride but rather is filled with false starts, dead ends, and wrong turns.

Developing a Research Hypothesis to Be Tested

The initial stage—and the first step—of the research process is to develop a research hypothesis to be tested. A **research hypothesis** may be defined as a statement regarding an expected or predicted relationship between variables. A **variable** is a property or characteristic of an object, event, or person that can take on different values. One example of a variable is "U.S. state," a variable with 50 possible values (Alabama, Arkansas, etc.).

Research hypotheses are usually developed through the completion of several tasks:

identifying a question or issue to be examined,

reviewing and evaluating relevant theories and research, and

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

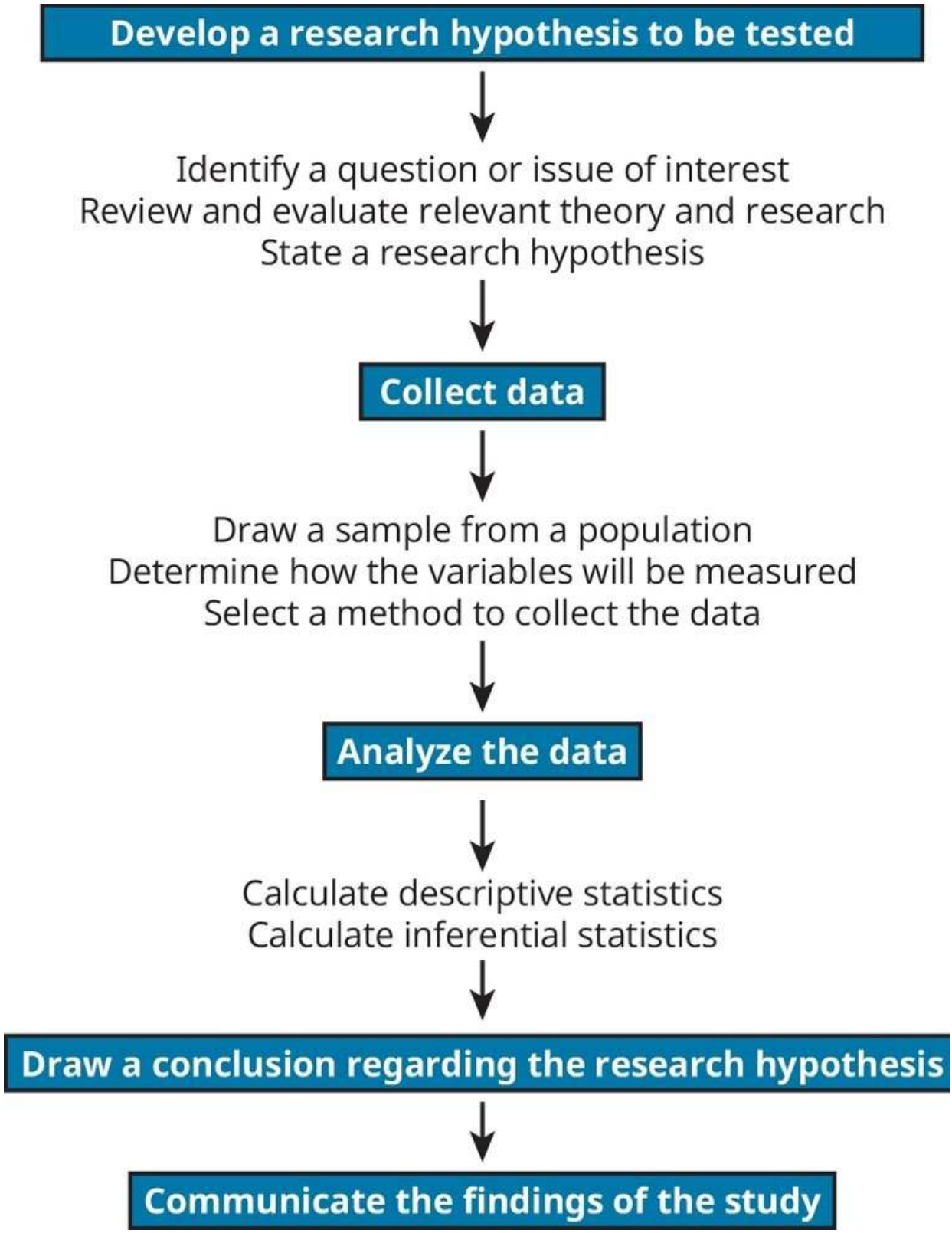
stating a research hypothesis.

Identifying a Question or Issue to Be Examined

Most research starts with a question posed by the researcher that often comes from the researcher's own ideas and daily observations. Although this may not seem terribly scientific, there's an advantage in using one's own experience as a starting point: People are generally much more motivated to explore issues that concern them personally. In teaching statistics, we advise students developing their own research projects to study something of interest to them. Conducting research can be tedious, difficult, and frustrating. At various points, you may ask yourself, "Why am I doing this?" Being able to provide a satisfactory answer to this question will help you overcome obstacles you encounter along the way.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

▶ Long Description: Figure 1.1 Steps in the Research Process Within the Scientific Method

Figure 1.1 Steps in the Research Process Within the Scientific Method

Reviewing and Evaluating Relevant Theories and Research

Beyond your own curiosity, research questions often arise from studying the theories, ideas, and research of others. A **theory** is a set of propositions used to describe or explain a phenomenon. The purpose of a theory is to summarize and explain specific facts using a logically consistent framework. Placing a question within a theoretical framework provides guidance and structure to research.

Reviewing and evaluating existing theories and research helps you decide whether it's worth the time and energy required to conduct a study. By seeing what others have done, you may conclude that the current way of thinking on a topic is incomplete or mistaken. By doing this review, you're able to ensure that the research you conduct adds to and improves upon an existing body of knowledge.

Stating a Research Hypothesis: Independent and Dependent Variables

Understanding and evaluating an existing literature not only helps develop a question of interest, but it may also lead to a predicted answer to that question. Within the scientific method, this answer is stated as a research hypothesis, defined earlier as a statement regarding an expected or predicted relationship between variables. Table 1.1 lists the research questions and research hypotheses for the studies listed at the beginning of this section. For example, the first research hypothesis states that "consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos."

One characteristic of research hypotheses such as those listed in Table 1.1 is that they identify the variables that are the focus of their research. As mentioned earlier, a variable is a property or characteristic with different values. Variables can be classified in several ways. In specifying a research hypothesis, researchers often speak in terms of "independent" and "dependent" variables. An **independent variable** may be defined as a variable manipulated or controlled by the researcher. A **dependent variable** , on the other hand, is a variable measured by the researcher that is expected to change or vary as a function of the independent variable. The independent variable is seen as the "cause" or "treatment," whereas the dependent variable is the "effect" or "outcome."

Researchers are interested in examining the effect of the independent variable on the dependent variable. Consider the first research hypothesis provided in Table 1.1 : "Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos." Here, the independent variable is the type of appeal consumers are exposed to: FOMO or non-FOMO. The dependent variable is the intention to invest in cryptos. In this study, the effect of the independent variable on the dependent variable is that

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

differences in the intention to invest in cryptos (the dependent variable) may “depend” upon which type of appeal (FOMO or non-FOMO) a consumer receives. Table 1.2 lists the independent and dependent variables for each of the research hypotheses in Table 1.1 .

In addition to identifying variables, a second characteristic of research hypotheses is that they specify the nature and direction of the relationship between variables. For example, the first research hypothesis in Table 1.1 states that “Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *higher* intention to invest in cryptos.” The word *higher* indicates the nature and direction of the relationship between the independent variable (type of appeal) and the dependent variable (intention to invest in cryptos). The direction of the relationship would not have been stated if the hypothesis had included the less specific phrase “will have a *different* intention,” which simply indicates that the two groups of consumers are not expected to be the same. Table 1.3 provides directional and nondirectional research hypotheses for the research studies from Table 1.1 .

Table 1.1 Examples of Research Questions and Hypotheses

Research Question	Research Hypothesis
Does appealing to one’s fear of missing out (FOMO) affect the intention to invest in cryptocurrencies? (Friederich et at., 2024)	Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.
Do pro-cannabis messages on social media lead to cannabis use? (Willoughby et al., 2024)	Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.
Do parents differ in their awareness of their children’s online activities? (Symons et al., 2017)	Mothers have more accurate knowledge of their children’s online activities than do fathers.
Do nonbinary college students have the same experiences as cis men or cis women? (Marx et al., 2024)	Nonbinary college students experience higher levels of victimization than do cis men or cis women.

The ability to form a directional research hypothesis is dependent on the state of the literature on the question of interest. If little or perhaps conflicting research has been conducted, researchers may not be able to form a directional hypothesis. For example, one study examining the relationship between exercise deprivation (not allowing people to exercise) and depression. reported, “We did not have a directional hypothesis, and when participants asked what we expected to find in this study, we replied: ‘We really are not sure since the results of earlier work on exercise deprivation are mixed’” (Mondin et al., 1996, p. 1200).

Table 1.2 Research Hypotheses and Their Independent and Dependent Variables

Research Hypothesis	Independent Variable	Dependent Variable
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher	Type of appeal	Intention to invest in cryptos

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Independent Variable	Dependent Variable
intention to invest in cryptos.		
Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.	Exposure to pro-cannabis messages	Intention to use cannabis
Mothers have more accurate knowledge of their children's online activities than do fathers.	Parent	Knowledge of children's online activities
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	Gender identity	Amount of victimization

Collecting Data

Once a research hypothesis has been formulated, researchers are ready to proceed to the second stage in the research process: collecting data. This step is seen as being composed of three tasks:

drawing a sample from a population,

determining how the variables will be measured, and

selecting a method by which to collect the data.

Table 1.3 Directional and Nondirectional Research Hypotheses

Directional Research Hypothesis	Nondirectional Research Hypothesis
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *higher* intention to invest in cryptos.	Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a *different* intention to invest in cryptos.
Exposure to pro-cannabis messages on social media is related to *increased* intentions to use cannabis.	Exposure to pro-cannabis messages on social media is *related to* intentions to use cannabis.
Mothers have *more* accurate knowledge of their children's online activities than do fathers.	Mothers have *different* knowledge of their children's online activities than do fathers.
Nonbinary college students experience *higher* levels of victimization than do cis men or cis women.	Nonbinary college students experience *different* levels of victimization than do cis men or cis women.

Drawing a Sample From a Population

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The first step in collecting data is to identify the group of participants to which the research hypothesis applies. The group to which the results of a study may be applied or generalized is called a population. A **population** is the total number of possible units or elements that could potentially be included in a study. Although people often think the word "population" refers to everyone in the world, researchers typically define their populations more narrowly. For example, the population for the first research hypothesis in Table 1.1 could be "consumers," "U.S. consumers," "Georgia consumers," or "University of Georgia consumers." Researchers typically try to define their populations as broadly as possibly (e.g., "consumers" rather than "University of Georgia consumers") to maximize the applications or implications of their research.

Learning Check 1

Reviewing What You've Learned So Far



1. Review questions
    a. What are the five steps involved in the research process?
    b. Why is it useful to review and evaluate theories and research before conducting a study?
    c. What are the two main characteristics of research hypotheses?
    d. What is the difference between an independent variable and a dependent variable?
2. The following are several research hypotheses from published studies. For each research hypothesis, identify the independent and dependent variable.
    a. "College students will rate instructors who dress formally (i.e., business suit and tie) as having more expertise than instructors who dress casually (i.e., slacks and shirt)" (Sebastian & Bristow, 2008, p. 197).
    b. "It was expected tha. . . . greater amounts of television viewin. . . . would predict greate. . . . posttraumatic stress symptoms" (McLeish & Del Ben, 2008, p. 417).
    c. "Do borrowers who did not complete college have greater financial distress compared with . . . non-borrowers who did not complete college [or] borrowers who earned a degree?" (Jabbari et al., 2023, p. 6).
    d. "We expected that drivers would believe that texting while driving is more detrimental to their driving performance than browsing social media" (Hashash et al., 2019, p. 70).

It's extremely difficult to collect data from all members of a population. Imagine, for example, the time and money needed to collect information from every consumer in the United States! For this reason, researchers typically draw conclusions about populations based on information collected from a sample of the population. A **sample** is a subset or portion of a

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

population. Table 1.4 describes the samples used in the four studies introduced in Table 1.1 . As you can see from Table 1.4 , samples vary greatly in terms of their targeted population and size (the number of participants). The extent to which the results of a study can be generalized to the target population depends on the extent to which the sample is representative of the population.

Determining How Variables Will Be Measured: Levels of Measurement

The research hypotheses described in Table 1.1 involve variables such as type of appeal, intentions to use cannabis, and gender identity. To conduct a research study, the researcher must decide how to measure the variables stated in the research hypothesis. **Measurement** is the assignment of numbers or categories to objects or events according to rules.

For example, to measure the variable “height,” a researcher might use “number of inches from the ground in bare feet” as a form of measurement. To measure “success in college,” a student’s grade point average (GPA) may be obtained from school transcripts. “Self-esteem” might be measured by having people complete a questionnaire and be categorized as having either “low” or “high” self-esteem. As these examples demonstrate, different types of variables have different forms of measurement. Consequently, researchers have identified four distinct levels for measuring variables: nominal, ordinal, interval, and ratio.

The values of variables measured at the **nominal level of measurement** differ in category or type. The word *nominal* implies having to do with “names,” such that we use first and last names to distinguish between people. Gender identity is an example of a nominal variable, in that it consists of categories or types (such as cis male and cis female) rather than numeric values. In the first research hypothesis in Table 1.1 , the independent variable of “type of appeal” is measured at the nominal level with two categories: FOMO and non-FOMO.

Variables measured at the **ordinal level of measurement** have values that can be placed in an order relative to the other values. Rankings (such as finishing first, second, or third in a race), size (small, medium, large, or extra-large), and ratings (below average, average, or above average) are examples of ordinal variables. The values of ordinal variables indicate that one value represents relatively more or less of the variable than other values, but they don’t specify the precise difference between values. For example, although you can say that a “medium” drink is smaller than a “large” drink which is smaller than an “extra-large” drink, you can’t specify the exact difference in size between the three drinks.

Table 1.4 Research Hypotheses and Study Samples

Research Hypothesis	Study Sample
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.	“We recruited 150 participants from Amazon MTurk” (Friederich et at., 2024, p. 105).
Exposure to pro-cannabis messages on social media is	“Teens (N = 350) . . . needed to be between 13 and 17 years old and from Washington state” (Willoughby et

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Study Sample
related to increased intentions to use cannabis.	al., 2024, p. 185).
Mothers have more accurate knowledge of their children's online activities than do fathers.	"triadic data including reports from children aged 13 to 18, mothers and fathers . . . recruited in Flanders, the northern, Dutch speaking region of Belgium . . . a total of 357 valid triads was achieved" (Symons et al., 2017, p. 405).
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	"undergraduate college students aged 18–30 who were enrolled at 4-year institutions. This yielded a sample of 46,510 students" (Marx et al., 2024, p. 334).

The values of variables measured at the **interval level of measurement** are numbers equally spaced along a continuum. One example of an interval variable is the Fahrenheit scale of temperature. Here, a difference of 5 degrees has the same meaning anywhere along the scale; for example, the difference between 45°F and 50°F is the same as the difference between 65°F and 70°F. Many variables studied in the behavioral sciences (e.g., personality characteristics or attitudes) are considered to be measured at the interval level of measurement.

Variables at the **ratio level of measurement** are identical to interval variables, with one exception: Ratio scales possess what is known as a true zero point, for which the value of zero (0) represents the complete absence of the variable. Variables that describe a physical dimension (such as height, weight, distance, and time duration) typically have a true zero point. In the last research hypothesis in Table 1.1 , the variable "number of times victimized in last 12 months" is a ratio variable because it has a true zero point, where zero would indicate the complete absence of victimizations.

One advantage of ratio measurement versus interval is that ratio variables allow for a greater number of comparisons among values. Consider, for example, the value 6 for the ratio variable "inches." Not only is the difference between 4 and 6 inches the same as the difference between 6 and 8 inches (thereby involving addition and subtraction, which also can be used with interval variables), 6 inches is twice as long as 3 inches and half as long as 12 inches (involving multiplication and division). Multiplication and division comparisons cannot be made with interval variables. You cannot, for example, say that a temperature of 90°F is three times as much temperature as 30°F.

The boldfaced and italicized text in Table 1.5 illustrates how the four studies described in the earlier tables measured their independent and dependent variables. As you can see, the variables in these studies have been measured in a variety of ways at different levels of measurement.

Table 1.5 Research Hypotheses and Measurement of Variables

Research Hypothesis	Independent Variable	Dependent Variable

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Research Hypothesis	Independent Variable	Dependent Variable
Consumers exposed to a FOMO appeal (vs. non-FOMO appeal) will have a higher intention to invest in cryptos.	Type of appeal **FOMO, non-FOMO (nominal)**	Intention to invest in cryptos **Likelihood of investing** **1-7 scale (1 = Highly unlikely to 7 = Highly likely) (ordinal)**
Exposure to pro-cannabis messages on social media is related to increased intentions to use cannabis.	Exposure to pro-cannabis messages on social media **How often saw pro-cannabis messages (ratio)**	Intentions to use cannabis **Likelihood of using cannabis in the future (ratio)**
Mothers have more accurate knowledge of their children's online activities than do fathers.	Type of parent **Mother, Father (nominal)**	Accuracy of online activities **Accurate, somewhat accurate, inaccurate (nominal)**
Nonbinary college students experience higher levels of victimization than do cis men or cis women.	Gender identity **Nonbinary, cis man, cis woman** **(nominal)**	Number of times victimized in last 12 months **(ratio)**

What difference does the level of measurement make?

A variable's level of measurement has important implications for researchers in that it influences how research hypotheses are stated and how data are analyzed. For example, imagine you study the variable "success in college." You could measure this variable at the *nominal* level of measurement by classifying students into one of two groups: successful or unsuccessful. Researchers typically use nominal variables when they are interested in studying differences between groups, such as, "Do successful and unsuccessful students differ in their study habits?"

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

You could instead measure success in college as an *ordinal* variable by ranking students from top to bottom, which enables a researcher to study relative differences with less concern for the precise size of these differences. For example, using ranked data, you could ask, “How similar are younger and older faculty members’ rankings of their students?”

To study success in college at the *interval* level, you could rate each student on a scale from 1 (low) to 100 (high). You could then use these ratings to address a research question such as, “Is there a relationship between faculty ratings of their students and students’ ratings of themselves?”

Finally, to measure success in college using the *ratio* level of measurement, you could use students’ salary after graduation (in dollars) as an indicator of success. Salary is a ratio variable because it contains a true zero point. In practice, ratio variables can be used to ask many of the same questions as those involving interval variables. For example, you could ask, “Is there a relationship between faculty members’ ratings of students and students’ salaries after graduation?”

As you see, how researchers measure their variables influences how they state their research questions and hypotheses. The level of measurement has another important implication: It helps determine the statistics researchers use to analyze their data. Certain statistics can be applied to variables measured at the interval or ratio levels of measurement, whereas others are appropriate for nominal or ordinal variables. As we move from chapter to chapter of this book, the level of measurement of the variables being analyzed will be noted.

Selecting a Method to Collect the Data: Experimental and Nonexperimental Research Methods

In addition to drawing a sample from a population and determining how the variables will be measured, the third step in collecting data is to determine the type of research method to use to collect the data on these variables. Research methods may be classified into two main types:

experimental research methods, and

nonexperimental research methods.

Experimental research methods.

Experimental research methods are methods designed to test causal relationships between variables—more specifically, whether changes in independent variables produce or cause changes in dependent variables. To make inferences about cause-effect relationships, researchers conducting an experiment must first eliminate other possible causes or explanations for changes in the dependent variable besides the independent variable. If it can be claimed that a variable other than the independent variable created the changes in the dependent variable, the research results are said to be confounded. A **confounding variable** is a variable related to an independent variable that provides an alternative explanation for the relationship between the independent and dependent variables.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

To illustrate the concept of confounding variables, consider the following question: Is there a relationship between a mother's ethnicity and the birth weight of her baby? Although research has shown differences in the birth weights of babies of different ethnicities, one study noted that some of this research did not eliminate the possibility that these differences may be partly due to differences between ethnic groups on factors associated with low birth weights, such as the mother's average age at the time of pregnancy and behaviors such as smoking and drinking (Ma, 2008). These other factors are considered confounding variables in that they provide alternative explanations for the relationship between ethnicity and babies' birth weights. Confounding variables make it difficult for researchers to draw firm conclusions regarding the relationship between the independent and dependent variable.

Learning Check 2

Reviewing What You've Learned So Far



1. Review questions
    a. What is the difference between a population and a sample?
    b. Within the research process, what is the relationship between populations and samples?
    c. What are the four levels of measurement? How do they differ?
2. For each of the following variables, name the scale of measurement (nominal, ordinal, interval, or ratio).
    a. Type of school attended (public, private)
    b. Probability of graduating college in 4 years (0% to 100%)
    c. Rating of television (good, better, best)
    d. Number of computers in home

Researchers minimize the influence of confounding variables in two main ways. First, researchers exercise experimental control by making the research setting (i.e., characteristics of the research participants, the instruments or measures administered, the methods used to collect data) the same for all participants. In the birth weight study, for example, the research design might control for the effect of parental smoking on birth weight by only including nonsmokers in the study sample, thereby excluding those who smoke.

Another way researchers exert control over the research setting is by including a condition known as a **control group** , a group of participants in an experiment not exposed to the independent variable. For example, if you conducted an experiment examining the effects of caffeine on one's health, one group (the experimental group) might be instructed to drink coffee while a second group (the control group) drinks decaffeinated coffee. By comparing the results of the two groups, researchers can then assess the impact of caffeinated coffee on the health of those who drink it.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Because researchers cannot possibly identify and control for all potential confounding variables, a second strategy used to minimize the influence of confounding variables is **random assignment** , which is assigning participants to each category of an independent variable in such a way that each participant has an equal chance of being assigned to each category. For example, a researcher could randomly assign a participant to the experimental or control condition by flipping a coin: heads for experimental, tails for control. The purpose of random assignment is to equalize or neutralize the effects of confounding variables by distributing them equally over the levels of the independent variable.

Nonexperimental (correlational) research methods.

Experimental research designs are one of the best tools researchers have for making causal inferences. However, the ability to make causal inferences requires a great deal of control that may not always be possible or desirable. For this reason, researchers often employ **nonexperimental research methods** , which are research methods designed to measure relationships between variables without having the ability to infer cause-effect relationships. These research methods are sometimes referred to as *correlational* research methods. It is important to understand that correlational research methods are different from correlational statistics we'll introduce in Chapter 13. Correlational research methods pertain to how data are *collected* ; correlational statistics pertain to how data are *analyzed* . Some of the most common types of nonexperimental research designs include quasi-experiments, survey research, observational research, and archival research.

Quasi-experimental research compares naturally formed or preexisting groups rather than randomly assigning participants to conditions. One example of this type of research was a study examining the effects of the COVID-19 pandemic on three groups of students: those living on campus, those living off campus but within commuting distance to school, and those living far away in another city, state, or country (Bisconer & McGill, 2023). This study is a quasi-experiment because the students were not randomly assigned to one of the three living conditions before collecting data.

Survey research methods obtain information directly from a group of people regarding their opinions, beliefs, or behavior. The goal of survey research, which can include using questionnaires and interviews, is to obtain information from a sample to represent or estimate the views of a larger population. One example of survey research examined college students' mental health by having them fill out a survey in which they reported their spirituality (the desire for truth, meaning, and significance of one's life) and their self-acceptance (knowledge and acceptance of one's self and limitations) (Nonis et al., 2024). Survey research is nonexperimental in that the variables being measured are not manipulated by the researcher.

Observational research is the systematic and objective observation of naturally occurring behavior or events. Observational research is often used to study phenomena that the researcher cannot or should not deliberately manipulate. For example, researchers studying interpersonal conflict among children would never force children to fight each other. One study resolved this issue by videotaping groups of adolescents discussing topics such as competition and peer pressure and then counting the number of times the adolescents teased or mocked each other (Connolly et al., 2015).

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Rather than observing or measuring behavior directly, **archival research** is the use of archives (records or documents of the activities of individuals, groups, or organizations) to examine research questions or hypotheses. One example of an archival research study studied the question of whether men are perceived to be more creative than women (Proudfoot et al., 2015). In this study, the 100 most viewed talks on TED.com were identified, and the researchers compared viewers' posted ratings of the talks given by men versus women.

Labeling variables in nonexperimental research.

One issue that often confuses students when learning about nonexperimental research is the practice of referring to the variables as the "independent variable" and the "dependent variable." For example, in the study described previously that involved biological sex and TED.com ratings, biological sex may be called the "independent variable." However, because biological sex isn't manipulated by the researcher, it would be inappropriate to say that biological sex "causes" differences in TED.com ratings. To avoid confusion, researchers sometimes refer to the independent variable and the dependent variable in a nonexperimental study as the **predictor variable** and **criterion variable**, respectively.

Analyzing the Data

Once data collection has been completed, the next step in the research process involves data analysis. This part of the research process addresses the primary topic of this book: statistics. Because this is the first chapter of this book, we won't describe analyzing data in detail. Instead, we introduce the notion that there are two main purposes of analyzing data: (1) to organize, summarize, and describe data that has been collected and (2) to test and draw conclusions about ideas and hypotheses. These two purposes are met by calculating two main types of statistics: descriptive statistics and inferential statistics.

Calculating Descriptive Statistics

Descriptive statistics are statistics used to summarize and describe a set of data for a variable. For example, you've probably heard of statistics used to describe or summarize aspects of our society, such as crime statistics and unemployment statistics. Using a research-related example, Caitlin Abar, a researcher at Brown University, conducted a study examining increases in students' alcohol-related behaviors after entering college (Abar, 2012). To measure students' level of alcohol use, she created a variable called "typical weekend drinking," which "was measured as the sum of drinks consumed on a typical Friday and Saturday within the past 30 days" (p. 22). One way to summarize students' responses to the typical weekend drinking variable (0, 1, 2, etc.) would be to calculate the mean, which is the mathematical average of a set of scores. The mean, described in Chapter 3, is an example of a descriptive statistic. The first part of this book describes a variety of descriptive statistics used by researchers to summarize data they've collected.

Calculating Inferential Statistics

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Besides summarizing and describing data, a second purpose of statistics is to test hypotheses. **Inferential statistics** are statistical procedures used to test hypotheses and draw conclusions from data collected during research studies. Using inferential statistics, researchers make inferences about what's believed to exist in populations based on data gathered from samples of the population.

As an example of an inferential statistic, let's return to the example of students' alcohol-related behaviors (Abar, 2012). This study was interested in seeing whether these behaviors were related to aspects of these students' relationships with their parents. One aspect was "alcohol communications," defined as "how often they discussed alcohol related topics with their parents" (p. 22). To test a hypothesis regarding the relationship between students' alcohol use and their alcohol communications, we could use an inferential statistic known as the Pearson correlation coefficient, which will be discussed in Chapter 13. The chapters in the last half of this book discuss a broad variety of inferential statistics, along with detailed instructions for analyzing and drawing conclusions from statistical data and presenting research findings.

Drawing a Conclusion Regarding the Research Hypothesis

Once statistical analyses have been completed, the next step is to interpret the results of the analyses as they relate to the study's research hypothesis. More specifically, do the findings of the analyses support or not support the research hypothesis? The word *support* in the previous sentence is very important. Students and researchers are sometimes tempted to say their findings "prove" their hypotheses are either "true" or "false." However, as we'll discuss in Chapter 5, because researchers typically don't collect data from the entire population, they can't know with 100% certainty whether their hypothesis has been proven to be true or false. A *huge* part of this book is learning how to draw appropriate conclusions from statistical analyses.

Communicating the Findings of the Study

Conducting research requires a variety of different skills: conceptual skills to develop research hypotheses, methodological skills to collect data, and mathematical skills to analyze these data. Another integral part of the research process is the communication skills needed to inform others about a study. Researchers must be not only skilled scientists but also effective writers.

Learning Check 3

Reviewing What You've Learned So Far



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1. Review questions
    a. What are the main differences between experimental and nonexperimental research methods?
    b. What are confounding variables? What do researchers do to minimize the effects of confounding variables?
    c. What are the main types of nonexperimental research methods?
    d. What are the relative strengths and weaknesses of experimental and nonexperimental research?
    e. What are the main purposes of descriptive and inferential statistics?

For many years, the American Psychological Association (APA), the professional association for psychologists, has provided researchers guidance on how to communicate the results of their research. In 1929, the APA published a seven-page article entitled, “Instructions in Regard to Preparation of Manuscript.” In contrast, the seventh edition of the *Publication Manual of the American Psychological Association* , published in 2020, consists of 428 pages! This increase in length highlights the challenge of communicating research effectively and appropriately.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

1.4 Plan of the Book

The primary purpose of this chapter was to introduce statistics and place it within the larger research process. The remainder of this book discusses different statistics used by researchers to examine various questions of interest. Although it's critical for you to learn to correctly calculate statistics, it is equally important to understand the role of statistics within the research process and appreciate issues related to the use (and sometimes misuse) of statistics.

The first half of the book (Chapters 2 through 6) introduces conceptual and mathematical issues that are the foundation of statistical analyses. Chapters 2, 3, and 4 discuss how data may be examined, described, summarized, and presented in numeric, visual, and graphic form. Descriptive statistics are introduced and discussed in these chapters. Chapters 5 and 6 introduce characteristics of the inferential statistics used to test research hypotheses. The chapters in the second half of the book cover a number of different inferential statistics used with different types of variables.

In most cases, the chapters in this book share a uniform format and structure and are centered on the research process, using examples of actual published research studies that address questions with which you may be familiar. For each example, we'll clearly state the research hypothesis, briefly describe how the data in the study were collected, illustrate the calculation and interpretation of descriptive and inferential statistics, discuss the extent to which the statistical analyses support the research hypothesis, and show how to communicate one's research to a broader audience.

1.5 Looking Ahead

We began this chapter by defining statistics and providing reasons why learning about them may be of value to you. Next, we placed statistics within the stages of the larger research process, a process that will be the foundation of this textbook. As we've explained, the research process is centered on a research hypothesis, which is a predicted relationship between variables. Once a hypothesis has been stated, the next step is to collect data to be analyzed. However, before conducting statistical analyses on a set of data, there is preliminary work that must be completed. The next chapter will focus on methods used by researchers to examine data.

1.6 Summary

Statistics may be defined as a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data. As such, statistics not only is concerned about how data are analyzed but also recognizes the importance of understanding how data are collected and how the results of analyses are interpreted and communicated. There are a variety of reasons why students should learn statistics: Statistics are encountered in a wide variety of daily activities, students may be asked or required to read and interpret the results of statistical analyses in their courses, students may use statistics in conducting their own research, and statistics may help one's career.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Researchers conduct research using the **scientific method** of inquiry, a method of investigation that uses the objective and systematic collection and analysis of empirical data to test theories and hypotheses. The research process used within the scientific method of inquiry consists of five main steps: developing a research hypothesis to be tested, collecting data, analyzing the data, drawing a conclusion regarding the research hypothesis, and communicating the findings to the study.

A **research hypothesis** is a statement regarding an expected or predicted relationship between variables. Developing a research hypothesis involves identifying a question or issue to be examined, reviewing and evaluating relevant theories and research, and stating the research hypothesis to be tested in the study. A **theory** is a set of propositions used to describe or explain a phenomenon. A research hypothesis contains **variables** , which are properties or characteristics of some object, event, or person that can take on different values. More specifically, a research hypothesis states the nature and direction of a proposed relationship between an **independent variable** (a variable manipulated or controlled by the researcher) and a **dependent variable** (a variable measured by the researcher that is expected to change or vary as a function of the independent variable).

Collecting data involves drawing a sample from a population, determining how the variables will be measured, and selecting a method to collect the data. A **population** is the total number of possible units or elements that could potentially be included in a study; a **sample** is a subset or portion of a population. Variables can be measured at one of four levels of measurement: **nominal** (values differing in category or type), **ordinal** (values placed in an order relative to the other values), **interval** (values equally spaced along a numeric continuum), or **ratio** (values equally spaced along a numeric continuum with a true zero point).

There are two main types of methods used to collect data: **experimental research methods** , designed to test whether changes in independent variables produce or cause changes in dependent variables, and **nonexperimental (correlational) research methods** , designed to examine the relationship between variables without the ability to infer cause-effect relationships.

In experimental research, researchers are concerned about possible **confounding variables** , which are variables related to independent variables that provide an alternative explanation for the relationship between independent and dependent variables. To minimize the impact of confounding variables, a research study may exercise experimental control over various aspects of the situation, including the use of a **control group** , which is a group of participants not exposed to the independent variable the research is designed to test, or **random assignment** , which involves assigning participants to each category of an independent variable in such a way that each participant has an equal chance of being assigned to each category.

Examples of nonexperimental research methods are **survey research** , in which information is directly obtained from a group of people regarding their opinions, beliefs, or behavior; **observational research** , which involves the systematic and objective observation of naturally occurring events; and **archival research** , which uses archives (records or documents of the activities of individuals, groups, or organizations) to examine research questions or hypotheses. In nonexperimental research, the independent variable and

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

dependent variable may be referred to as the **predictor variable** and **criterion variable**, respectively.

Once data have been collected, the next step is to analyze them using two main types of statistics: **descriptive statistics**, which summarize and describe a set of data, and **inferential statistics**, which are statistical techniques used to test hypotheses and draw conclusions from data collected during research studies.

Once the statistical analyses have been completed, the results of the analyses are interpreted regarding whether they support or not support the study's research hypothesis. The final step in the research process is to communicate the study to a broader audience.

1.7 Important Terms

archival research

confounding variable

control group

criterion variable

dependent variable

descriptive statistics

experimental research methods

independent variable

inferential statistics

interval level of measurement

level of measurement (nominal, ordinal, interval, ratio)

measurement

nonexperimental research methods

observational research

population

predictor variable

quasi-experimental research

random assignment

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

research hypothesis

sample

scientific method

statistics

survey research

theory

variable

1.8 Exercises

1. Identify the independent and dependent variable for each of the following research questions and hypotheses from published articles.
    a. “We hypothesize that finance majors will embody narcissistic traits more than other business majors” (Wood et al., 2021, p. 924).
    b. “Men are expected to employ physical characteristics (particularly those that are directly related to sex) more often than women in selecting dating candidates” (Hetsroni, 2000, p. 91).
    c. “It is hypothesized that perceptions of emotional pain and support needs will be higher for the community of the Black victim than those of the officers’” (Kahn et al., 2024).
    d. “Gen Z students are more likely to believe that ‘Group projects bring out the worst in people’ than Millennials” (Schlee et al., 2020, p. 145).
    e. “We hypothesized to find . . . higher stress and anxiety post COVID in the Portuguese group in contrast with the British ” (Penalba-Sanchez et al., 2022, p. 322).
    f. “We hypothesized that consumers would think that a ‘sale price’ presentation would generate a greater monetary savings than an ‘everyday low price’ presentation” (Tom & Ruiz, 1997, p. 403).
2. The following are a number of research questions and hypotheses from actual published articles. For each hypothesis, identify the independent and dependent variable.
    a. “The use of color in a Yellow Pages advertisement will increase the perception of quality of the products for a particular business when compared with noncolor advertisements” (Lohse & Rosen, 2001, p. 75).
    b. “We hypothesized that parents who use more frequent corporal and verbal punishmen. . . . will report more problem behaviors in their children” (Brenner & Fox, 1998, p. 252).
    c. “It was hypothesized that adolescents with anorexia nervosa woul. . . . be more respectful when compared with peers with bulimia nervosa” (Pryor & Wiederman, 1998, p. 292).
    d. “The purpose of the present research was to assess brand name recognition as a function of humor in advertisements. . . . It was predicted that participants would recognize product brand names that had been presented with humorous

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

advertisements more often than brand names presented with nonhumorous advertisements" (Berg & Lippman, 2001, p. 197).

e. "The purpose of this study was to explore the relation between time spent in daycare and the quality of exploratory behaviors in 9-month-old infant. . . . it was hypothesized tha. . . . infants who spent greater amounts of time in center-based care would demonstrate more advanced exploratory behaviors than infants who did not spend as much time in center-based care" (Schuetze et al., 1999, p. 269).

f. "It was hypothesized that students would score higher on test items in which the narrative contains topics and elements that resonate with their daily experiences versus test items comprised of material that is unfamiliar" (Erdodi, 2012, p. 172).

3. Identify the independent and dependent variable for each of the following research questions and hypotheses from published articles.
    a. "It was hypothesized that the number of selfies posted . . . would be positively correlated with dimensions of narcissism" (Barry et al., 2017, p. 51).
    b. "We hypothesized that preschoolers' self-reported preferences for aggressive media . . . would be positively related to parent reports of children's aggressive behaviors" (Jamnik & DiLalla, 2018, p. 159).
    c. "We hypothesized that helicopter parenting would lead to higher levels of depression" (Darlow et al., 2017, p. 3).
    d. "We predicted that weight-related shame would be associated with greater binge eating" (Craven & Fekete, 2019, p. 45).
    e. "We hypothesize that men with strong gender identities will be more sensitive to manipulated threats to their masculinity compared with men with gender identities that are less central to their identity" (Harrison & Michelson, 2019, p. 65).
    f. "After exposure to fake news story with user comments identifying the news as fake, people are less likely to share the fake news than after exposure to the fake news story with user comments supportive of the content" (Colliander, 2019, p. 204).
4. Name the scale of measurement (nominal, ordinal, interval, ratio) for each of the following variables:
    a. The amount of time needed to react to a sound
    b. Gender
    c. Score on the Scholastic Aptitude Test (SAT)
    d. Political orientation (not at all conservative, conservative, very conservative)
    e. Political affiliation (Democrat, Republican, Independent)
5. Name the scale of measurement (nominal, ordinal, interval, ratio) for each of the following variables:
    a. One's age (in years)
    b. Size of soft drink (small, medium, large, extra large)
    c. Voting behavior (in favor vs. against)
    d. IQ score
    e. Parent (mother vs. father)
6. A faculty member wishes to assess the relationship between students' scores on the SAT and their performance in college.
    a. What is a possible research hypothesis in this situation?
    b. What are the independent and dependent variables?
    c. How could you measure the variable "performance in college" at each of the four levels of measurement?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

7. A researcher hypothesizes that drivers who use cellular phones will get into a greater number of traffic accidents than drivers who do not use these phones. For this example,
    a. What are the independent and dependent variables?
    b. How could you measure the dependent variable?
    c. How could you conduct this study using an experimental research method? How could you conduct this study using a nonexperimental method?

Answers to Learning Checks

Learning Check 1

1.

    a. IV: Instructor's dress; DV: Expertise

    b. IV: Amount of television viewing; DV: Posttraumatic stress symptoms

    c. IV: Student group; DV: Financial distress

    d. IV: Activity while driving; DV: Driving performance

Learning Check 2

2.

    a. Nominal

    b. Ratio

    c. Ordinal

    d. Ratio

Answers to Odd-Numbered Exercises

1.

    a. IV: Type of business major; DV: Narcissistic traits

    b. IV: Biological sex; DV: Employment of physical characteristics

    c. IV: Community; DV: Perceptions of emotional pain and support needs

    d. IV: Generation; DV: Beliefs about group projects

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

e. IV: Country; DV: Stress and anxiety

f. IV: Price presentation; DV: Perception of savings

3.

a. IV: Number of selfies posted; DV: Narcissism

b. IV: Preference for aggressive media; DV: Parent reports of children's aggressive behaviors

c. IV: Helicopter parenting; DV: Depression

d. IV: Weight-related shame; DV: Binge eating

e. IV: Strength of gender identity; DV: Sensitivity to threat to masculinity

f. IV: Type of user comments to news story; DV: Sharing of news story

5.

a. Ratio

b. Ordinal

c. Nominal

d. Interval

e. Nominal

7.

a. IV: Cellular phone use; DV: Number of traffic accidents

b. Example: Determine the number of traffic accidents each person has experienced in the past year.

c. Example of experimental: Use a driving simulation with the "driver" talking on a phone and measure the number of potential driving mistakes or accidents.

d. Example of nonexperimental: Take a survey of the number of accidents people have been in and whether or not they were using a phone during the accident.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CHAPTER 2 EXAMINING DATA: TABLES AND FIGURES

Learning Objectives

2.1 Describe the main purpose and the research hypothesis of the “lottery” study.

2.2 Describe different reasons for examining data.

2.3 Describe the steps used to create a frequency distribution table.

2.4 Describe the steps used to create a grouped frequency distribution table.

2.5 Explain the differences between bar charts, pie charts, histograms, and frequency polygons.

2.6 Describe the three aspects and types of distributions of data.

Chapter 1 described a process by which research may be conducted, which begins by stating a research hypothesis. The next step involves collecting data that will ultimately be used in statistical analyses designed to test the research hypothesis. However, researchers don’t conduct analyses until they’ve first examined the data to make sure appropriate conclusions can be drawn from the analyses. This chapter describes *why* researchers examine data and *how* data may be examined using tables, charts, and graphs. As with many of the chapters in this book, this discussion will be guided by the findings from a published research study.

2.1 An Example From the Research: Winning the Lottery

A friend of yours flips a coin and asks you to guess whether it’s “heads” or “tails.” You guess “tails,” but the answer is “heads.” Your friend flips the coin a second time, and again you predict “tails.” But again the answer is “heads.” On the third coin flip, you guess “tails” once more—but, once again, you’re wrong. As your friend prepares to flip the coin a fourth time, you tell yourself there’s little chance the coin could be “heads” again. But the reality is that the fourth coin flip is equally likely to be “heads” or “tails” because the outcome of the first three coin flips has absolutely no impact on the fourth flip. In other words, the outcome of a coin flip is a random event that can’t be controlled or predetermined. But do you really believe this?

The impact of random events can take on greater significance than the outcome of a coin flip. For example, government-run lotteries provide the opportunity to win huge amounts of money through the random selection of numbers. In 2001, Dr. Karen Hardoon, a researcher at McGill University in Canada, studied the thought processes people use when purchasing lottery tickets (Hardoon et al., 2001). As previous research had found gamblers believe they can control random events such as the rolling of dice or the spinning of a roulette wheel, Dr. Hardoon and her colleagues were interested in

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

determining “whether individuals with gambling problems perceive the purchase of lottery tickets in a similar manner as non-problem gamblers” (p. 752). On the basis of earlier findings, Dr. Hardoon hypothesized that problem gamblers are less likely to believe lottery outcomes are random than are nonproblem gamblers.

To test their research hypothesis, Dr. Hardoon and her associates located a group of gamblers and asked whether they had ever experienced problems as a result of their gambling. Each participant was classified into one of two groups: problem gamblers and nonproblem gamblers.

The participants in the study, which we’ll refer to as the lottery study, were shown four lottery tickets and asked, “If you were to buy a ticket to play in the lottery, which one would you select?” Each ticket contained six numbers selected from among the numbers 1 through 49; the four tickets were given a particular label. The *sequence* ticket had numbers in consecutive order (30-31-32-33-34-35); numbers on the *pattern* ticket increased by 5s (5-10-15-20-25-30); the *nonequilibrated* numbers (35-37-40-43-44-49) were clustered at the upper end of the 49 numbers, and the *random* ticket (7-8-23-34-36-42) shared none of the characteristics of the other three tickets. We’ll call the variable in this study “lottery ticket,” a variable with four categories: sequence, pattern, nonequilibrated, and random.

The researchers recorded which of the four tickets each participant chose. Table 2.1 lists the data for the lottery ticket variable for the nonproblem gamblers (the data for the problem gamblers are provided in the exercises at the end of this chapter). Looking at this table, we see that the nonproblem gamblers differed in their lottery ticket choices; however, it’s difficult to describe these differences given the unorganized nature of the collected data. Consequently, before conducting statistical analyses on their data, researchers typically organize and examine them. In the following sections, we will discuss why and how data are organized and examined.

2.2 Why Examine Data?

There are a variety of reasons why researchers examine data they’ve collected before analyzing the data. These reasons include the following:

To gain an initial sense of the data

To identify outliers

To evaluate research methodology

To determine whether data meet statistical criteria and assumptions

Each of these reasons is discussed in the following sections.

Table 2.1 The Lottery Ticket Choices of 22 Nonproblem Gamblers

Participant	Lottery Ticket
1	**Random**
2	**Random**
3	**Nonequilibrated**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Participant	Lottery Ticket
4	**Pattern**
5	**Random**
6	**Sequence**
7	**Random**
8	**Random**
9	**Pattern**
10	**Pattern**
11	**Random**
12	**Pattern**
13	**Random**
14	**Nonequilibrated**
15	**Random**
16	**Pattern**
17	**Nonequilibrated**
18	**Random**
19	**Random**
20	**Pattern**
21	**Random**
22	**Random**

Gaining an Initial Sense of the Data

Because researchers spend a great amount of time and energy designing their research studies, they're eager to analyze the data they've collected. Students taking courses in statistics, on the other hand, face a different challenge: They're given sets of data with which they have no prior experience. Although both researchers and students may be motivated to analyze their data as quickly as possible, it's important to first examine the data in order to gain an initial sense of them.

As an example, imagine you conduct a study regarding the fuel efficiency of automobiles and collect data on the miles per gallon (MPG) of different types of cars. You might expect the scores in your data to range from 10 to 50 MPG, with the majority between 25 and 35 MPG. Examining data helps researchers gain an initial sense of whether the data are in line with their expectations. Examining data also provides a way of confirming the results of later statistical analyses. We've found that students' errors in their calculations could have been avoided had they first looked at the set of data. For example, examining data will help you avoid making statements such as, "The average MPG in my sample was 127 MPG."

Identifying Outliers

Researchers also examine their data to identify **outliers** , defined as rare, extreme scores that lie outside of the range of the majority of scores in a set of data. Imagine, for example, that a group of 15 students is asked to take a 20-item test of mathematical calculations; each student's score is the number of test items answered correctly. In scoring the tests, a researcher finds that 14 students

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

correctly answered between 8 and 17 questions. However, one student provided the wrong answer to every question for a score of 0. In this case, the score of 0 would be considered an outlier.

If not identified, outliers may distort conclusions researchers draw about their sample and their data, particularly when the sample is relatively small. In the previous example, the single score of 0 might lead an observer to conclude the class as a whole performed more poorly on the test than they actually did.

Evaluating Research Methodology

Examining data is also useful in assessing the effectiveness of the methods used to collect the data. For the 20-item math test, for example, it might be reasonable to expect the majority of scores to fall in the middle of the 0 to 20 range. But what if every student answered at least 17 of the 20 questions correctly? You might conclude the test was too easy and that more difficult questions are needed to accurately assess students' math skills. By comparing the collected data for a variable with the range of possible values, researchers are able to evaluate and modify their measurement tools.

Determining Whether Data Meet Statistical Criteria and Assumptions

Researchers also examine their data to see whether they meet certain statistical criteria and assumptions. Many of the statistics in this book are based on assumptions regarding the shape and nature of a set of data that has been collected. For example, these procedures assume that in most data sets, the majority of scores will fall in the middle of the range of possible values, with a small amount of scores at the highest and lowest possible values (you may have heard this referred to as "the bell curve"). In the example of the 20-item math test, it might be assumed that most scores will be between 7 and 13, with smaller number of scores less than 7 or greater than 13. When assumptions such as this are not met, it's possible to draw inappropriate conclusions from statistical analyses that are conducted.

2.3 Examining Data Using Tables

One part of examining data that have been collected for a variable involves answering a series of questions:

> What are the values of the variable?
>
> How many participants in the sample have each value of the variable?
>
> What percentage of the sample has each value of the variable?

As it may be difficult to answer questions such as these just by looking at the data in Table 2.1 , researchers typically organize their data by creating a table known as a **frequency distribution table** , a table that summarizes the number and percentage of participants for the different values of a variable. This section discusses the steps involved in creating frequency distribution tables using these questions.

What Are the Values of the Variable?

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The first step in organizing the data for a variable is to identify all of the possible values for the variable. Table 2.2 begins the construction of a frequency distribution table by identifying the four values for the lottery ticket variable: sequence, pattern, nonequilibrated, and random.

How Many Participants in the Sample Have Each Value of the Variable?

Once the values of the variable have been identified, the next step in organizing a set of data is to determine the **frequency** of each value, which is the number of participants in a sample corresponding to a value of a variable. These frequencies help indicate the nature and shape of the distribution of scores for a variable in a sample.

Table 2.2 Identifying the Values of the Lottery Ticket Variable

Number	Lottery Ticket
1	Sequence
2	Pattern
3	Nonequilibrated
4	Random

Table 2.3 Using the Tick Mark Method to Determine the Frequencies of the Lottery Ticket Variable

Lottery Ticket	Tick Marks	*f*
Sequence	|	1
Pattern	||||— |	6
Nonequilibrated	|||	3
Random	||||— ||||— ||	12
Total	||||— ||||— ||||—||||—||	22

One way to determine the frequencies for a variable is to sort the data into the different values and then count the number of participants having each value using a “tick mark method,” where a tick mark is placed next to a value of the variable each time that value appears in the sample. As shown in Table 2.3 , the tick marks are then counted to determine the frequency for each value. For example, the number of tick marks and therefore the frequency (f) for the sequence ticket is 1 because it was picked by one participant (Participant 6 in Table 2.1). Similarly, the tick marks for each of the other three types of tickets can be recorded and counted until all of the data in the sample are accounted for.

What Percentage of the Sample Has Each Value of the Variable?

Although it’s important to determine the frequency for each value of a variable, these frequencies by themselves are not always particularly informative. For example, how you interpret “There are 35 Democrats in my sample” is different when the sample consists of 40 voters rather than 400 voters.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Because the interpretation of frequencies depends on the size of the sample, it's useful to calculate the *percentage* of the sample having each value of a variable. These percentages can be calculated using the following formula:

$$\% = \frac{f}{total\,number\,of\,scores} * 100$$

where *f* is the frequency for a value of a variable.

Looking at Table 2.3 , 1 of the 22 nonproblem gamblers chose the sequence ticket. Therefore, the percentage of the sample choosing this ticket is

$$\begin{aligned} \% &= \frac{f}{total\,number\,of\,scores} * 100 \\ &= \frac{1}{22} * 100 = .05 * 100 \\ &= 5\% \end{aligned}$$

After calculating the percentage for each of the four tickets, Table 2.4 provides the frequency distribution table for the lottery ticket variable.

Once a frequency distribution table has been constructed for a variable, researchers can begin to draw conclusions about their sample. For the lottery ticket variable, the random ticket had the highest frequency and percentage (*f* = 12, 54%) compared to the pattern (*f* = 6, 27%), nonequilibrated (*f* = 3, 14%), and sequence (*f* = 1, 5%) tickets. As a result, the researchers in the lottery study made the following observation:

"The results of the present study indicated that, for the entire sample . . . the greatest percentage of tickets selected by the entire sampl. . . . were random tickets." (Hardoon et al., 2001, p. 760)

Table 2.4 Frequency Distribution Table for the Lottery Ticket Variable

Lottery Ticket	***f***	**%**
Sequence	1	**5%**
Pattern	6	**27%**
Nonequilibrated	3	**14%**
Random	12	**54%**
Total	22	**100%**

Learning Check 1

Reviewing What You've Learned So Far

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



1. Review questions
    a. What are some reasons for examining data before conducting statistical analyses?
    b. Why are outliers of concern to researchers?
    c. What questions do you address about a set of data for a variable in constructing frequency distribution tables?
    d. Why is it useful to calculate percentages in addition to frequencies for each value of a variable?
2. For each of the following situations, create a frequency distribution table.
    a. A college counselor asks a group of 116 seniors what their plans are after graduating from college; 71 say they are going to “work,” 34 are planning on going to “graduate school,” and 11 say they are “not sure” what their plans are.
    b. A high school instructor teaching a course finds that of the students she teaches, 69 are Freshmen, 18 are Sophomores, 12 are Juniors, and 9 are Seniors.
    c. A pollster stops 12 people and asks if they are either “in favor” or “against” a local proposition in Table 2.5 :

Table 2.5 Example of People’s Position

Person	Position
1	**In favor**
2	**In favor**
3	**Against**
4	**In favor**
5	**Against**
6	**Against**
7	**In favor**
8	**Against**
9	**In favor**
10	**Against**
11	**In favor**
12	**In favor**

    d. The company that makes M&Ms candy (www.mms.com) conducted a survey asking people which new color they would like to have added: purple, aqua, or pink. Table 2.6 are the votes of a hypothetical sample:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Table 2.6
Hypothetical Sample of New M&M Color

Person	Color
1	**Pink**
2	**Purple**
3	**Pink**
4	**Aqua**
5	**Purple**
6	**Purple**
7	**Purple**
8	**Aqua**
9	**Purple**
10	**Pink**
11	**Aqua**
12	**Pink**
13	**Pink**
14	**Purple**
15	**Aqua**
16	**Pink**
17	**Purple**
18	**Purple**
19	**Aqua**
20	**Purple**
21	**Purple**
22	**Pink**
23	**Pink**
24	**Purple**

2.4 Grouped Frequency Distribution Tables

The frequency distribution table in Table 2.4 was sufficient for presenting the frequencies for the lottery ticket variable. However, other situations may involve variables with a large number of possible values. Consider, for example, the variable grade point average (GPA). The values for GPA typically range from 0.00 to 4.00, with hundreds of possible values in between. Creating a frequency distribution table for GPA could result in an extremely large table, with many values having frequencies of zero.

To examine a variable with a large number of values, it is useful to construct a **grouped frequency distribution table** , a table that groups the values of a variable measured at the interval or ratio level

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

of measurement into a small number of intervals and then provides the frequency and percentage for each interval. To illustrate how grouped frequency distribution tables are constructed, Table 2.7 (a) provides a sample of 42 hypothetical GPAs. In Table 2.7 (b), each of these 42 GPAs has been placed into one of eight intervals that comprise a grouped frequency distribution table.

To illustrate how to read and interpret a grouped frequency distribution table, the lowest interval in Table 2.7 (b) consists of students with GPAs less than (<) 2.00. The frequency in this interval (*f* = 1) consists of one student: Student 29 (GPA = 1.96). The next interval consists of students with GPAs ranging from 2.00 to 2.29 and contains three students (4, 22, and 38). This grouping continues until all of the GPAs in the sample have been accounted for.

To create a grouped frequency distribution table, the real limits of each interval should be identified. The **real limits** of an interval are the values of the variable that are halfway between the top of one interval and the bottom of the next interval. For example, because of how values of GPA are rounded, the lowest value of GPA that would lead someone to be placed in the interval 2.30–2.59 is 2.295, which is halfway between the top of the 2.00–2.29 interval and the bottom of the 2.30–2.59 interval. The **real lower limit** is the smallest value of a variable that would be grouped into a particular interval. In this case, 2.295 represents the real lower limit for the 2.30–2.59 interval. On the other hand, the **real upper limit** is the largest value of a variable that would be grouped into a particular interval. For the 2.30–2.59 interval, the real upper limit is 2.594.

Once the grouped frequency table for the GPA example has been created (Table 2.7(b)), what preliminary conclusions might be drawn about these students? First, because the 2.90–3.19 interval has the greatest frequency (*f* = 10), you might conclude that the typical student in this sample has a GPA close to 3.00. Second, moving in both directions away from the 2.90–3.19 interval, the frequencies of the different intervals grow progressively smaller, with fewer and fewer students having either very high or very low GPAs. Table 2.8 provides a list of guidelines for creating grouped frequency distribution tables.

Cumulative Percentages

In addition to determining the percentage of the sample that has each individual value or interval for a variable, it is sometimes useful to combine these percentages. For the GPA example, you may be interested in knowing, “What percentage of the sample had GPAs less than 2.60?” To answer this question, the percentages in the “%” column of the frequency distribution table must be combined. This type of percentage is referred to as a **cumulative percentage** , defined as the percentage of a sample at or below a particular value of a variable.

Table 2.7a Grade Point Average (GPA), 42 Students: Individual GPAs

Student	GPA
1	**3.05**
2	**2.83**
3	**3.26**
4	**2.19**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Student	GPA
5	**3.52**
6	**3.34**
7	**3.02**
8	**2.97**
9	**3.71**
10	**2.50**
11	**2.77**
12	**3.10**
13	**3.06**
14	**3.83**
15	**2.39**
16	**3.16**
17	**3.92**
18	**2.37**
19	**3.65**
20	**3.70**
21	**3.00**
22	**2.26**
23	**2.65**
24	**3.40**
25	**2.87**
26	**2.92**
27	**2.89**
28	**3.41**
29	**1.96**
30	**3.77**
31	**3.20**
32	**2.86**
33	**2.94**
34	**2.71**
35	**3.43**
36	**3.95**
37	**3.32**
38	**2.28**
39	**3.08**
40	**3.25**
41	**2.40**
42	**3.22**

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Table 2.7b Grade Point Average (GPA), 42 Students: Grouped Frequency Distribution Table

GPA	*f*	%
3.80+	3	7%
3.50−3.79	5	12%
3.20−3.49	9	21%
2.90−3.19	10	24%
2.60−2.89	7	17%
2.30−2.59	4	10%
2.00−2.29	3	7%
<2.00	1	2%
Total	42	100%

Table 2.8 Guidelines for Creating Grouped Frequency Distribution Tables

Guideline	Guideline Description
1. *Variables are grouped into approximately 10 intervals.*	The exact number of intervals depends on the data in each sample. For example, data that fall within a small range of values may be accurately represented with a relatively small number of intervals, whereas data extending across a wide range may require a larger number of intervals.
2. *The number of intervals should accurately represent the data.*	The intervals should represent the nature of the data as accurately as possible. For example, there should not be so large a number of intervals that many intervals have frequencies of zero (e.g., 3.00−3.05) or so few intervals that a large majority of the sample falls into only one or two intervals (e.g., 3.00−3.99).
3. *Intervals should be of equal size.*	To make the intervals comparable, they should be of equal size or width. For example, you would not want one interval of 2.01−2.50 (a width of .50) and another of 2.51−2.70 (a width of .20). One exception to this guideline is the lowest or highest interval (e.g., less than 2.00 or greater than 3.70), which is typically wider than the other intervals because it contains a relatively small frequency.
4. *Intervals should not overlap.*	The intervals should be distinct from each other, such that each score may be placed in only one interval. For example, intervals of 2.30− *2.60* and *2.60*− 2.90 are a problem because the GPA 2.60 is in both intervals leading to confusion regarding within which interval to place someone with a GPA of 2.60.

Table 2.9 includes the cumulative percentages for the GPA example. As an example of a cumulative percentage, the cumulative percentage for the 2.30–2.59 interval (19%) has been calculated by combining the percentages for three intervals: < 2.00, 2.00–2.29, and 2.30–2.59: 2% + 7% + 10% = 19%. As a result, we could conclude that 19% of this sample had GPAs less than 2.60. The accumulating of percentages continues until all (100%) of the scores in the sample are combined. As

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

shown in Table 2.9 , cumulative percentages are typically located in the final column of a frequency distribution table and labeled “Cum %.”

Table 2.9 Frequency Distribution Table With Cumulative Percentages for the GPA Variable

GPA	*f*	%	Cum %
3.80+	3	7%	**100%**
3.50−3.79	5	12%	**93%**
3.20−3.49	9	21%	**81%**
2.90−3.19	10	24%	**60%**
2.60−2.89	7	17%	**36%**
2.30−2.59	4	10%	**19%**
2.00−2.29	3	7%	**9%**
< 2.00	1	2%	**2%**
Total	42	100%	

2.5 Examining Data Using Figures

In addition to organizing data into frequency distribution tables, researchers also examine data visually using figures such as charts or graphs. There are four types of figures often used to visually portray data for variables: bar charts, pie charts, histograms, and frequency polygons. Which type of figure a researcher may use depends on the variable’s level of measurement (nominal, ordinal, interval, ratio). Data for variables measured at the nominal or ordinal level of measurement are typically displayed using bar charts or pie charts. The values of interval and ratio variables are graphed using histograms and frequency polygons. Each of these types of visual illustration is discussed in the following sections.

Displaying Nominal and Ordinal Variables: Bar Charts and Pie Charts

The values of variables measured at the *nominal* level of measurement differ in category or type; one example of a nominal variable is the lottery ticket variable discussed earlier, which consisted of four categories: sequence, pattern, nonequilibrated, and random. The values of variables measured at the *ordinal* level of measurement can be placed in an order relative to the other values; age (young, middle age, elderly) is one example of an ordinal variable.

One way to display nominal and ordinal variables is to use a bar chart. A **bar chart** is a figure that uses bars to represent the frequency or percentage of a sample corresponding to each value of a variable, with the different bars not touching each other. Figure 2.1 shows a bar chart for the data from the lottery ticket variable. In this figure, each bar corresponds to one of the four types of lottery tickets; the height of each bar corresponds to the frequency (*f*) of nonproblem gamblers who selected that type of ticket.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

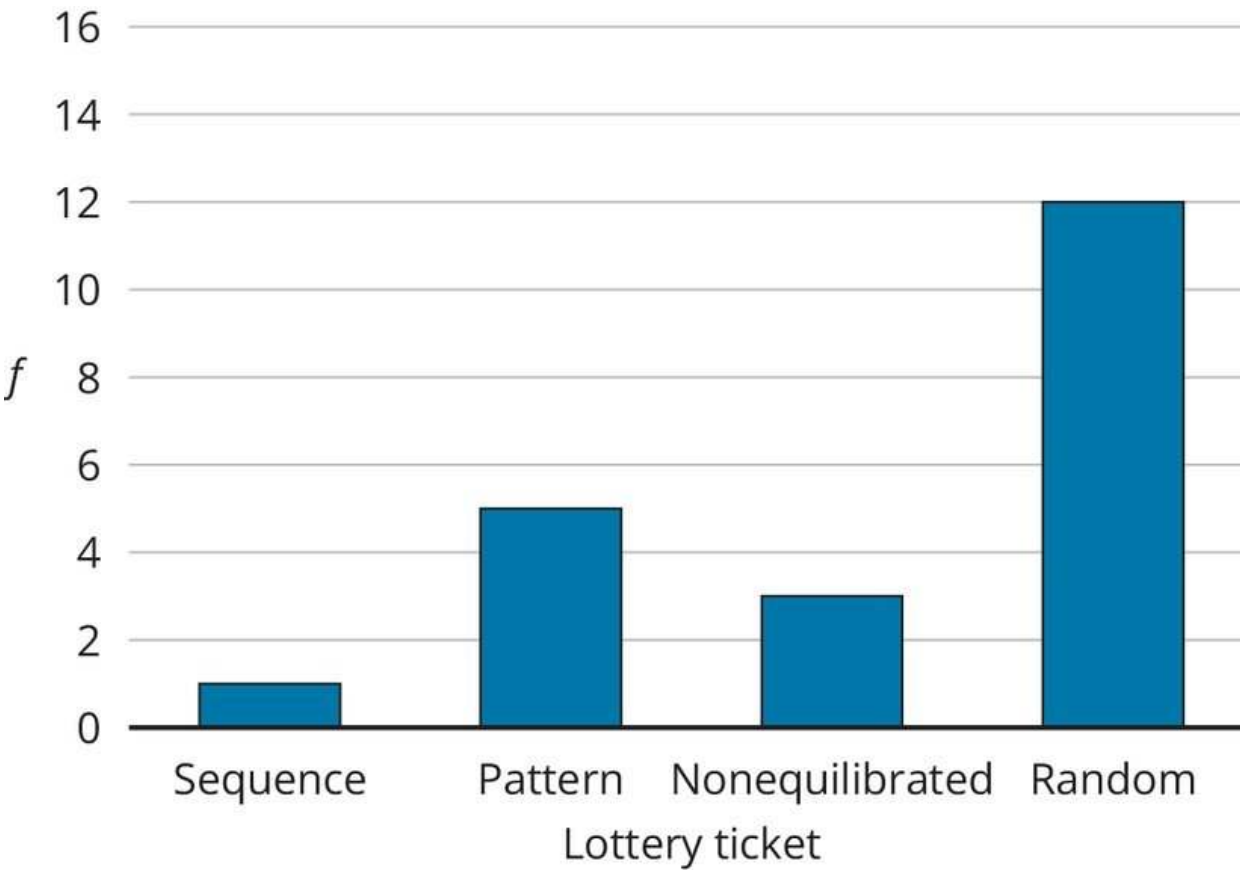


Figure 2.1 Bar Chart Showing the Frequencies of the Lottery Ticket Variable for the Nonproblem Gamblers

When creating a bar chart, the horizontal (X) axis contains the name of the variable and the values of the variable; the vertical (Y) axis contains the frequency (*f*) or percentage (%) for each value. Notice that the bars in the bar chart in Figure 2.1 do not touch each other; the gap between the bars indicates that the values of the variable represent distinct groups or categories that can't be placed along a numeric continuum.

A second type of figure used to display nominal or ordinal variables is a pie chart. A **pie chart** is a figure that uses a circle divided into proportions (slices) that represent the percentage of the sample for each value of a variable. Pie charts are created by dividing the 360 degrees of a circle into proportions corresponding to the percentage of each value.

Figure 2.2 shows a pie chart for the data for the lottery ticket variable. To illustrate how pie charts are constructed, the frequency distribution table in Table 2.4 shows that 5% of nonproblem gamblers chose the sequence ticket. Given that a pie chart consists of 360 degrees, 5% of the 360 degrees are needed to represent the sequence ticket. As 5% * 360 = 18, 18 degrees of the pie is assigned to the sequence ticket. Similarly, 97 degrees of the pie chart (27% of 360) are used to represent the pattern ticket. This process continues until all 360 degrees of the pie chart are accounted for.

The decision whether to use a bar chart or pie chart to illustrate a variable may depend on the number of values for the variable. For example, a pie chart for a variable that has 12 values may

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

result in many of the 12 “slices” being very thin and small, making them hard to see, meaning that a bar chart may be a better choice to illustrate this variable.

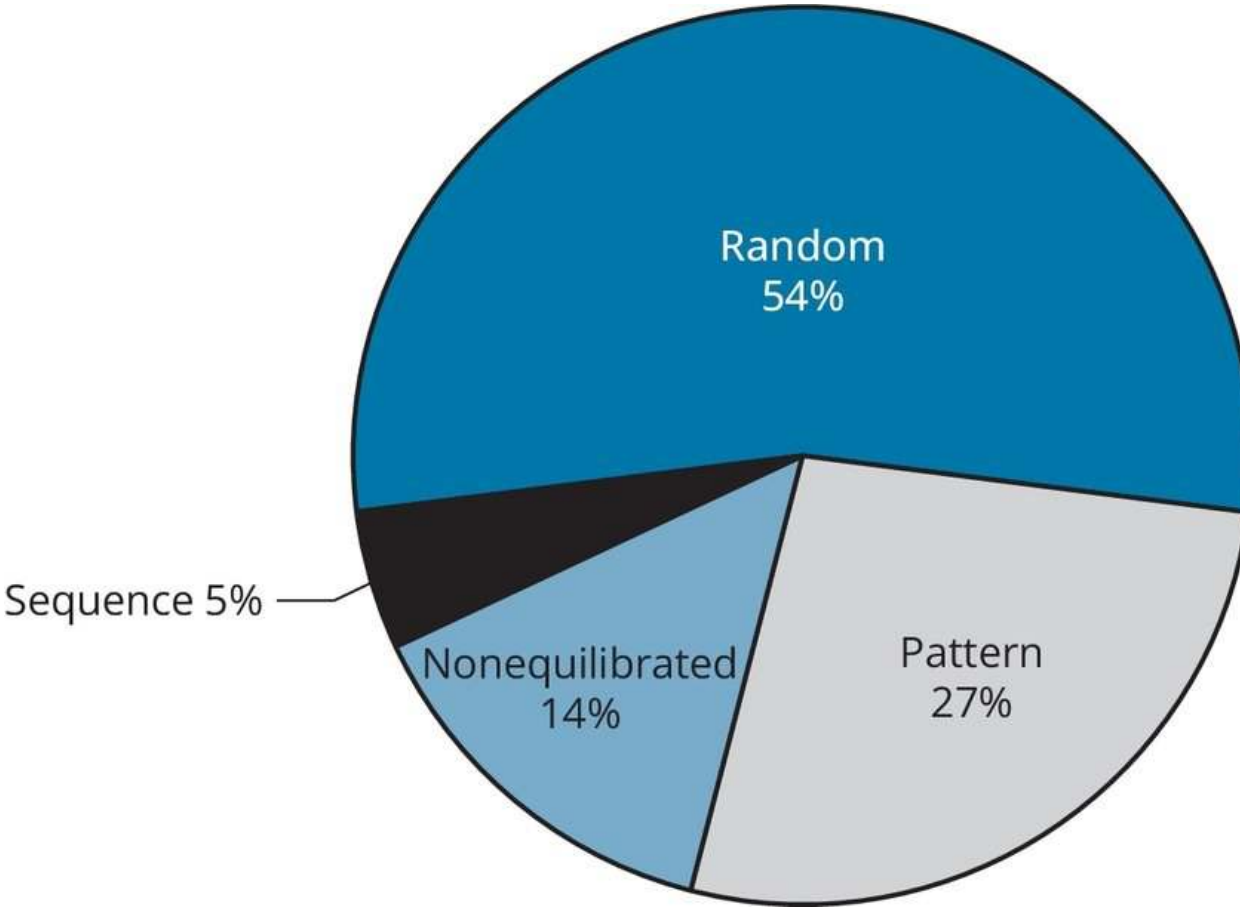


Figure 2.2 Pie Chart Showing the Percentages of the Lottery Ticket Variable for the Non−Problem Gamblers

Displaying Interval and Ratio Variables: Histograms and Frequency Polygons

Interval and ratio variables have values equally spaced along a numeric continuum and are identical to each other with one exception: Ratio scales possess a true zero point, for which the value of zero (0) represents the complete absence of the variable. The values of interval and ratio variables may be illustrated using histograms and frequency polygons, the choice of which depends on the number of values for the variable.

A **histogram** is a figure in which bars are used to represent the frequency of each value of a variable, with the different bars touching each other. The name and values of the variable are located on the X-axis, and the frequency or percentage for each value is on the Y-axis. Histograms are similar to bar charts, with one important difference: The bars in a histogram touch each other, indicating that the values of the variable are connected to each other along a numeric continuum.

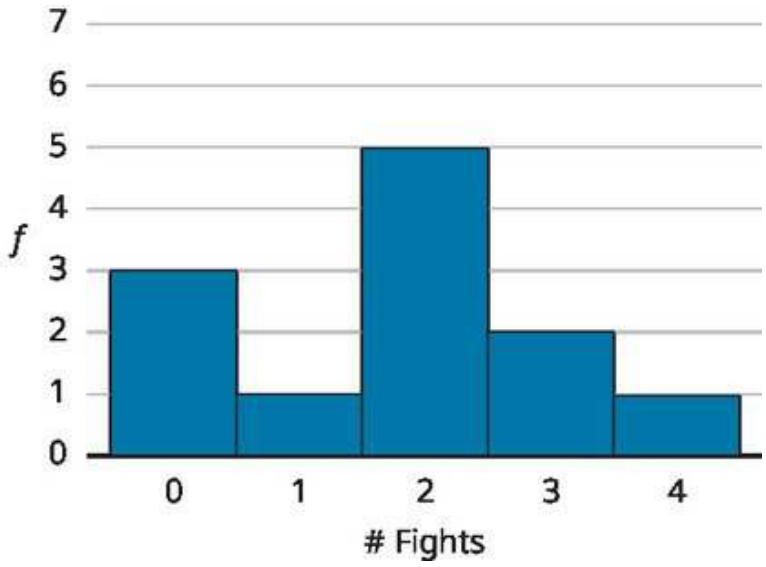
join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

As an example, imagine a teacher is interested in studying aggressive behavior in children and measures this variable by counting the number of fights each of her students is involved in over a 3-day period. “Number of fights” is a ratio variable rather than an interval variable because the value of 0 represents the complete absence of fights. Figure 2.3 provides a frequency distribution table and histogram for this variable.

A histogram is used when an interval or ratio variable consists of a relatively small number of values. But what if, for example, the number of fights could be anywhere from 0 to 20 rather than 0 to 5? In order to have the histogram fit on one page, the 21 bars would need to be very thin and crowded together, making the histogram confusing and difficult to read. In situations such as this, the values of the variable may be more clearly illustrated using a figure known as a **frequency polygon**, which is a line graph that uses data points to represent the frequency of each value of a variable, with straight lines connecting the data points. In a frequency polygon, the X-axis represents the name and values of the variable, while the Y-axis represents the frequency for each value.

# Fights	*f*
4	1
3	2
2	5
1	1
0	3
Total	12



▶ Long Description: Figure 2.3 Frequency Distribution Table and Histogram for the Number of Fights Variable

Figure 2.3 Frequency Distribution Table and Histogram for the Number of Fights Variable

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

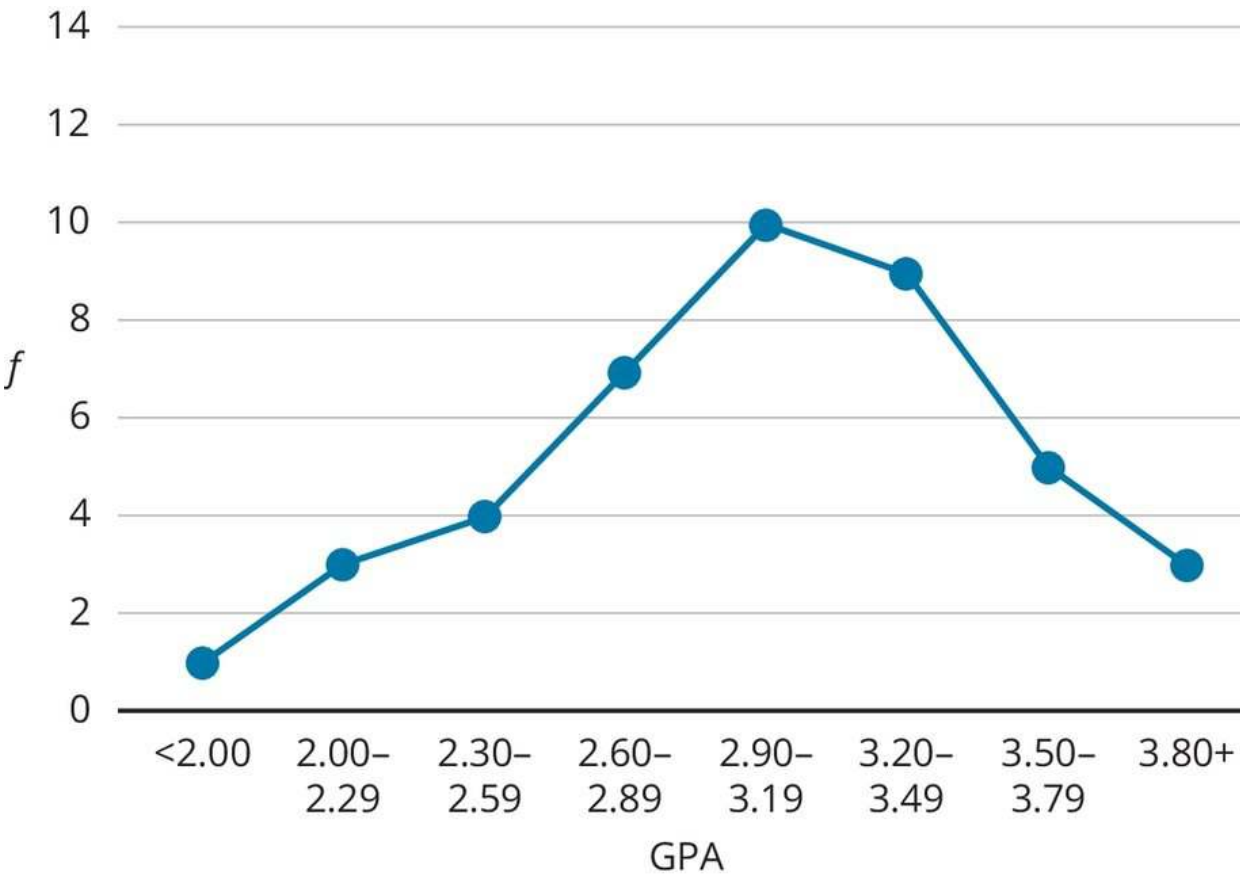


Figure 2.4 Frequency Polygon Showing the Frequency of Students in Each Interval of the Grade Point Average (GPA) Variable

Figure 2.4 shows a frequency polygon based on the grouped frequency distribution table of the earlier example of students' GPA (Table 2.7(b)). In this frequency polygon, the data points representing the frequencies of the eight GPA intervals are connected using straight, sharp lines. These lines indicate that the values of the variable can be connected together along a numeric continuum.

Although a frequency polygon like the one in Figure 2.4 may look jagged as it moves from one data point to the next, it provides a general sense of the shape of the distribution of scores for a variable. For example, looking at Figure 2.4 , you might conclude that the majority of GPAs in this sample are in the middle of the 1.00 to 4.00 range, with relatively low frequencies of students with either low or high GPAs.

Learning Check 2

Reviewing What You've Learned So Far

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH



1. Review questions
    a. What is the difference between a frequency distribution table and a grouped frequency distribution table?
    b. For what levels of measurement would you create a pie chart, bar chart, histogram, or a frequency polygon?
    c. What is the difference between a bar chart and a histogram?
    d. Why would it be inappropriate to create a histogram or frequency polygon for a nominal or ordinal variable?
    e. Under what circumstances might you use a frequency polygon rather than a histogram to graph an interval or ratio variable?
2. For each of the following variables, identify whether you could use a bar chart, pie chart, histogram, and/or a frequency polygon to visually display the data.
    a. Miles per gallon (MPG)
    b. Living situation (on-campus, off-campus)
    c. Number of brothers and sisters
    d. Marital status (single, married, divorced)

Drawing Inappropriate Conclusions From Figures

A well-constructed figure can greatly facilitate the understanding of complex information. However, a poorly designed figure may not only be confusing, it can lead to inappropriate or misleading conclusions. Box 2.1 provides guidelines for constructing figures based on recommendations from the American Psychological Association (APA).

A simple example may demonstrate how altering the features of a figure can lead to different interpretations of the same set of data. Imagine a pollster assesses how voters with different political affiliations plan to vote on a ballot measure. She asks a sample of Democrats, Republicans, and Independents the question, “Do you plan on voting in favor of Measure A?” In examining her data, the pollster finds 56% of Democrats, 47% of Republicans, and 52% of Independents intend to vote “yes” on the measure.

Figure 2.5 graphs these three percentages using two bar charts; bar charts are used to display the data because the variable of interest (political affiliation) is a nominal variable (Democrat, Republican, or Independent). The bars don’t touch each other because the three political affiliations are distinct from each other and can’t be placed along a numeric continuum.

Although the two bar charts are similar, there is one important difference in how they are constructed. Following the guidelines in Box 2.1, the bar chart in Figure 2.5 (a) starts the Y-axis with the value 0% and ends with 100%. However, the bar chart in Figure 2.5 (b) starts with the value 45% and ends with 65%. By limiting the range of values of the percentages, the bar chart in Figure 2.5 (b) exaggerates differences in the heights of the three bars. As a result, someone looking at Figure 2.5 (b) may interpret the differences between the three political affiliations as being larger than someone looking at Figure 2.5 (a).

The saying “a picture is worth a thousand words” is particularly relevant for the presentation of statistical data. Because figures can provide a vivid representation of data, it’s the responsibility of researchers to portray their data in an accurate manner and to avoid creating images that may mislead or confuse.

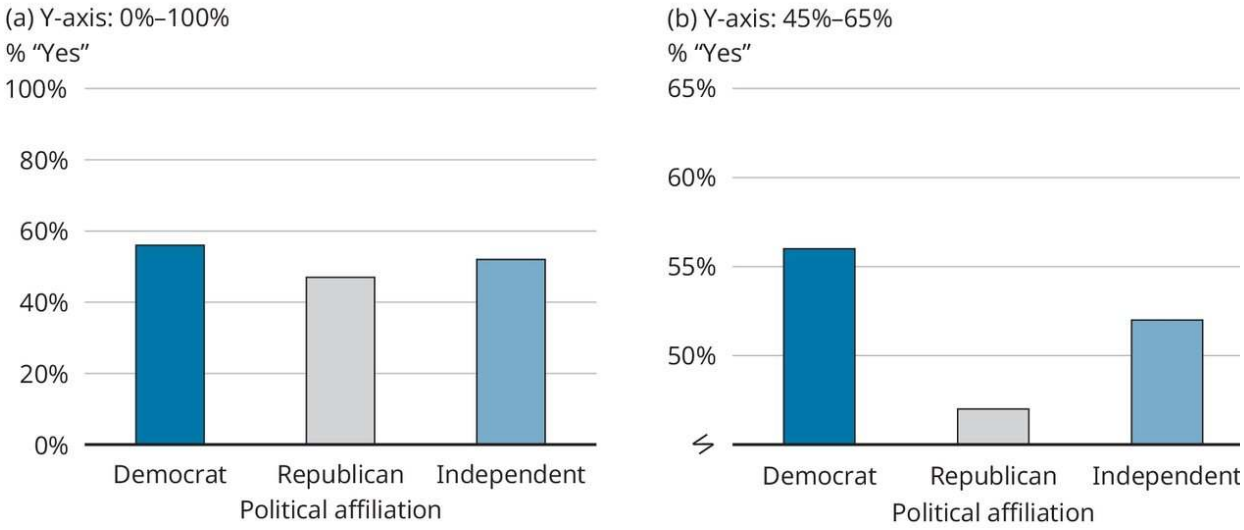
join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Box 2.1 Some APA Guidelines for Creating Figures

1. The figure should be formatted in a portrait rather than landscape page orientation (the reader should not turn the page sideways to read the table).
2. The figure is numbered using Arabic numbers (Figure 1, not Figure *I* or Figure *A*).
3. The figure includes a title that describes the contents of the figure. For example, it is not enough to name a figure simply " Figure 2.1 " or " Figure 2.1 . Bar chart."
4. Figures have a rectangular shape, with the height of the vertical (Y) axis two thirds to three fourths the length of the horizontal (X) axis.
5. The horizontal (X) axis contains the values of the variable; the vertical (Y) axis contains the frequency (*f*) or percentage (%) of the sample having each value of the variable.
6. The *X* - and *Y* -axes include descriptive labels written in both upper- and lowercase letters (as opposed to all capital letters).
7. The labels for the *X* - and *Y* -axes are short enough to fit on one line along the axes.
8. The label of the Y-axis is often placed parallel to the Y-axis (the letters face sideways rather than vertically) but can sometimes be seen on top of the y-axis for accessibility.
9. The Y-axis starts with the value zero (0) and is divided into equally spaced values or intervals.
10. The Y-axis is divided into 6 to 10 values (just enough values to accommodate bar length). The bars should neither be flattened along the bottom of the Y-axis nor bumped up against the top of the Y-axis.

Political affiliation	% "Yes"
Democrat	56%
Republican	47%
Independent	52%



▶ Long Description: Figure 2.5 Different Displays of the Same Data

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Figure 2.5 Different Displays of the Same Data

2.6 Examining Data: Describing Distributions

We've discussed how researchers construct tables and figures to organize their data. Looking at a table or figure enables researchers to describe the distribution of values or scores for a variable. Describing a distribution involves addressing questions such as, "Which values of the variable occur most often?," "How do the frequencies of the values of the variable change in relation to the most common values?," and "What is the amount of differences among the different values?"

Questions such as these pertain to three aspects of distributions: modality, symmetry, and variability. These three aspects of distributions are introduced in the following sections; they'll also be discussed in greater detail in the next two chapters of this book.

Modality

The first feature of a distribution of data for a variable is its modality. The **modality** of a distribution refers to the value or values of the variable that have the highest frequency or occur most often in a set of data. Using the lottery study as an example, identifying the modality of the lottery ticket variable involves answering the question, "Which of the four lottery tickets was chosen the most often?"

The modality of a distribution can be identified by looking at a frequency distribution table; the modality is the value(s) of the variable with the highest frequency (*f*) or percentage (%). The modality may also be determined by examining a figure and locating the tallest bar in a bar chart or histogram, the highest point in a frequency polygon, or the biggest "slice" in a pie chart. For the lottery ticket variable, the modality is the random ticket.

In identifying the modality of a distribution, it may be that one value or a limited number of values adjacent to one another clearly have the highest frequency. In such cases, the distribution is referred to as a unimodal distribution. A **unimodal distribution** is a distribution where one value of a variable occurs with the greatest frequency. Looking at the frequency distribution table (Table 2.4) or figure (Figure 2.1) for the lottery ticket variable, this distribution would be considered unimodal as one value, the random ticket, clearly occurred more often than the other three ticket choices.

In contrast to a unimodal distribution, it's possible that two distinct values may both have relatively high frequencies; if so, the distribution is referred to as a **bimodal distribution** , which is a distribution where two values clearly distinct from each other have the greatest frequency. A bimodal distribution sometimes resembles the two humps on a camel's back. As an example of a bimodal distribution, in April 2024 the Pew Research Center (www.pewresearch.org) asked a sample of 7,000 voters whether they felt Donald Trump had the mental fitness needed to be president, with five responses ranging from "extremely" to "not at all." The majority of voters chose one of the two extreme choices, either "extremely" (23%) or "not at all" (35%), with only 14% choosing the middle response of "somewhat."

And, in addition to unimodal and bimodal distributions, it's possible for a distribution to have more than two values with relatively equally high frequencies; this type of distribution is known as a **multimodal distribution** —a distribution where more than two values have the greatest frequency. In examining people's attitudes toward gun control, in May 2023, the news network CNN asked 1,227 people the question, "Do you think allowing gun owners to carry their guns in public places makes those places safer, less safe, or makes no difference in how safe a public place is?" They found people differed greatly in their responses, with 32% saying "makes public places safer," 36% saying

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

"makes public places less safe," and 32% saying "makes no difference," Figure 2.6 presents an example of a unimodal, bimodal, and multimodal distribution.

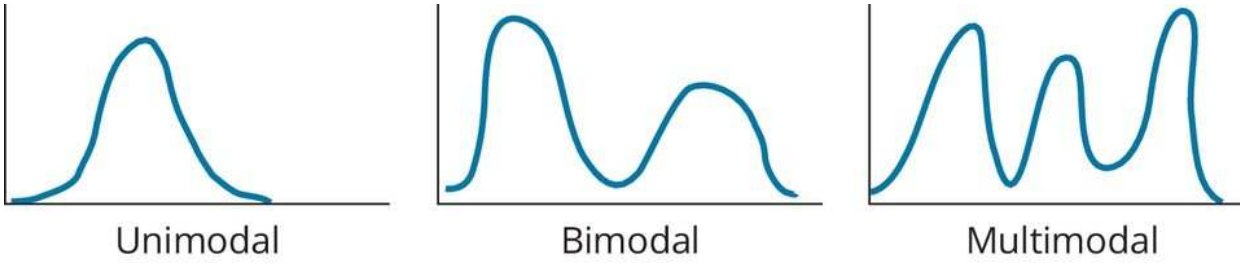


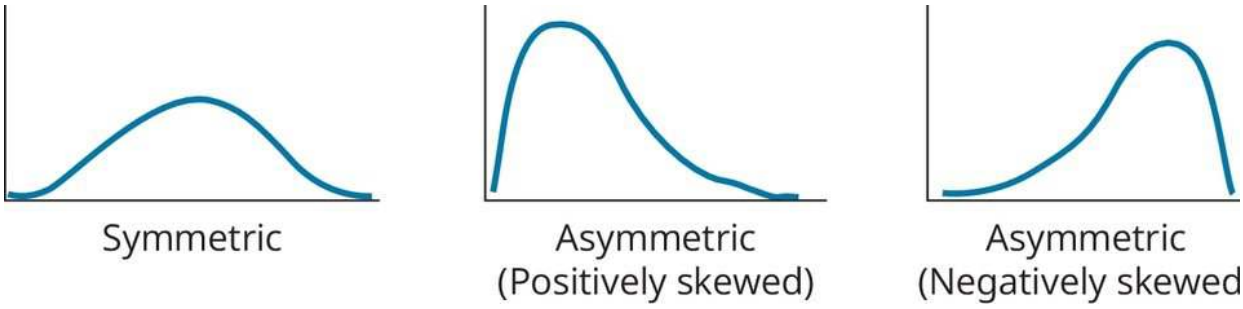
Figure 2.6 Distributions With Different Modalities (Unimodal, Bimodal, and Multimodal)

Symmetry

The second feature of a distribution of data is its symmetry. The **symmetry** of a distribution refers to how the frequencies of values of a variable change in relation to the most common or frequently occurring values. Describing the symmetry of a distribution means answering the question, "What is the shape of the distribution?"

The symmetry of a distribution can be determined by looking at the frequencies in a frequency distribution table or the bars or points in a figure. By doing so, a distribution may be described as being either symmetric or asymmetric. A **symmetric distribution** is one in which the frequencies change in a similar manner moving away in both directions from the most frequently occurring values. An example of a symmetric distribution is illustrated in Figure 2.7 . One way to determine whether a distribution is symmetric is to draw the distribution on a piece of paper and then fold the distribution vertically down the middle so that the two halves are on top of each other. A distribution is symmetric when the two halves are similar in shape such that they resemble mirror images of each other.

However, a distribution may also be asymmetric. An **asymmetric distribution** (also known as a *skewed* distribution) is one in which the frequencies change in a different manner moving away in both directions from the most frequently occurring values. In asymmetric distributions, the highest frequencies are located at one end of the distribution rather than in the middle, such that if you fold an asymmetric distribution down the middle, the shapes of the two halves do not look the same.



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Figure 2.7 Distributions With Different Symmetries (Symmetric, Positively Skewed, and Negatively Skewed)

In an asymmetric distribution, the frequencies take on the shape of a long tail as the values move toward the end of the distribution; scores in the tail of the distribution are often considered to be outliers. For example, in the 2022 Current Population Survey, the U.S. Census Bureau (www.census.gov) estimated that most Americans (63%) had annual incomes less than $50,000. However, a very small percentage (3%) were greater than $200,000; these higher incomes would be considered outliers.

Depending on the location of its tail, asymmetric distributions are referred to either as positively skewed or negatively skewed. A **positively skewed distribution** is one in which the higher frequencies are at the lower end of the distribution, with the tail on the upper (right) end of the distribution; a **negatively skewed distribution** is one in which the higher frequencies are at the upper end of the distribution, with the tail on the lower (left) end of the distribution. Figure 2.7 provides an illustration of a symmetric distribution and two asymmetric distributions (one positively skewed and one negatively skewed). Annual income, illustrated by the U.S. Census Bureau example mentioned earlier, is an example of a positively skewed distribution.

Variability

The third way in which distributions may be described is in terms of the amount or nature of differences among the different values for a variable. The **variability** of a distribution refers to the amount of differences in a distribution of data for a variable. The variability in a distribution is described by answering the question, “To what extent are the scores in the distribution different from each other?”

The amount of variability in a distribution can be estimated by looking at a frequency distribution table or figure. Imagine, for example, an instructor gives an 11-item quiz to her class and finds almost all of the students answer six, seven, or eight of the items correctly. In this situation, the distribution would resemble the shape of a tall mountain: thin and tall, with many of the scores residing in a few values of the variable. A distribution with a small amount of variability such as this one is referred to as a **peaked distribution** , defined as a distribution where much of the data are in a small number of values of a variable.

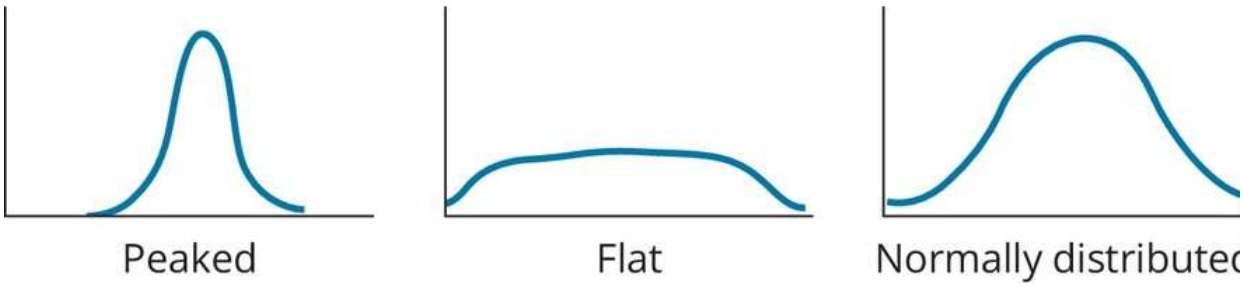


Figure 2.8 Distributions With Different Variability (Peaked, Flat, and Normally Distributed)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

On the other hand, what would the distribution look like if she had found her students' scores were distributed relatively evenly across the entire range of possible scores? In this situation, the distribution would not be tall and thin like a mountain but rather would be low and wide like a plateau. This type of distribution is referred to as a **flat distribution** , which is a distribution where the data are spread evenly across the values of a variable. The CNN gun control example described earlier is an example of a flat distribution.

In discussing modality, symmetry, and variability, distributions may have more than one mode, may be asymmetric (skewed), and may be peaked or flat. But what if a distribution has none of these qualities? A distribution that is unimodal, symmetric, and neither peaked nor flat may be described as "bell shaped." If you look at a bell, you'll notice that it is the highest in the middle, and as you move away from the top of a bell in both the left and right directions, the shape of the bell slopes downward in the same manner. A bell-shaped distribution for a variable is given a special name: a normally distributed variable.

A **normally distributed variable** is a distribution for a variable that is considered unimodal, symmetric, and neither peaked nor flat. Note that the word *normal* is a statistical concept and not a value judgment. In other words, asymmetric, peaked, or flat distributions are not considered "abnormal." The concept of normal distributions will be discussed in greater detail in Chapter 5. Figure 2.8 provides an illustration of three distributions with different amounts of variability: peaked, flat, and normally distributed.

Learning Check 3

Reviewing What You've Learned So Far



1. Review questions
    a. What are the three aspects used to describe distributions of variables?
    b. What is the difference between unimodal, bimodal, and multimodal distributions?
    c. What is the difference between symmetric and asymmetric (skewed) distributions?
    d. What is the difference between positively and negatively skewed distributions?
    e. What is the difference between peaked and flat distributions?
    f. What are the characteristics of a normally distributed variable?
2. For each of the following sets of data, determine whether the distribution of scores is unimodal, bimodal, or multimodal.
    a. 4, 2, 11, 6, 10, 2, 6, 13, 2, 10, 6, 10, 1
    b. 15, 17, 13, 18, 17, 14, 20, 18, 17, 14, 17
    c. 20, 10, 50, 10, 80, 30, 70, 10, 60, 10, 80, 0, 30, 80, 10, 100, 80
3. For each of the following sets of data, determine whether the distribution of scores is symmetric or asymmetric (skewed).
    a. 4, 5, 1, 4, 2, 4, 3, 5, 3, 4, 4, 3
    b. 220, 230, 250, 230, 200, 240, 210, 230, 270, 220, 240, 230, 240, 230
    c. 3, 1, 9, 2, 3, 6, 2, 7, 3, 5, 2, 6, 3, 2, 2
4. For each of the following sets of data, determine whether the distribution of scores is peaked or flat.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

a. 16, 10, 19, 12, 13, 15, 14, 17, 15, 20, 11, 13, 19, 16, 15
b. 12, 15, 12, 11, 13, 11, 12, 12, 14, 10, 12, 11, 16, 12, 13, 12, 11
c. 14, 12, 16, 14, 13, 15, 12, 14, 14, 15, 14, 14, 13, 14

2.7 Looking Ahead

This chapter has discussed why and how researchers examine data. Examining data by creating tables and figures requires a number of different skills, including organizational skills (to prepare a set of data for statistical analysis), communication skills (to present data in tables or figures), and conceptual skills (to describe a distribution). The next two chapters focus on mathematical skills related to the statistical analysis of data, describing how researchers describe distributions numerically rather than visually. Given that this is a statistics textbook, we will focus on calculating statistics to describe and analyze data. However, and we cannot say this strongly enough, it is *extremely* important to examine a set of data before conducting statistical analyses as this will prevent errors both in your calculations of statistics as well as your interpretation of the results of these calculations.

2.8 Summary

Before conducting statistical analyses, researchers typically examine their data to gain an initial sense of the data, detect data coding or data entry errors, identify ***outliers*** (rare, extreme scores that lie outside of the range of the majority of scores in a set of data), evaluate research methodology, and determine whether the data meet statistical criteria and assumptions.

One common way of examining variables involves creating tables known as ***frequency distribution tables*** (tables that summarize the number and percentage of participants that express the different values for a variable) and ***grouped frequency distribution tables*** (tables that group the values of a variable measured at the interval or ratio level of measurement into a small number of intervals and then provide the frequency and percentage within each interval). To create a grouped frequency table, the ***real limits*** of each interval (the values of the variable that fall halfway between the top of one interval and the bottom of the next interval) need to be identified.

Data for a variable may also be examined by creating figures such as bar charts, pie charts, histograms, and frequency polygons. ***Bar charts*** and ***pie charts*** are used for variables measured at the nominal or ordinal level of measurement; a bar chart uses bars to represent the frequency or percentage of a sample corresponding to each value of a variable, with the different bars not touching each other; a pie chart is a circle divided into proportions that represent the percentage of the sample corresponding to each value of a variable. ***Histograms*** and ***frequency polygons*** are used for variables measured at the interval or ratio level of measurement; a histogram is a figure in which bars are used to represent the frequency of each value of a variable, with the different bars touching each other; a frequency polygon is a line graph that uses data points to represent the frequency of each value of a variable, with lines connecting the data points.

From examining their data, researchers can describe the distribution of values or scores for a variable. There are three main aspects of distributions: modality, symmetry, and variability.

The ***modality*** of a distribution is the value(s) of the variable that have the highest frequencies or occur most often in a set of data. ***Unimodal distributions*** are distributions where one value occurs with the greatest frequency, ***bimodal distributions*** have two values with the greatest frequency, and ***multimodal distributions*** have more than two values with the greatest frequencies.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

The ***symmetry*** of a distribution refers to how the frequencies of values of a variable change in relation to the most common or frequently occurring values of the variable. In ***symmetric distributions*** , the frequencies change in a similar manner, moving away in both directions from the most common values of the variable. In ***asymmetric (skewed) distributions*** , the frequencies change in a different manner, moving away in both directions from the most common values of the variable.

The ***variability*** of a distribution refers to the amount of differences in a distribution of data. In ***peaked distributions*** , much of the data are in a few values of a variable; in ***flat distributions*** , the data are spread evenly across the values of a variable. A ***normally distributed variable*** has a symmetric distribution that is neither peaked nor flat.

2.9 Important Terms

asymmetric (skewed) distribution

bar chart

bimodal distribution

cumulative percentage

flat distribution

frequency (*f*)

frequency distribution table

frequency polygon

grouped frequency distribution table

histogram

modality

multimodal distribution

negatively skewed distribution

normally distributed variable

outlier

peaked distribution

pie chart

positively skewed distribution

real limits

real lower limit

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

real upper limit

symmetric distribution

symmetry

unimodal distribution

variability

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

2.10 Formula Introduced in This Chapter

Percentage (%)

$$\% = \frac{f}{total\ number\ of\ scores} * 100$$

2.11 Using IBM SPSS Software *

Creating Data Files: Defining Variables and Entering Data

* IBM SPSS software. SPSS is a registered trademark of International Business Machines Corporation.

1. Open a new data file (See Figure 2.9).

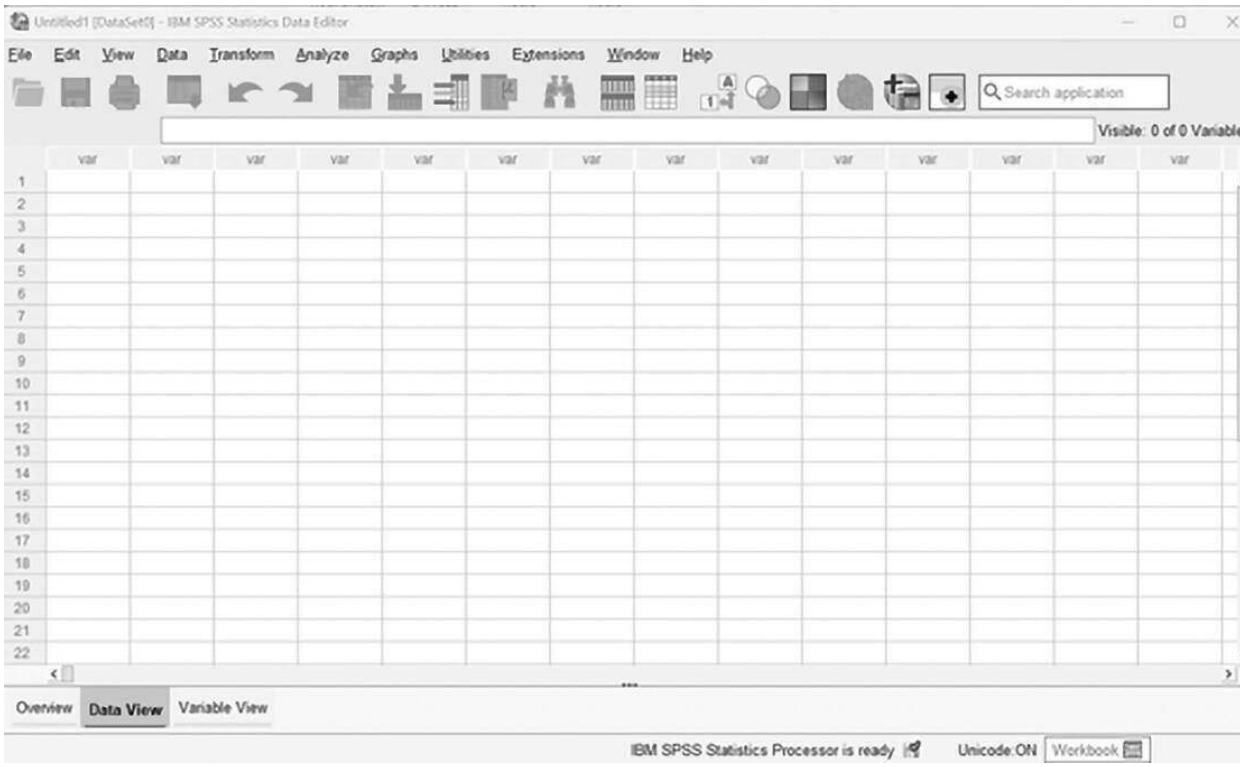


Figure 2.9 Data View

2. Define and name variable(s) (See Figure 2.10).

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

How? (1) Click **Variable View** tab, (2) click below **Name** , and (3) type name of variable in the cell.

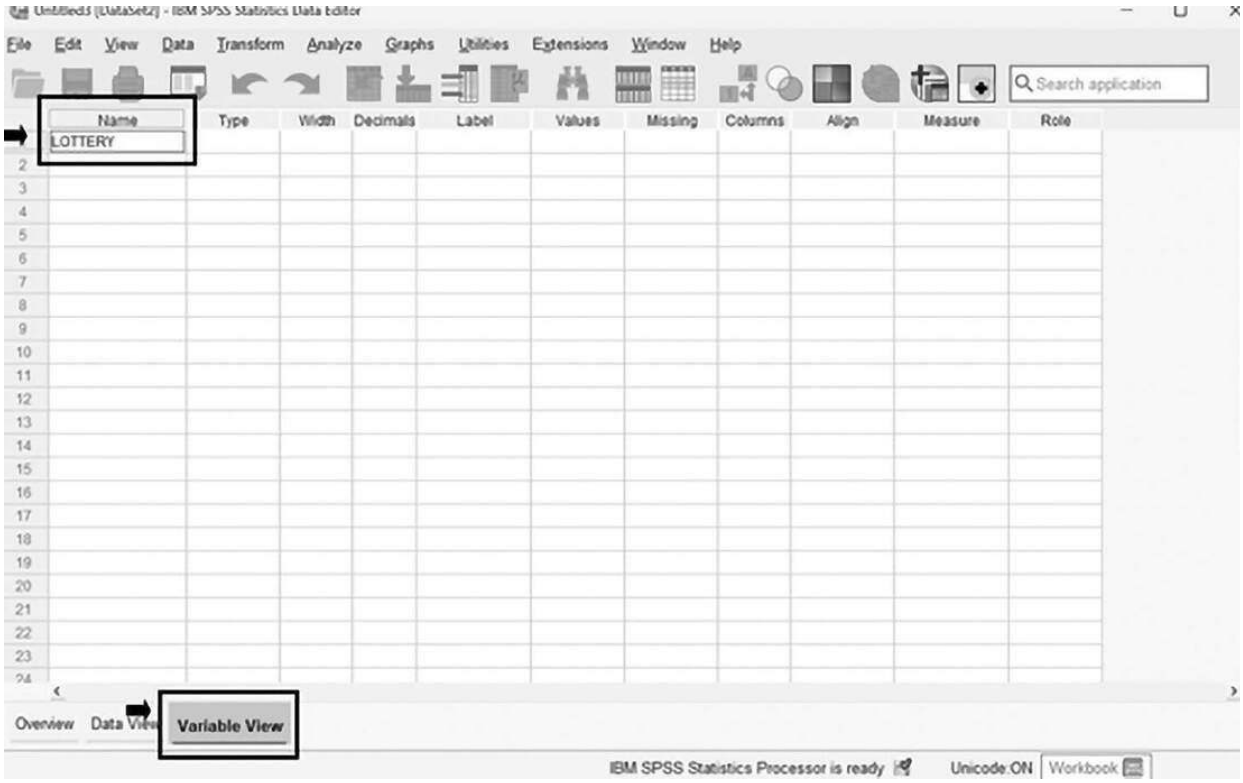


Figure 2.10 Variable View

3. Determine number of decimals for your variable(s) (See Figure 2.11).

How? Click desired number of decimal places in the **Decimals** box.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

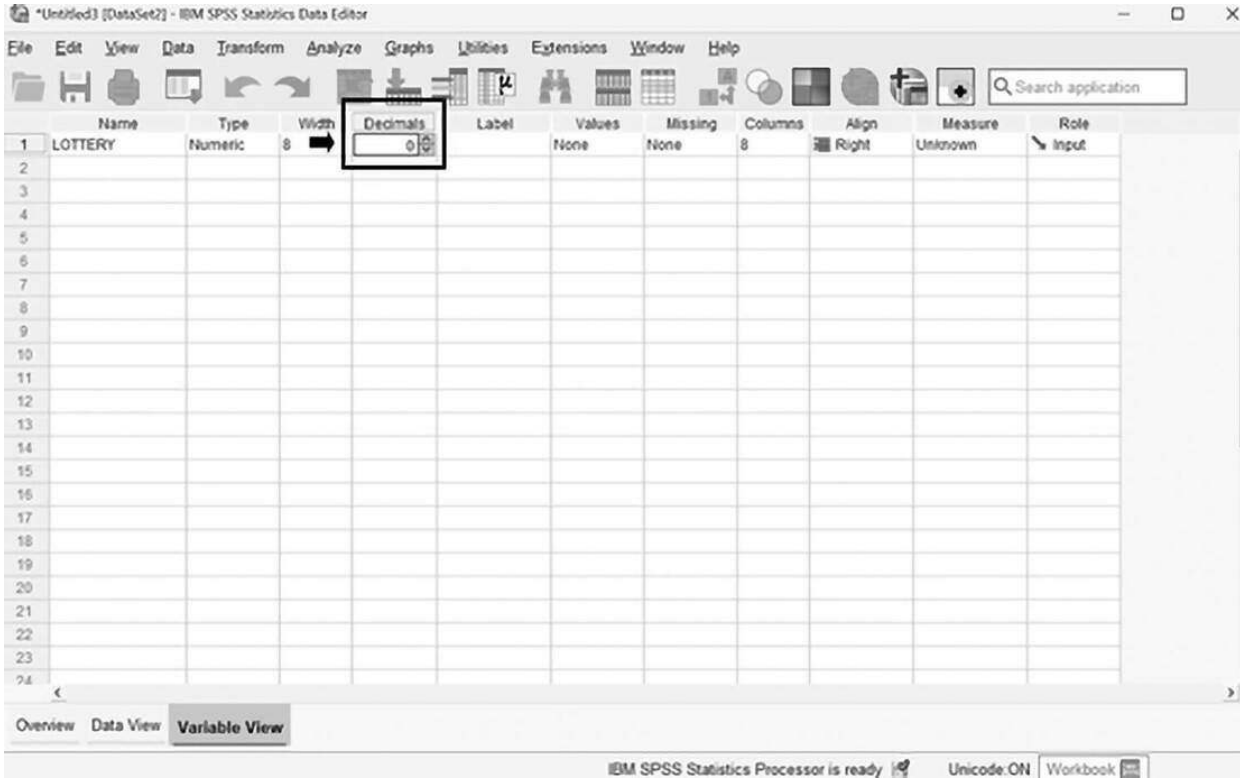


Figure 2.11 Decimals Box

4. Provide labels for your variable(s) (See Figure 2.12).

How? (1) Click below **Label** , and (2) type the label for the variable in the cell.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

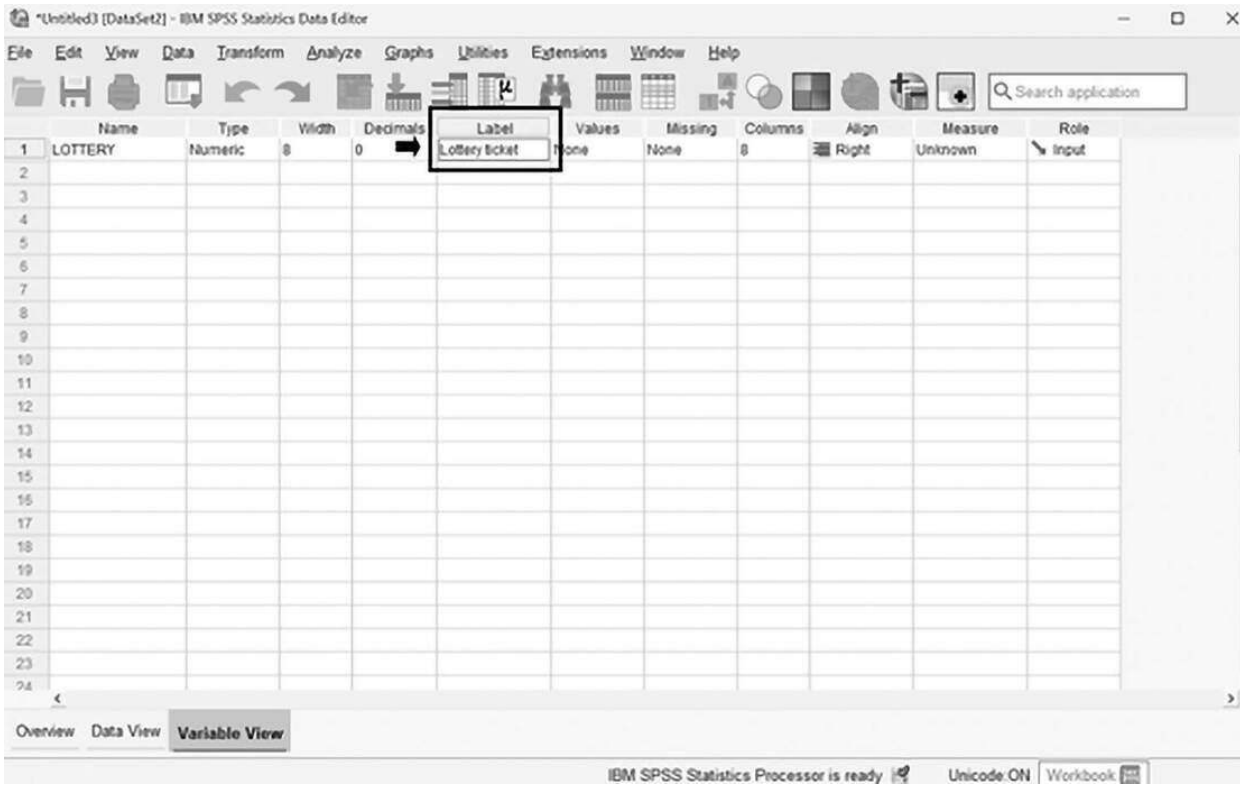


Figure 2.12 Label Box

5. Provide labels for values of nominal (categorical) variable(s) (See Figure 2.13).

How? (1) Click below **Values** , and (2) type labels for coded values of the variable.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

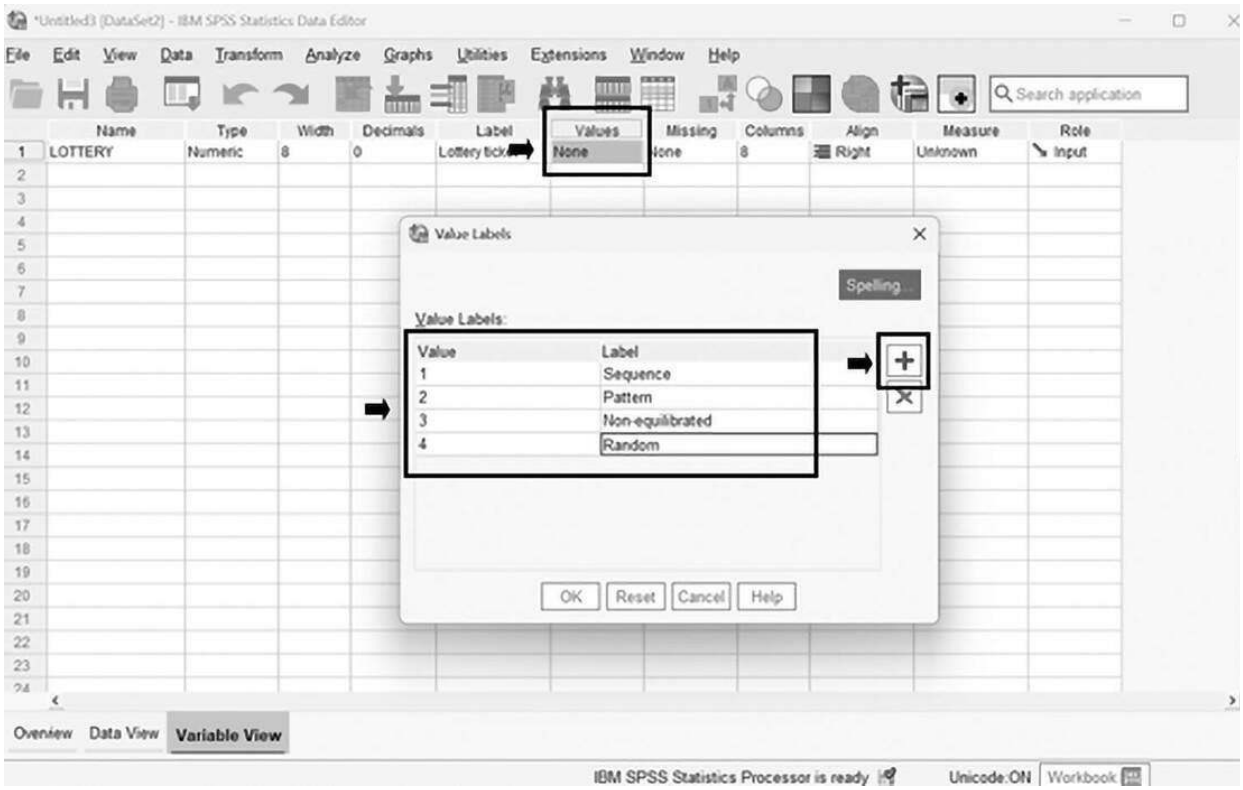


Figure 2.13 Nominal Variables

6. Enter data (See Figure 2.14).

How? (1) Click **Data View** tab and (2) enter data for the variable in the cells.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

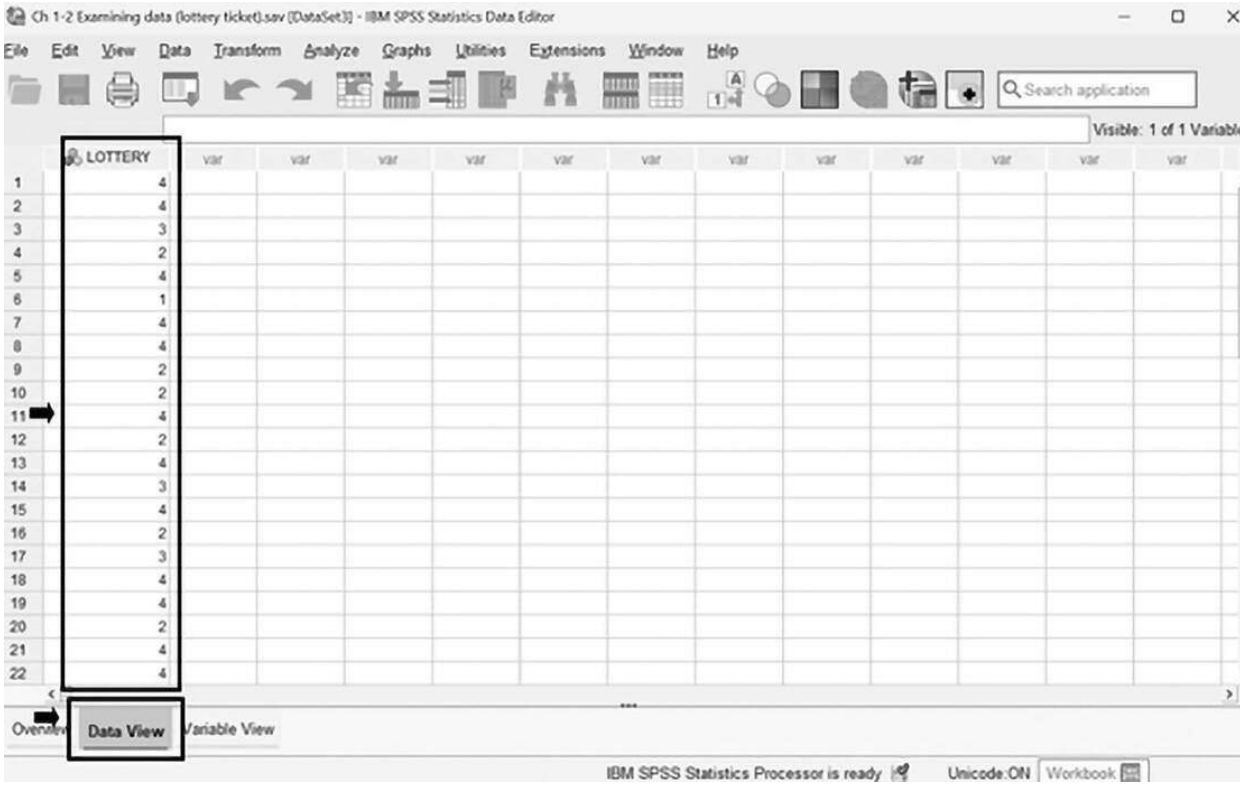


Figure 2.14 Enter Data View

Examining Data: Frequency Distribution Tables and Figures: The Lottery Study (2.1)

1. Select the frequency distribution procedure within SPSS (See Figure 2.15).

   How? (1) Click **Analyze** menu, (2) click **Descriptive Statistics** , and (3) click **Frequencies** .

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

To Jennifer , my kids Meagan and Eika , and my parents, Katsumi and Grayce Tokunaga.

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

As a global academic publisher, Sage is driven by the belief that research and education are critical in shaping society. Our mission is building bridges to knowledge—supporting the development of ideas into scholarship that is certified, taught, and applied in the real world.

Sage's founder, Sara Miller McCune, transferred control of the company to an independent trust, which guarantees our independence indefinitely. This enables us to support an equitable academic future over the long term by building lasting relationships, championing diverse perspectives, and co-creating social and behavioral science resources that transform teaching and learning.

▶ Long Description:

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

FUNDAMENTAL STATISTICS FOR THE SOCIAL AND BEHAVIORAL SCIENCES

Third Edition

Howard T. Tokunaga

San Jose State University



join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

eISBN: 978-1-0718-6160-8

Copyright © 2027 by Sage.

All rights reserved. Except as permitted by U.S. copyright law, no part of this work may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without permission in writing from the publisher.

All third party trademarks referenced or depicted herein are included solely for the purpose of illustration and are the property of their respective owners. Reference to these trademarks in no way indicates any relationship with, or endorsement by, the trademark owner.



FOR INFORMATION:

2455 Teller Road, Thousand Oaks, California 91320, E-mail: orders@sagepub.com

3rd floor, HYLO, 103-105 Bunhill Row, London, EC1Y 8LZ, United Kingdom

10th Floor, Emaar Capital Tower 2, MG Road Sikanderpur, Sector 26, Gurugram, Haryana – 122002, India

18 Cross Street #10-10/11/12, China Square Central, Singapore 048423

Library of Congress Control Number: 2026943277

ISBN: 978-1-0718-6144-8

Printed in the United States of America

This book is printed on acid-free paper.

26 27 28 29 30 10 9 8 7 6 5 4 3 2 1

Acquisitions Editor: Anna Villarruel

Content Development Editor: Kenzie Offley

Editorial Assistant: Daria Terry

Production Editor: Syeda Aina Rahat Ali

Copy Editor: Michelle Ponce

Typesetter: Lumina Datamatics

Cover Designer: Candice Harman

Marketing Manager: Victoria Velasquez

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

CONTENTS

Preface
Acknowledgments
About the Author
Chapter 1 Introduction to Statistics
Chapter 2 Examining Data: Tables and Figures
Chapter 3 Measures of Central Tendency
Chapter 4 Measures of Variability
Chapter 5 Normal Distributions
Chapter 6 Probability and Introduction to Hypothesis Testing
Chapter 7 Testing One Sample Mean
Chapter 8 Estimating the Mean of a Population
Chapter 9 Testing the Difference Between Two Means
Chapter 10 Errors in Hypothesis Testing, Statistical Power, and Effect Size
Chapter 11 One-Way Analysis of Variance (ANOVA)
Chapter 12 Two-Way Analysis of Variance (ANOVA)
Chapter 13 Correlation
Chapter 14 Linear Regression and Multiple Correlation
Chapter 15 Chi-Square
Appendix A: Tables
Appendix B: Review of Basic Mathematics
Glossary
References
Index

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

DETAILED CONTENTS

Preface

Acknowledgments

About the Author

Chapter 1 Introduction to Statistics

- 1.1 What Is Statistics?
- 1.2 Why Learn Statistics?
- 1.3 Introduction to the Stages of the Research Process
  - Developing a Research Hypothesis to Be Tested
    - Identifying a Question or Issue to Be Examined
    - Reviewing and Evaluating Relevant Theories and Research
    - Stating a Research Hypothesis: Independent and Dependent Variables
  - Collecting Data
    - Drawing a Sample From a Population
    - Determining How Variables Will Be Measured: Levels of Measurement
    - Selecting a Method to Collect the Data: Experimental and Nonexperimental Research Methods
  - Analyzing the Data
    - Calculating Descriptive Statistics
    - Calculating Inferential Statistics
  - Drawing a Conclusion Regarding the Research Hypothesis
  - Communicating the Findings of the Study
- 1.4 Plan of the Book
- 1.5 Looking Ahead
- 1.6 Summary
- 1.7 Important Terms
- 1.8 Exercises

Answers to Learning Checks

Learning Check 1

Learning Check 2

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Odd-Numbered Exercises

Chapter 2 Examining Data: Tables and Figures

2.1 An Example From the Research: Winning the Lottery

2.2 Why Examine Data?

Gaining an Initial Sense of the Data

Identifying Outliers

Evaluating Research Methodology

Determining Whether Data Meet Statistical Criteria and Assumptions

2.3 Examining Data Using Tables

What Are the Values of the Variable?

How Many Participants in the Sample Have Each Value of the Variable?

What Percentage of the Sample Has Each Value of the Variable?

2.4 Grouped Frequency Distribution Tables

Cumulative Percentages

2.5 Examining Data Using Figures

Displaying Nominal and Ordinal Variables: Bar Charts and Pie Charts

Displaying Interval and Ratio Variables: Histograms and Frequency Polygons

Drawing Inappropriate Conclusions From Figures

2.6 Examining Data: Describing Distributions

Modality

Symmetry

Variability

2.7 Looking Ahead

2.8 Summary

2.9 Important Terms

2.10 Formula Introduced in This Chapter

Percentage (%)

2.11 Using IBM SPSS Software *

Creating Data Files: Defining Variables and Entering Data

Examining Data: Frequency Distribution Tables and Figures: The Lottery Study (2.1)

2.12 Exercises

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 3 Measures of Central Tendency
3.1 An Example From the Research: The 10% Myth
3.2 Understanding Central Tendency
3.3 The Mode
Determining the Mode
3.4 The Median
Calculating the Median: Odd Number of Scores
Calculating the Median: Even Number of Scores
Calculating the Median When Multiple Scores Have the Median Value
3.5 The Mean
Calculating the Mean
Reporting Calculated Numbers
Rounding Calculated Numbers
Calculating the Mean From Frequency Distribution Tables
The Mean as a Balancing Point in a Distribution
The Sample Mean vs. the Population Mean: Statistics vs. Parameters
3.6 Comparison of the Mode, Median, and Mean
Strengths and Weaknesses of the Mode
Strengths and Weaknesses of the Median
Strengths and Weaknesses of the Mean
Choosing a Particular Measure of Central Tendency
3.7 Measures of Central Tendency: Drawing Conclusions
3.8 Looking Ahead
3.9 Summary
3.10 Important Terms
3.11 Formulas Introduced in This Chapter
Median (Odd Number of Scores)
Median (Even Number of Scores)

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Mean
Mean (Frequency Distribution Table)
Mean (Balancing Point in a Distribution)
3.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 4 Measures of Variability
4.1 An Example From the Research: How Many Sometimes" in an Always"?
4.2 Understanding Variability
4.3 The Range
Strengths and Weaknesses of the Range
4.4 The Interquartile Range
Strengths and Weaknesses of the Interquartile Range
4.5 The Variance (s^2)
Definitional Formula for the Variance
Computational Formula for the Variance
Why Not Use the Absolute Value of the Deviation in Calculating the Variance?
Why Divide by N − 1 Rather Than N in Calculating the Variance?
4.6 The Standard Deviation (s)
Definitional Formula for the Standard Deviation
Computational Formula for the Standard Deviation
Eyeball Estimating" the Mean and Standard Deviation
4.7 Measures of Variability for Populations
The Population Variance (σ^2)
The Population Standard Deviation (σ)
Measures of Variability for Samples vs. Populations
4.8 Measures of Variability for Nominal Variables
The Index of Qualitative Variation
4.9 Measures of Variability: Drawing Conclusions

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

4.10 Looking Ahead
4.11 Summary
4.12 Important Terms
4.13 Formulas Introduced in This Chapter
Range
Interquartile Range
Variance (s^2) (Definitional Formula)
Variance (s^2) (Computational Formula)
Standard Deviation (s) (Definitional Formula)
Standard Deviation (s) (Computational Formula)
Population Variance (σ^2) (Definitional Formula)
Population Standard Deviation (σ) (Definitional Formula)
Index of Qualitative Variation (IQV)
4.14 Using SPSS
Calculating Measures of Central Tendency and Variability: The Frequency Rating Study (4.1)
4.15 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 5 Normal Distributions
5.1 Normal Distributions
What Are Normal Distributions?
Why Are Normal Distributions Important?
Is There Only One Normal Distribution?
5.2 The Standard Normal Distribution
What Is the Standard Normal Distribution?
Interpreting z-Scores: Distance From the Mean
Interpreting z-Scores: Relative Position Within the Distribution
Finding the Area Between the Mean and a z-Score

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Finding the Area Less or Greater Than a z-Score
Finding the Area Between Two z-Scores
5.3 Applying z-Scores to Normal Distributions
Transforming Normal Distributions Into the Standard Normal Distribution
Understanding the z-Score Formula
Example: Transforming SAT Scores Into z-Scores
Interpreting z-Scores: Distance From the Mean
Interpreting z-Scores: Relative Position Within Distribution
Finding the Area Between the Mean and a Score
Finding the Area Less or Greater Than a Score
Finding the Area Between Two Scores
Transforming Scores Into z-Scores: Drawing Conclusions
5.4 Standardizing Frequency Distributions
Example: Measuring Politicians' Behavior
Standardizing a Frequency Distribution
Interpreting Standardized Scores
Using Standardized Scores to Compare or Combine Variables
Characteristics of Standardized Distributions
When Can You Apply the Normal Curve Table to a Standardized Distribution?
5.5 Looking Ahead
5.6 Summary
5.7 Important Terms
5.8 Formulas Introduced in This Chapter
z-Score
Standardized Score
5.9 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 6 Probability and Introduction to Hypothesis Testing

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

6.1 A Brief Introduction to Probability
- What Is Probability?
- Why Is Probability Important to Researchers?
- Applying Probability to Normal Distributions
- Applying Probability to Binomial Distributions

6.2 Example: Making Heads or Tails of the Super Bowl

6.3 Introduction to Hypothesis Testing
- State the Null and Alternative Hypotheses (H_0 and H_1)
  - The Null Hypothesis (H_0)
  - The Alternative Hypothesis (H_1)
- Make a Decision About the Null Hypothesis
  - Defining a Low" Probability of a Statistic: Setting Alpha (α)
  - Identifying the Values of the Statistic With a Low Probability
  - Stating a Decision Rule
  - Calculating a Value of a Statistic
  - Making the Decision Whether to Reject the Null Hypothesis
- Draw a Conclusion From the Analysis
- Relate the Result of the Analysis to the Research Hypothesis
- Summary of Steps in Hypothesis Testing

6.4 Issues Related to Hypothesis Testing: An Introduction
- The Issue of Proof" in Hypothesis Testing
- Errors in Decision Making
- Factors Influencing the Decision About the Null Hypothesis
  - Sample Size
  - Alpha (α)
  - Directionality of the Alternative Hypothesis

6.5 Looking Ahead

6.6 Summary

6.7 Important Terms

6.8 Formula Introduced in This Chapter

Probability

6.9 Exercises

Answers to Learning Checks

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 1
Learning Check 2
Learning Check 3
Answers to Odd-Numbered Exercises
Chapter 7 Testing One Sample Mean
7.1 An Example From the Research: Do You Read Me?
7.2 The Sampling Distribution of the Mean
Characteristics of the Sampling Distribution of the Mean
7.3 Inferential Statistics: Testing One Sample Mean (σ Known)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: z-Test for One Mean
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the z-Test for One Mean
Summary
7.4 A Second Example From the Research: Unique Invulnerability
7.5 Introduction to the t-Distribution
Characteristics of the Student t-Distribution
7.6 Inferential Statistics: Testing One Sample Mean (σ Not Known)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: t-Test for One Mean
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the t-Test for One Mean

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Summary
7.7 Factors Affecting the Decision About the Null Hypothesis
Sample Size
Alpha (α)
Directionality of the Alternative Hypothesis
7.8 Looking Ahead
7.9 Summary
7.10 Important Terms
7.11 Formulas Introduced in This Chapter
z-Test for One Mean
Population Standard Error of the Mean $\left(\mathbf{i}t\sigma_{\mathbf{i}t\overline{X}}\right)$
Degrees of Freedom (df) for One Sample
t-Test for One Mean
Standard Error of the Mean $\left(\mathbf{i}ts_{\mathbf{i}t\overline{X}}\right)$
7.12 Using SPSS
Testing One Sample Mean (σ Not Known): The Unique Invulnerability Study (7.6)
7.13 Exercises
Answers to Learning Checks
Learning Check 2
Learning Check 3
Learning Check 5
Learning Check 6
Answers to Odd-Numbered Exercises
Chapter 8 Estimating the Mean of a Population
8.1 An Example From the Research: Salary Survey
8.2 Introduction to the Confidence Interval for the Mean
Using the Sampling Distribution of the Mean to Create the Confidence Interval for the Mean
8.3 The Confidence Interval for the Mean (σ Not Known)
State the Desired Level of Confidence
Calculate the Confidence Interval and Confidence Limits
Calculate the Standard Error of the Mean $\left(s_{\overline{X}}\right)$

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Calculate the Degrees of Freedom (df) and Identify the Critical Value of t

Calculate the Confidence Interval and Confidence Limits

Draw a Conclusion About the Confidence Interval

Probability of an Interval or the Probability of a Population Mean?

Summary

8.4 The Confidence Interval for the Mean (σ Known)

State the Desired Level of Confidence

Calculate the Confidence Interval and Confidence Limits

Calculate the Population Standard Error of the Mean $\left(\sigma_{\overline{X}}\right)$

Identify the Critical Value of z

Calculate the Confidence Interval and Confidence Limits

Draw a Conclusion About the Confidence Interval

Summary

8.5 Factors Affecting the Width of the Confidence Interval for the Mean

Sample Size

Estimating the Sample Size Needed for a Desired Confidence Interval

Level of Confidence

Choosing Between Different Levels of Confidence: Probability Versus Precision

8.6 Interval Estimation and Hypothesis Testing

Concerns About Hypothesis Testing

The Interpretation of Statistical Significance and Nonsignificance

The Influence of Sample Size on the Decision About the Null Hypothesis

Interval Estimation as an Alternative to Hypothesis Testing

8.7 Looking Ahead

8.8 Summary

8.9 Important Terms

8.10 Formulas Introduced in This Chapter

Confidence Interval for the Mean (σ Not Known)

Confidence Interval for the Mean (σ Known)

Estimated Sample Size for the Confidence Interval for the Mean

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

8.11 Using SPSS
Calculating the confidence interval for the mean (The salary survey study (8.1))
8.12 Exercises
Answers to Learning Checks
Learning Check 1
Learning Check 2
Learning Check 3
Learning Check 4
Answers to Odd-Numbered Exercises
Chapter 9 Testing the Difference Between Two Means
9.1 An Example From the Research: You Can Just Wait
9.2 The Sampling Distribution of the Difference
Characteristics of the Sampling Distribution of the Difference
9.3 Inferential Statistics: Testing the Difference Between Two Sample Means
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Calculate the Degrees of Freedom (df)
Set Alpha (α), Identify the Critical Values, and State a Decision Rule
Calculate a Statistic: t-Test for Independent Means
Make a Decision Whether to Reject the Null Hypothesis
Determine the Level of Significance
Draw a Conclusion From the Analysis
Relate the Result of the Analysis to the Research Hypothesis
Assumptions of the t-Test for Independent Means
Summary
9.4 Inferential Statistics: Testing the Difference Between Two Sample Means (Unequal Sample Sizes)
An Example From the Research: Kids' Motor Skills and Fitness
Inferential Statistics: Testing the Difference Between Two Sample Means (Unequal Sample Sizes)
State the Null and Alternative Hypotheses (H_0 and H_1)
Make a Decision About the Null Hypothesis
Draw a Conclusion From the Analysis

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Relate the Result of the Analysis to the Research Hypothesis

Assumptions of the t-Test for Independent Means (Unequal Sample Sizes)

Summary

9.5 Inferential Statistics: Testing the Difference Between Paired Means

Research Situations Appropriate for Within-Subjects Research Designs

An Example From the Research: Web-Based Family Interventions

Inferential Statistics: Testing the Difference Between Paired Means

Calculate the Difference Between the Paired Scores

State the Null and Alternative Hypotheses (H_0 and H_1)

Make a Decision About the Null Hypothesis

Draw a Conclusion From the Analysis

Relate the Result of the Analysis to the Research Hypothesis

Assumptions of the t-Test for Dependent Means

Summary

9.6 Looking Ahead

9.7 Summary

9.8 Important Terms

9.9 Formulas Introduced in This Chapter

Degrees of Freedom (df), Difference Between Two Sample Means

t-Test for Independent Means

Standard Error of the Difference $\left(s_{\overline{X}_1-\overline{X}_2}\right)$

Standard Error of the Difference (Unequal Sample Sizes)

Degrees of Freedom (df), Difference Between Paired Means

t-Test for Dependent Means

Standard Error of Difference Scores (s D)

9.10 Using SPSS

Testing the Difference Between Two Sample Means: The Parking Lot Study (9.1)

Testing the Difference Between Paired Means: The Web-Based Intervention Study (9.5)

9.11 Exercises

Answers to Learning Checks

Learning Check 1

join this discord server to get full pdf: discord.gg/EVabegZqJH

join this discord server to get full pdf: discord.gg/EVabegZqJH

Learning Check 2
Learning Check 3
Learning Check 4
Learning Check 5
Answers to Odd-Numbered Exercises
Chapter 10 Errors in Hypothesis Testing, Statistical Power, and Effect Size
10.1 Hypothesis Testing vs. Criminal Trials
10.2 An Example From the Research: Truth or Consequences
Correct Decisions and Mistakes in Criminal Trials
10.3 Two Errors in Hypothesis Testing: Type I and Type II Error
Type I Error: The Risk in Rejecting the Null Hypothesis
What Is the Probability of Making a Type I Error?
What Is the Probability of Not Making a Type I Error?
Why Is Type I Error a Concern?
Type II Error: The Risk in Not Rejecting the Null Hypothesis
What Is the Probability of Making a Type II Error?
What Is the Probability of Not Making a Type II Error?
Why Is Type II Error a Concern?
Why Does Type II Error Occur?
Type I and Type II Error: Summary
10.4 Controlling Type I and Type II Error
Controlling Type I Error
Concerns About Controlling Type I Error
Controlling Type II Error
Increasing Sample Size
Raising Alpha (α)
Using a Directional Alternative Hypothesis (H1)
Increasing Between-Group Variability
Decreasing Within-Group Variability
Controlling Type I and Type II Error: Summary
10.5 Measures of Effect Size
Interpreting Inferential Statistics: The Influence of Sample Size
The Inappropriateness of “Highly Significant”

join this discord server to get full pdf: discord.gg/EVabegZqJH