CENGAGE

Introduction to

STATISTICS & DATA ANALYSIS

Roxy Peck ◆ *Tom Short* ◆ *Chris Olsen*

Sixth Edition

Index of Applications in Examples and Activities

Act: Activity; Ex: Example

Education and Child Development

Childcare for preschoolers: Ex 4.15
College plans of high school seniors: Ex 7.4
Combining exam scores: Ex 7.16
Hand gestures in learning: Ex 2.7, Ex 2.9
IQ scores: Ex 4.16, Ex 7.28
Is education worth the cost?: Ex 3.1
Note-taking methods: Ex 2.4
Predictors of writing competence: Ex 14.4
School enrollment in Northern and Central Africa: Ex 3.11
Standardized test scores: Ex 4.14, Ex 10.16
Students' knowledge of geography: Act 3.1

Environmental Science

Bison of Yellowstone Park: Ex 13.4
Cosmic radiation: Ex 9.7
Lead in tap water: Ex 10.7
Rainfall frequency distributions for Albuquerque: Ex 3.19
River water velocity and distance from shore: Ex 5.17
Soil and sediment characteristics: Ex 14.12, Ex 14.14, Ex 14.16 (online)
Water conservation: Ex 10.10
Water quality: Ex 1.2

Food Science

Calorie consumption at fast food restaurants: Ex 2.2
Effect of exercise on food intake: Ex 4.1
Fat content of hot dogs: Ex 8.6

Leisure and Popular Culture

Babies on social media: Ex 9.5
Car preferences: Ex 6.1
Do U Txt?: Ex 1.6
Facebook and academic performance: Ex 11.1
iPod shuffles: Ex 7.7
Jeopardy! nerds: Ex 12.1, Ex 12.2
Life insurance for cartoon characters: Ex 3.2
Number of trials required to complete game: Ex 7.2
Probability a Hershey's Kiss will land on its base: Act 6.1
Selecting cards: Ex 6.20
Selection of contest winners: Ex 6.7
Tossing a coin: Ex 6.8
Video game performance: Ex 4.9, Ex 4.10, Ex 6.2, Ex 6.4
Word cloud representation of a document: Ex 11.5, Ex 11.6, Ex 11.8

Manufacturing and Industry

Bottled soda volumes: Ex 8.5
Computer configurations: Ex 6.19
Computer sales: Ex 7.19
Corrosion of underground pipe coatings: Ex 15.14 (online)
Customer hold time: Ex 10.17
Durable press rating of cotton fabric: Ex 14.18 (online)
Engineering stress test: Ex 7.3
Ergonomic characteristics of stool designs: Ex 15.10, Ex 15.11 (online)
Garbage truck processing times: Ex 7.30
GFI switches: Ex 6.12
Lifetime of compact florescent lightbulbs: Ex 10.2
On-time package delivery: Ex 10.18
Paint flaws: Ex 7.6
Smart phone warranties: Ex 6.24
Strength of bark board: Ex 16.4 (online)
Testing for flaws: Ex 7.11, Ex 7.12

Marketing and Consumer Behavior

Budgets and tracking spending: Ex 7.21
Car choices: Ex 6.10
Energy efficient refrigerators: Ex 7.5
High-pressure sales tactics: Ex 16.13 (online)
Impact of food labels: Ex 10.8
Satisfaction with cell phone service: Ex 4.6

Medical Science

Affect of long work hours on sleep: Ex 11.10
Anti-clotting medications after hip or knee surgery: Ex 11.15
Apgar scores: Ex 7.10, Ex 7.13
Blood platelet volume: Ex 8.2
Blood pressure and kidney disease: Ex 16.5 (online)
Blue light exposure and blood glucose level: Ex 11.12
Body mass index (BMI): Ex 7.14
Chronic airflow obstruction: Ex 16.9 (online)
Contracting hepatitis from blood transfusion: Ex 8.8, Ex 8.9
Cooling treatment after oxygen deprivation in newborns: Ex 2.6
Diagnosing tuberculosis: Ex 6.15
Drive-through medicine: Ex 9.8
Early detection of lung cancer: Ex 10.6
Effects of ethanol on sleep time: Ex 15.6
Effect of school start time on sleep patterns: Ex 3.16
Evaluating disease treatments: Ex 10.3
Exercise and sleep quality: Ex 12.8
Facial expression and self-reported pain level: Ex 12.7
Growth hormone levels and diabetes: Ex 16.10 (online)
Heart attacks in high-rise buildings: Ex 12.4, Ex 12.5
Hip-to-waist ratio and risk of heart attack: Ex 14.2
Hormones and body fat: Ex 15.4, Ex 15.5
Lead exposure and brain volume: Ex 5.12
Liver injuries in newborns: Ex 9.15
Lyme disease: Ex 6.27
Markers for kidney disease: Ex 7.33
Maternal age and baby's birth weight: Ex 13.2
Medical errors: Ex 6.9
Oxytocin nasal spray and social interaction: Ex 11.16
Parental smoking and infant health: Ex 16.2, Ex 16.3 (online)

Physical Sciences

Politics and Public Policy

Psychology, Sociology, and Social Issues

Public Health and Safety

Sports

Surveys and Opinion Polls

Transportation

EDITION
6

Introduction to Statistics and Data Analysis

Roxy Peck
California Polytechnic State University, San Luis Obispo

Tom Short
West Chester University of Pennsylvania

Chris Olsen
Grinnell College

CENGAGE

Australia • Brazil • Mexico • Singapore • United Kingdom • United States

CENGAGE

Introduction to Statistics and Data Analysis, **Sixth Edition**
Roxy Peck, Tom Short, Chris Olsen

Product Director: Mark Santee

Product Manager: Catherine Van Der Laan

Product Assistant: Amanda Rose

Marketing Manager: Mike Saver

Learning Designer: Elinor Gregory

Subject Matter Expert: Morgan Johnson

Content Managers: Brendan Killion and Abby DeVeuve

Manufacturing Planner: Doug Bertke

IP Analyst: Reba Frederics

IP Project Manager: Carly Belcher

Art Director: Vernon T. Boes

Design and Production Services/Compositor: MPS Limited

Cover Image: Tawatchai Prakobit/EyeEm/ Getty Images

Library of Congress Control Number: 2018949851

Student Edition:
ISBN: 978-1-337-79361-2

Loose-leaf Edition:
ISBN: 978-1-337-79432-9

Annotated Instructor's Edition:
ISBN: 978-1-337-79415-2

Cengage
20 Channel Center Street
Boston, MA 02210
USA

Cengage is a leading provider of customized learning solutions with employees residing in nearly 40 different countries and sales in more than 125 countries around the world. Find your local representative at **www.cengage.com**.

Cengage products are represented in Canada by Nelson Education, Ltd.

To learn more about Cengage platforms and services, register or access your online learning solution, or purchase materials for your course, visit **www.cengage.com**.

Printed in the United States of America
Print Number: 01 Print Year: 2018

To Lygia and Kyle

Roxy Peck

To my children: Bob, Kathy, Peter, and Kellen

Tom Short

To my wife, Sally, and my daughter, Anna

Chris Olsen

Author Bios

ROXY PECK is Emerita Associate Dean of the College of Science and Mathematics and Professor of Statistics Emerita at California Polytechnic State University, San Luis Obispo. A faculty member at Cal Poly from 1979 until 2009, Roxy served for 6 years as Chair of the Statistics Department before becoming Associate Dean, a position she held for 13 years. She received an M.S. in Mathematics and a Ph.D. in Applied Statistics from the University of California, Riverside. Roxy is nationally known in the area of statistics education, and she was presented with the Lifetime Achievement Award in Statistics Education at the U.S. Conference on Teaching Statistics in 2009. In 2003 she received the American Statistical Association's Founder's Award, recognizing her contributions to K–12 and undergraduate statistics education. She is a Fellow of the American Statistical Association and an elected member of the International Statistics Institute. Roxy served for 5 years as the Chief Reader for the Advanced Placement (AP®) Statistics Exam and has chaired the American Statistical Association's Joint Committee with the National Council of Teachers of Mathematics on Curriculum in Statistics and Probability for Grades K–12 and the Section on Statistics Education. In addition to her texts in introductory statistics, Roxy is also co-editor of *Statistical Case Studies: A Collaboration Between Academe and Industry* and a member of the editorial board for *Statistics: A Guide to the Unknown,* 4th edition. Outside the classroom, Roxy likes to travel and spends her spare time reading mystery novels. She also collects Navajo rugs and heads to Arizona and New Mexico whenever she can find the time.

TOM SHORT is an Associate Professor in the Statistics Program within the Department of Mathematics at West Chester University of Pennsylvania. He previously held faculty positions at Villanova University, Indiana University of Pennsylvania, and John Carroll University. He is a Fellow of the American Statistical Association and received the 2005 Mu Sigma Rho Statistics Education Award. Tom is part of the leadership team for readings of the Advanced Placement® Statistics Exam, and was a member of the AP® Statistics Development Committee. He has also served on the Board of Directors of the American Statistical Association. Tom treasures the time he shares with his children and the many adventures experienced with his wife, Darlene.

CHRIS OLSEN taught statistics in Cedar Rapids, Iowa, for over 25 years, and at Cornell College and Grinnell College. Chris is a past member (twice!) of the Advanced Placement® Statistics Test Development Committee and has been a table leader and question leader at the AP® Statistics reading for 11 years. He is a long-time consultant to the College Board, and Chris has led workshops and institutes for AP® Statistics teachers in the United States and internationally. Chris was the Iowa recipient of the Presidential Award for Excellence in Science and Mathematics Teaching in 1986, a regional awardee of the IBM Computer Teacher of the Year in 1988, and received the Siemens Award for Advanced Placement in mathematics in 1999. Chris is a frequent contributor to and is moderator of the AP® Statistics Teacher Community online. He is currently a member of the editorial board of *Teaching Statistics*. Chris graduated from Iowa State University with a major in mathematics. While acquiring graduate degrees at the University of Iowa, he concentrated on statistics, computer programming and psychometrics. In his spare time he enjoys reading and hiking. He and his wife have a daughter, Anna, a Caltech graduate in Civil Engineering. She is a risk modeler at RMS, the world's leading catastrophe risk modeling company.

Brief Contents

Sections and/or chapter numbers shaded in color can be found at www.cengage.com

Contents

Sections and/or chapter numbers shaded in color can be found at www.cengage.com

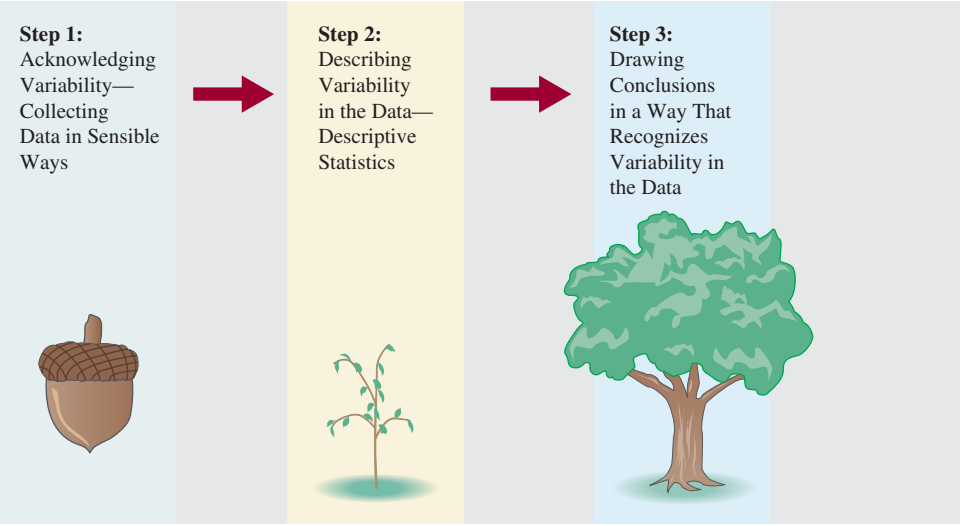
Preface

In a nutshell, statistics is about understanding the role that variability plays in drawing conclusions based on data. *Introduction to Statistics and Data Analysis,* Sixth Edition, develops this crucial understanding of variability through its focus on the data analysis process.

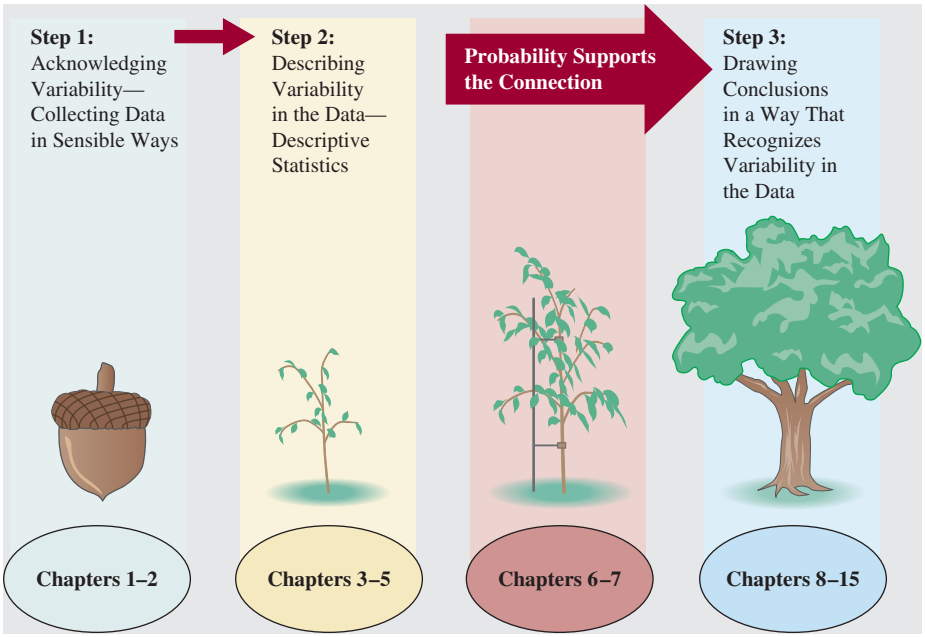
An Organization That Reflects the Data Analysis Process

Students are introduced early to the idea that data analysis is a process that begins with careful planning, followed by data collection, data description using graphical and numerical summaries, data analysis, and finally interpretation of results. This process is described in detail in Chapter 1, and the ordering of topics in the first ten chapters of the book mirrors the process: data collection, then data description, then statistical inference.

The logical order in the data analysis process can be pictured as shown in the following figure.



Unlike many introductory texts, *Introduction to Statistics and Data Analysis,* Sixth Edition, is organized in a way that is consistent with the natural order of the data analysis process:



The Importance of Context and Real Data

Statistics is not about numbers; it is about data—numbers in context. It is the context that makes a problem meaningful and something worth considering. For example, exercises that ask students to calculate the mean of 10 numbers or to construct a dotplot or boxplot of 20 numbers without context are arithmetic and graphing exercises. They become statistics problems only when a context gives them meaning and allows for interpretation. While this makes for a text that may appear "wordy" when compared to traditional mathematics texts, it is a critical and necessary component of a modern statistics text.

Examples and exercises with overly simple settings do not allow students to practice interpreting results in authentic situations or give students the experience necessary to be able to use statistical methods in real settings. We believe that the exercises and examples are a particular strength of this text, and we invite you to compare the examples and exercises with those in other introductory statistics texts.

Many students are skeptical of the relevance and importance of statistics. Contrived problem situations and artificial data often reinforce this skepticism. A strategy that we have employed successfully to motivate students is to present examples and exercises that involve data extracted from journal articles, newspapers, and other published sources. Most examples and exercises in the book are of this nature; they cover a very wide range of disciplines and subject areas. These include, but are not limited to, health and fitness, consumer research, psychology and aging, sports, environmental research, law and criminal justice, and entertainment.

A Focus on Interpretation and Communication

Most chapters include a section titled "Interpreting and Communicating the Results of Statistical Analyses." These sections include advice on how to best communicate the results of a statistical analysis and also consider how to interpret statistical summaries found in journals and other published sources. Subsections titled "A Word to the Wise" reminds readers of things that must be considered in order to ensure that statistical methods are employed in reasonable and appropriate ways.

Consistent with Recommendations for the Introductory Statistics Course Endorsed by the American Statistical Association

In 2016, the American Statistical Association updated the report "Guidelines in Assessment and Instruction for Statistics Education (GAISE) College Report 2016," which included the following six recommendations for the introductory statistics course:

1. Teach statistical thinking.
2. Focus on conceptual understanding.
3. Integrate real data with a context and purpose.
4. Foster active learning.
5. Use technology to explore concepts and analyze data.
6. Use assessments to improve and evaluate student learning.

Introduction to Statistics and Data Analysis, Sixth Edition, is consistent with these recommendations and supports the GAISE guidelines in the following ways:

1. **Teach statistical thinking.**
   Statistical thinking and statistical literacy literacy are promoted throughout the text in the many examples and exercises that are drawn from the popular press. In addition, a focus on the role of variability, consistent use of context, and an emphasis on interpreting and communicating results in context work together to help students develop skills in statistical thinking.
2. **Focus on conceptual understanding.**
   Nearly all exercises in *Introduction to Statistics and Data Analysis,* Sixth Edition, are multipart and ask students to go beyond just computation. They focus on interpretation and communication, not just in the chapter sections specifically devoted to this topic, but throughout the text. The examples and explanations are designed to promote conceptual understanding. Hands-on activities in each chapter are also constructed to strengthen conceptual understanding. Which brings us to . . .
3. **Integrate real data with a context and a purpose.**
   The examples and exercises from *Introduction to Statistics and Data Analysis,* Sixth Edition, are context driven, and the reference sources include the popular press as well as journal articles.
4. **Foster active learning.**
   While this recommendation speaks more to pedagogy and classroom practice, *Introduction to Statistics and Data Analysis,* Sixth Edition, provides more than 30 hands-on activities in the text and additional activities in the accompanying instructor resources that can be used in class or assigned to be completed outside of class.
5. **Use technology to explore concepts and analyze data.**
   The computer brings incredible statistical power to the desktop of every investigator. The wide availability of statistical computer packages such as Minitab, JMP, and SPSS, and the graphical capabilities of the modern microcomputer have transformed both the teaching and learning of statistics. To highlight the role of the computer in contemporary statistics, we have included sample output throughout the book. In addition, numerous exercises contain data that can easily be analyzed using statistical software. However, access to a particular statistical package is not assumed. Technology manuals for specific software packages and for the graphing calculator are available in the online materials that accompany this text. The sixth edition of *Introduction to Statistics and Data Analysis* also includes a number of Shiny web apps that can be used to illustrate statistical concepts and to implement the simulation-based inference methods covered in new optional sections.

   The appearance of handheld calculators with significant statistical and graphing capability has also changed statistics instruction in classrooms where access to computers is still limited. There is not universal or even wide agreement about the proper role for the graphing calculator in college statistics classes, where access to a computer is more common. At the same time, for tens of thousands of students in

Advanced Placement Statistics in our high schools, the graphing calculator is the only dependable access to statistical technology.

This text allows the instructor to balance the use of computers and calculators in a manner consistent with his or her philosophy. As with computer packages, our exposition avoids assuming the use of a particular calculator and presents the calculator capabilities in a generic format. For those using a TI graphing calculator, there is a technology manual available in the online materials that accompany this text.

6. **Use assessments to improve and evaluate student learning.** Assessment materials in the form of a test bank, quizzes, and chapter exams are available in the instructor resources that accompany this text. The items in the test bank reflect the data-in-context philosophy of the text's exercises and examples.

Advanced Placement® Statistics

We have designed this book with a particular eye toward the syllabus of the Advanced Placement® Statistics course and the needs of high school teachers and students. Concerns expressed and questions asked in teacher workshops and on the AP® Statistics teacher community have strongly influenced our explanation of certain topics, especially in the area of experimental design and probability. We have taken great care to provide precise definitions and clear examples of concepts that Advanced Placement® Statistics instructors have acknowledged as difficult for their students. We have also expanded the variety of examples and exercises, recognizing the diverse potential futures envisioned by very capable students who have not yet focused on a college major. The AP® edition of this text also contains a collection of multiple choice and free response questions that can be used to help students review for the AP® Statistics exam.

Topic Coverage

Our book can be used in courses as short as one quarter or as long as one year in duration. Particularly in shorter courses, an instructor will need to be selective in deciding which topics to include and which to set aside. The book divides naturally into four major sections: collecting data and descriptive methods (Chapters 1–5), probability (Chapters 6–8), the basic one- and two-sample inferential techniques (Chapters 9–12), and more advanced inferential methodology (Chapters 13–16).

We include an early chapter (Chapter 5) on descriptive methods for bivariate numerical data. This early exposure introduces questions and issues that should stimulate student interest in the subject. It is also advantageous for those teaching courses in which time constraints preclude covering advanced inferential material. However, this chapter can easily be postponed until the basics of inference have been covered, and then combined with Chapter 13 for a unified treatment of regression and correlation.

With the possible exception of Chapter 5, Chapters 1 through 10 should be covered in order. We anticipate that most instructors will then continue with two-sample inference (Chapter 11) and methods for categorical data analysis (Chapter 12), although regression could be covered before either of these topics. Optional portions of Chapter 14 (multiple regression) and Chapter 15 (analysis of variance) and Chapter 16 (nonparametric methods) are included in the online materials that accompany this text.

A Note on Probability

The content of the probability chapters is consistent with the Advanced Placement® Statistics course description. It includes both a traditional treatment of probability and probability distributions at an introductory level, as well as a section on the use of simulation as a tool for estimating probabilities. For those who prefer a more informal treatment of probability, see the text *Statistics: Learning from Data*, Second Edition, by Roxy Peck and Tom Short.

In This Edition

Look for the following in the Sixth Edition:

- **NEW Updated Examples and Exercises.** In our continuing effort to keep things interesting and relevant, the sixth edition contains many updated examples and exercises that use data from recent journal articles, newspapers, and web posts, on topics of interest to students.
- **NEW Sections on Randomization-Based Inference Methods.** Research indicates that randomization-based instruction in statistical inference may help learners to better understand the concepts of confidence and significance. The sixth edition includes new optional sections on randomization-based inference methods. These methods are also particularly useful in that they provide an alternative method of analysis that can be used when the conditions required for normal distribution-based inference are not met. The inference chapters (Chapters 9–11) now contain new optional sections on randomization-based inference that include bootstrap methods for simulation-based confidence intervals and randomization-based tests of hypotheses. These new sections are accompanied by online Shiny apps, which can be used to construct bootstrap confidence intervals and to carry out randomization tests.
- **Helpful hints in exercises.** To help students who might be having trouble getting started, hints have been added to many of the exercises directing students to relevant examples in the text.
- **Margin notes to highlight the importance of context and the process of data analysis.** Margin notations appear in appropriate places in the examples. These include *Understanding the context*, *Consider the data, Formulate a plan*, *Do the work*, and *Interpret the results*. These notes are designed to increase student awareness of the steps in the data analysis process.
- **Activities at the end of each chapter.** These activities can be used as a chapter capstone or can be integrated at appropriate places as the chapter material is covered in class.
- **Advanced topics** that are often omitted in a one-quarter or one-semester course, such as survey design (Section 2.6), logistic regression (Section 5.6), inference based on the estimated regression line (Sections 13.4 and 13.5), inference and variable selection methods in multiple regression (Sections 14.3 and 14.4), analysis of variance for randomized block and two-factor designs (Sections 15.3 and 15.4), and distribution-free procedures (Chapter 16) **are available in the online materials that accompany this text.**
- **Updated materials for instructors** are included on the Instructor Companion Site. In addition to the usual instructor supplements such as a complete solutions manual and a test bank, the website contains examples that can be incorporated into classroom presentations and cross-references to resources such as Fathom, *Workshop Statistics*, and *Against All Odds*. Of particular interest to those teaching Advanced Placement Statistics, the website also includes additional data analysis questions of the type encountered on the free response portion of the Advanced Placement exam, as well as a collection of model responses.

Instructor and Student Resources

MindTap™

Available via WebAssign is MindTap™ Reader, Cengage Learning's next-generation eBook. MindTap Reader provides robust opportunities for students to annotate, take notes, navigate, and interact with the text (e.g., ReadSpeaker). Annotations captured in MindTap Reader are automatically tied to the Notepad app, where they can be viewed chronologically and in a cogent, linear fashion. Instructors also can edit the text and assets in the Reader, as well as add videos or URLs.

WebAssign

WebAssign for Peck/Short/Olsen's *Introduction to Statistics and Data Analysis*, 6th Edition, is a flexible and fully customizable online instructional solution that puts powerful tools in the hands of instructors, empowering you to deploy assignments, instantly assess individual student and class performance, and help your students master the course concepts. With WebAssign's powerful digital platform and Introduction to Probability and Statistics specific content, you can tailor your course with a wide range of assignment settings, add your own questions and content, and access student and course analytics and communication tools. Learn more at www.webassign.com.

JMP Statistical Software

Access to JMP is free with the purchase of a new book.

JMP is a statistics software for Windows and Macintosh computers from SAS, the market leader in analytics software and services for industry. JMP Student Edition is a streamlined, easy-to-use version that provides all the statistical analysis and graphics covered in this textbook. Once data is imported, students will find that most procedures require just two or three mouse clicks. JMP can import data from a variety of formats, including Excel and other statistical packages, and you can easily copy and paste graphs and output into documents.

JMP also provides an interface to explore data visually and interactively, which will help your students develop a healthy relationship with their data, work more efficiently with data, and tackle difficult statistical problems more easily. Because its output provides both statistics and graphs together, the student will better see and understand the application of concepts covered in this book as well. JMP Student Edition also contains some unique platforms for student projects, such as mapping and scripting. JMP functions in the same way on both Windows and Mac platforms and instructions contained with this book apply to both platforms.

Access to this software is available for free with new copies of the book and available for purchase standalone at Cengage.com or www.jmp.com/getse. Find out more at www.jmp.com.

Web Apps

A collection of easy to use web apps is available at statistics.cengage.com/PSO6e/Apps.html. This collection includes apps that support new sections on bootstrap confidence intervals and randomization tests, as well as apps that help students visualize the meaning of confidence level and to understand the concept of sampling variability.

Student Resources

Digital

To access additional course materials and companion resources, please visit www.cengage.com. At the Cengage.com home page, search for the ISBN of your title (from the back cover of your book) using the search box at the top of the page. This will take you to the product page where free companion resources can be found.

- Complete step-by-step instructions for JMP, TI-84 Graphing Calculators, Excel, Minitab, and SPSS.
- Data sets in Excel, ASCII-comma, ASCII-tab, JMP, Minitab, R, SAS, and SPSS file formats indicated by the ● icon throughout the text.
- Applets used in the Activities found in the text.

Print

Student Solutions Manual *(ISBN: 978-1-337-79417-6):* The Student Solutions Manual, prepared by Stephen Miller, contains fully worked-out solutions to all of the odd-numbered exercises in the text, giving students a way to check their answers and ensure that they took the correct steps to arrive at an answer.

Instructor Resources

Print

Annotated Instructor's Edition *(ISBN: 978-1-337-79415-2):* The Annotated Instructor's Edition contains answers for all exercises, including those not found in the answer section of the student edition. There also are suggested assignments and teaching tips for each section in the book, along with an annotated table of contents with comments written by Roxy Peck.

AP® Teacher's Resource Binder: The Teacher's Resource Binder, prepared by Chris Olsen, is full of wonderful resources for both college professors and AP® Statistics teachers. These include

- Additional examples from published sources (with references), classified by chapter in the text. These examples can be used to enrich your classroom discussions.
- Model responses—examples of responses that can serve as a model for work that would be likely to receive a high mark on the AP® exam.
- A collection of data explorations written by Chris Olsen that can be used throughout the year to help students prepare for the types of questions that they may encounter on the investigative task on the AP® Statistics Exam.
- Advice to AP® Statistics teachers on preparing students for the AP® Exam, written by Brian Kotz.
- Activity worksheets, prepared by Carol Marchetti, that can be duplicated and used in class.
- A list of additional resources for activities, videos, and computer demonstrations, cross-referenced by chapter.
- A test bank that includes assessment items, quizzes, and chapter exams written by Chris Olsen, Josh Tabor, and Peter Flanagan-Hyde.

Online

- **Instructor Companion Site:** Everything you need for your course in one place! This collection of book-specific lecture and class tools is available online via www.cengage.com/login. Access and download PowerPoint presentations, images, instructor's manual, and more.
- **Cengage Learning Testing Powered by Cognero (*ISBN: 978-1-337-79423-7*)** is a flexible, online system that allows you to author, edit, and manage test bank content, create multiple test versions in an instant, and deliver tests from your LMS, your classroom or wherever you want. This is available online via www.cengage.com/login.
- **Complete Solutions Manual** This manual contains solutions to all exercises from the text, including Chapter Review Exercises and Cumulative Review Exercises. This manual can be found on the Instructor Companion Site.

Acknowledgments

We are grateful for the thoughtful feedback from the following reviewers that has helped to shape this text over the last three editions:

Reviewers for the Sixth Edition

Weizhong Tian, Eastern New Mexico University
Greg Perkins, Hartnell College
Bambi Jones, Lake Land College
Paul Holmes, University of Georgia
Christina Cornejo, Erie Community College

Carl Brezovec, Franklin Pierce University
David Manley, Rowan University
John Racquet, University at Albany, Excelsior College
Charles Conrad, Volunteer State Community College
Zhongming Huang, Midland University
Chad Bemis, Pierce College Fort Steilacoom

Reviewers for the Fifth, Fourth, Third, and Second Editions

Arun K. Agarwal, Jacob Amidon, Holly Ashton, Barb Barnet, Eddie Bevilacqua, Piotr Bialas, Kelly Black, Jim Bohan, Pat Buchanan, Gabriel Chandler, Andy Chang, Jerry Chen, Richard Chilcoat, Mary Christman, Marvin Creech, Kathleen Dale, Ron Degged, Hemangini Deshmukh, Ann Evans, Guangxiong Fang, Sharon B. Finger, Donna Flint, Steven Garren, Mark Glickman, Rick Gumina, Debra Hall, Tyler Haynes, Sonja Hensler, Trish Hutchinson, John Imbrie, Bessie Kirkwood, Jeff Kollath, Christopher Lacke, Austin Lampros, Michael Leitner, Zia Mahmood, Art Mark, Pam Martin, David Mathiason, Bob Mattson, Kendra Mhoon, C. Mark Miller, Megan Mocko, Paul Myers, Kane Nashimoto, Helen Noble, Douglas Noe, Broderick Oluyede, Elaine Paris, Shelly Ray Parsons, Deanna Payton, Judy Pennington-Price, Michael Phelan, Alan Polansky, Mamunur Rashid, Leah Rathbun, Michael Ratliff, David Rauth, Kevin J. Reeves, Lawrence D. Ries, Hazel Shedd, Robb Sinn, Greg Sliwa, Angela Stabley, Jeffery D. Sykes, Yolanda Tra, Joe Ward, Nathan Wetzel, Mark Wilson, Yong Yu, and Toshiyuki Yuasa, Cathleen Zucco-Teveloff.

We would also like to express our thanks and gratitude to those whose support made this sixth edition possible:

- Cassie Van Der Laan, Product Manager
- Brendan Killion and Abby DeVeuve, Content Managers
- Elinor Gregory, Learning Designer
- Lori Hazzard, our senior project manager at MPS Limited
- Stephen Miller for his work in creating new student and instructor solutions manuals to accompany the text
- A. Palanisamy, for checking the accuracy of examples and solutions
- Carolyn Crockett and Molly Taylor, our former editors at Cengage, for their support on the previous editions of this book

And, as always, we thank our families, friends, and colleagues for their continued support.

Roxy Peck
Tom Short
Chris Olsen

Introduction to Statistics and Data Analysis

1 The Role of Statistics and the Data Analysis Process

ESB Professional/Shutterstock.com

Statistics is the scientific discipline that provides methods that help us make sense of data. Statistical methods offer a set of powerful tools for gaining insight into the world around us. The use of statistical analyses in fields such as business, medicine, agriculture, social science, natural science, and engineering has led to increased recognition that statistical literacy should be part of a well-rounded education.

The field of statistics helps us to make intelligent judgments and informed decisions in the presence of uncertainty and variability. In this chapter, we consider the role of variability in statistical settings, introduce some basic terminology, and look at some simple graphical displays for summarizing data.

LEARNING OBJECTIVES

Students will understand:

- The steps in the data analysis process.

Students will be able to:

- Distinguish between a population and a sample.
- Distinguish between categorical, discrete numerical, and continuous numerical data.
- Construct a frequency distribution and a bar chart, and describe the distribution of a categorical variable.
- Construct a dotplot and describe the distribution of a numerical variable.

SECTION 1.1 Why Study Statistics?

There is an old saying that "without data, you are just another person with an opinion." While anecdotes and coincidences may make for interesting stories, you wouldn't want to make important decisions on the basis of anecdotes alone. For example, just because a friend of a friend ate 16 apricots and then experienced relief from joint pain doesn't mean that this is all you need to know to help one of your parents choose a treatment for arthritis! Before recommending apricots, you would definitely want to consider relevant data—that is, data that would allow you to investigate the effectiveness of apricots as a treatment for arthritis.

It is challenging to function in today's world without a basic understanding of statistics. For example, here are just a few headlines from articles that draw conclusions based on data that appeared in a single newspaper on one day.

How many people does it take to build an airplane? The article **"Boeing Delivers Records" (*The Wall Street Journal*, January 10, 2018)** looked at the ways in which Boeing has been able to increase airplane production and introduce new airplane models. The article states that "Boeing has boosted output by two-thirds over the past seven years but cut the average number of employees needed to build each plane." This statement was supported by a graph that showed the number of employees per jet airplane produced over time. In 2017, this number was at its lowest, with 94 employees per jet produced.

"New Venture Fund Targets Autos" (*The Wall Street Journal*, January 10, 2018) is the title of an article that looked at funding for innovation in the auto industry. Venture capital is money invested in start-up companies exploring new technologies. The article included a graph of data that shows how the amount of venture capital funding for the automotive industry has been increasing over time, noting that this funding nearly tripled in 2017 compared with 2016. The hope is that this will increase the pace of introduction of new technologies in new cars.

The article **"Companies Take to the Sky in Race to Deliver on Time" (*The Wall Street Journal*, January 10, 2018)** notes that growth in online shopping and a healthy economy is creating an increased demand for shipping by air. Two graphs of data are included with the article. One shows the yearly change in air cargo volume for the years 2013 to 2017, noting increases in each of these years and a particularly large increase in 2017. The second graph shows how the cost of air cargo shipping has increased since 2015. Based on these trends, Amazon has started its own airline to handle the shipping of Amazon orders and is converting older passenger jets for cargo use.

The article **"Saudis Target Religious Extremism" (*The Wall Street Journal*, January 10, 2018)** reported on public reaction to Saudi Arabia's decision to lift a ban that prohibits women from driving. The article includes graphs summarizing data collected in a survey of 500 Saudis. The people surveyed were asked if they were pleased with the decision to lift the ban on women driving. The graphs showed that 74% of the women surveyed and 55% of the men surveyed said that they were pleased with the decision. The graphs also showed that the percentage of men who were not sure if they were pleased or if they were not pleased was greater than this percentage for women (17% of the men and 11% of the women). These data enable the Saudi government to assess support for social change.

As people approach retirement age, many are finding that they are not prepared financially. The article **"37% of Gen X Can't Afford to Retire, Poll Finds" (*The Wall Street Journal*, January 10, 2018)** summarized the results of a survey of 828 people born between 1965 and the late 1970s (known as "Generation X") and 990 people born between 1945 and 1964 (the "Baby Boomers"). They found that while 47% of Baby Boomers expect that they will be very secure in retirement, only 33% of Gen Xers expect that they will be very secure. They also found that 37% of Gen Xers would like to stop working someday, but fear that they will not be able to afford to, and that 49% are worried about running out of money once they leave the workforce. These findings have implications for those who provide social services to older Americans.

To be an informed consumer of reports such as those described above, you must be able to do the following:

1. Extract information from tables, charts, and graphs.
2. Follow numerical arguments.
3. Understand the basics of how data should be gathered, summarized, and analyzed in order to draw statistical conclusions.

Your statistics course will help prepare you to perform these tasks.

Studying statistics will also enable you to collect data in a sensible way and then use the data to answer questions of interest. In addition, studying statistics will allow you to critically evaluate the work of others by providing you with the tools you need to make informed judgments.

Throughout your personal and professional life, you will need to understand and use data to make decisions. To do this, you must be able to

1. Decide whether existing data are adequate or whether additional information is required.
2. If necessary, collect more information in a reasonable and thoughtful way.
3. Summarize the available data in a useful and informative manner.
4. Analyze the available data.
5. Draw conclusions, make decisions, and assess the risk of an incorrect decision.

These are the steps in the data analysis process. These steps are considered in more detail in Section 1.3.

We hope that this textbook will help you to understand the logic behind statistical reasoning, prepare you to apply statistical methods appropriately, and enable you to recognize when statistical arguments are faulty.

SECTION 1.2 The Nature and Role of Variability

Statistical methods allow us to collect, describe, analyze, and draw conclusions from data. If we lived in a world where all measurements were identical for every individual, these tasks would be simple. Imagine a population consisting of all students at a particular university. Suppose that *every* student is enrolled in the same number of courses, spent exactly the same amount of money on textbooks this semester, and favors increasing student fees to support expanding library services. For this population, there is *no* variability in number of courses, amount spent on books, or student opinion on the fee increase. A researcher studying students from this population in order to draw conclusions about these three variables would have a particularly easy task. It would not matter how many students the researcher studied or how the students were selected. In fact, the researcher could collect information on number of courses, amount spent on books, and opinion on the fee increase by just stopping the next student who happened to walk by. Because there is no variability in the population, this one individual would provide complete and accurate information about the population. The researcher could draw conclusions with no risk of error.

The situation just described is obviously unrealistic. Populations with no variability are rare. We need to understand variability to be able to collect, describe, analyze, and draw conclusions from data in a sensible way.

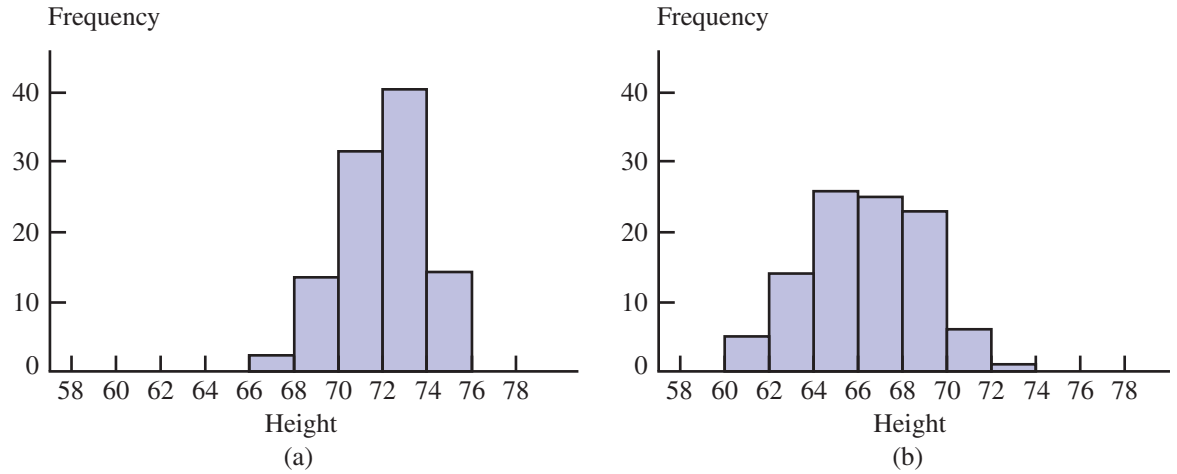
Examples 1.1 and 1.2 illustrate how describing and understanding variability are important.

Example 1.1 If the Shoe Fits

Understand the context ➤ The graphs in Figure 1.1 are examples of a type of graph called a histogram. (The construction and interpretation of histograms is discussed in Chapter 3.) Figure 1.1(a) shows the distribution of the heights of female basketball players who played at a particular university between 2005 and 2013. Each bar in the histogram represents a particular range of player heights. The height of each bar in the graph indicates how

Consider the data ➤ many players' heights were in the corresponding range. For example, 40 basketball players had heights between 72 inches and 74 inches, while only 2 players had heights between 66 inches and 68 inches. Figure 1.1(b) shows the distribution of heights for members of the women's gymnastics team. Both histograms are based on the heights of 100 women.

FIGURE 1.1
Histograms of heights (in inches) of female athletes: (a) 100 basketball players; (b) 100 gymnasts.



The first histogram shows that the heights of female basketball players varied, with most heights falling between 68 inches and 76 inches. In the second histogram we see that the heights of female gymnasts also varied, with most heights in the range of 60 inches to 72 inches. It is also clear that there is more variation in the heights of the gymnasts than in the heights of the basketball players, because the gymnast histogram spreads out more about its center than does the basketball histogram.

Interpret the results ➤ Now suppose that a tall woman (5 feet 11 inches) tells us she is looking for her sister who is practicing with her team at the gym. Should we direct her to where the basketball team is practicing or to where the gymnastics team is practicing? What reasoning could we use to decide? If we found a pair of size 6 shoes left in the locker room, should we first try to return them by checking with members of the basketball team or the gymnastics team?

We would probably send the woman looking for her sister to the basketball practice and we would probably try to return the shoes to a gymnastics team member. Reaching these conclusions requires statistical reasoning that combines knowledge of the relationship between heights of siblings and between shoe size and height with the information about the distributions of heights presented in Figure 1.1. We might have reasoned that heights of siblings tend to be similar and that a height as great as 5 feet 11 inches, although not impossible, would be unusual for a gymnast. On the other hand, a height as tall as 5 feet 11 inches would be common for a basketball player.

Similarly, we might have reasoned that tall people tend to have bigger feet and that short people tend to have smaller feet. The shoes found were a small size, so it is more likely that they belong to a gymnast than to a basketball player, because small heights are common for gymnasts and unusual for basketball players.

Example 1.2 Monitoring Water Quality

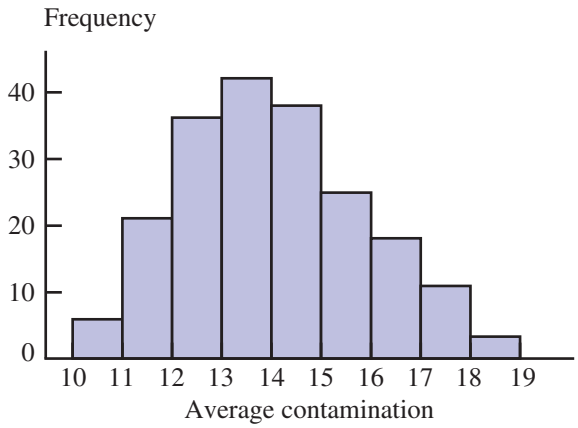

David Chasey/Photodisc/Getty Images

Understand the context ➤ As part of its regular water quality monitoring efforts, an environmental control board selects five containers of water from a particular well each day. The concentration of contaminants in parts per million (ppm) is measured for each of the five containers, and then the average of the five measurements is calculated. The histogram in Figure 1.2 summarizes the average contamination values for 200 days.

Now suppose that a chemical spill has occurred at a manufacturing plant 1 mile from the well. It is not known whether a spill of this nature would contaminate groundwater in the area of the spill and, if so, whether a spill this distance from the well would affect the quality of well water.

Consider the data ➤ One month after the spill, five containers of water are collected from the well, and the average contamination is 15.5 ppm. Considering the variability before the spill, should we interpret this as evidence that the well water was affected by the spill? What if the calculated average was 17.4 ppm? 22.0 ppm? How is the reasoning related to the histogram in Figure 1.2?

FIGURE 1.2
Average contamination concentration (in ppm) measured each day for 200 days.



Interpret the results ➤ Before the spill, the average contaminant concentration varied from day to day. An average of 15.5 ppm would not have been an unusual value, so seeing an average of 15.5 ppm after the spill isn't necessarily an indication that contamination has increased. On the other hand, an average as large as 17.4 ppm is less common, and an average as large as 22.0 ppm is not at all typical of the pre-spill values. In this case, we would probably conclude that the well contamination level has increased.

In these two examples, reaching a conclusion required an understanding of variability. Understanding variability allows us to distinguish between common and unusual values. The ability to recognize unusual values in the presence of variability is an important aspect of most statistical procedures. It also enables us to quantify the chance of being incorrect when a conclusion is based on data. These concepts will be developed further in later chapters.

SECTION 1.3 Statistics and the Data Analysis Process

Statistical studies are undertaken to answer questions about our world. Is a new flu vaccine effective in preventing illness? Is the use of bicycle helmets on the rise? Are injuries that result from bicycle accidents less severe for riders who wear helmets than for those who do not? Data collection and analysis allow researchers to answer questions like these.

The data analysis process can be viewed as a sequence of steps that lead from planning to data collection to making informed conclusions based on the resulting data. The process can be organized into the six steps described below.

The Data Analysis Process

1. **Understanding the nature of the problem.** Effective data analysis requires an understanding of the research problem. We must know the goal of the research and what questions we hope to answer. It is important to have a clear direction before gathering data to ensure that we will be able to answer the questions of interest using the data collected.
2. **Deciding what to measure and how to measure it.** The next step in the process is deciding what information is needed to answer the questions of interest. In some cases, the choice is obvious. For example, in a study of the relationship between the weight of a Division I football player and position played, we would need to collect data on player weight and position. In other cases the choice of information is not as straightforward. For example, in a study of the relationship between preferred learning style and intelligence, how should we define learning style and measure it? What measure of intelligence should we use? It is important to carefully define the variables to be studied and to develop appropriate methods for determining their values.
3. **Data collection.** The data collection step is very important. The researcher must first decide whether an existing data source is adequate or whether new data must be collected. If a decision is made to use existing data, it is important to understand how the data were collected and for what purpose, so that any resulting limitations are also fully understood. If new data are to be collected, a careful plan must be developed, because the type of analysis that is appropriate and the conclusions that can be drawn depend on how the data are collected.
4. **Data summarization and preliminary analysis.** After the data are collected, the next step is usually a preliminary analysis that includes summarizing the data graphically and numerically. This initial analysis provides insight into important characteristics of the data and provides guidance in selecting appropriate methods for further analysis.
5. **Formal data analysis.** The data analysis step requires the researcher to select appropriate statistical methods. Much of this textbook is devoted to methods that can be used to carry out this step.
6. **Interpretation of results.** Several questions should be addressed in this final step. Some examples are: What can we learn from the data? What conclusions can be drawn from the analysis? How can our results guide future research? The interpretation step often leads to the formulation of new research questions. These new questions lead back to the first step. In this way, good data analysis is often an iterative process.

To illustrate these steps, consider the following example. The admissions director at a large university might be interested in learning why some applicants who were accepted for the fall 2019 term failed to enroll at the university. The **population** of interest to the director consists of all accepted applicants who did not enroll in the fall 2019 term. Because this population is large and it may be difficult to contact all the individuals, the director might decide to collect data from only 300 selected students. These 300 students constitute a **sample**.

DEFINITIONS

Population: The entire collection of individuals or objects about which information is desired is called the **population** of interest.

Sample: A **sample** is a subset of the population, selected for study.

Deciding how to select the 300 students and what data should be collected from each student are steps 2 and 3 in the data analysis process. Step 4 in the process involves

organizing and summarizing data. Methods for organizing and summarizing data, such as the use of tables, graphs, and numerical summaries, make up the branch of statistics called **descriptive statistics**. A second major branch of statistics, **inferential statistics**, involves generalizing from a sample to the population from which it was selected. When we generalize in this way, we run the risk of an incorrect conclusion, because a conclusion about the population is based on incomplete information. An important aspect in the development of inferential techniques involves quantifying the chance of an incorrect conclusion.

DEFINITIONS

Descriptive statistics: The branch of statistics that includes methods for organizing and summarizing data.

Inferential statistics: The branch of statistics that involves generalizing from a sample to the population from which the sample was selected and assessing the reliability of such generalizations.

Example 1.3 illustrates the steps in the data analysis process.

Example 1.3 Chew More, Eat Less?

Understand the context ➤ The article **"Increasing the Number of Chews before Swallowing Reduces Meal Size in Normal-Weight, Overweight, and Obese Adults" (*Journal of the Academy of Nutrition and Dietetics* [2014]: 926–931)** describes a study that investigated whether chewing each bite of food more before swallowing would result in people eating less. Participants in the study were adults between the ages of 18 and 45 years. At the beginning of the study, each participant was observed as they each ate five pizza rolls, and the number of chews made before swallowing was observed in order to determine a baseline for that participant.

Participants were then invited back for a second session on a different day. They were asked to eat their usual breakfast on that day and to not eat anything after breakfast. At the second session, all participants were provided with a platter of pizza rolls and were told to eat until they were comfortably full. They were also told they could request more pizza rolls if they wanted more. Each participant was also told how many times to chew each pizza roll before swallowing. Then, each participant was assigned to one of three groups. The participants in group 1 were given a number of chews equal to their baseline. The participants in group 2 were given a number of chews that was 150% of (one and a half times as large as) their baseline. The participants in group 3 were assigned a number of chews that was 200% of (twice as large as) their baseline.

Interpret the results ➤ After analyzing data from this study, the researcher concluded that people ate about 10% less when they increased the number of chews by 50% (group 2) and about 15% less when they doubled the number of chews.

This study illustrates the nature of the data analysis process. A clearly defined research question and an appropriate choice of how to measure the variables of interest (the number of chews and how much people ate) preceded the data collection. Assuming that a reasonable method was used to collect the data (we will see how this can be evaluated in Chapter 2) and that appropriate methods of analysis were employed, the investigators reached the conclusion that increasing the number of chews before swallowing results in people tending to eating less.

EXERCISES 1.1 - 1.11

● **Data set available online**

1.1 Give brief definitions of the terms *descriptive statistics* and *inferential statistics.*

1.2 Give brief definitions of the terms *population* and *sample.*

1.3 The following conclusion from a study appeared in the article **"Smartphone Nation" (*AARP Bulletin*, September 2009)**: "If you love your smart phone, you are not alone. Half of all boomers sleep with their cell phone within arm's length. Two of three

people age 50 to 64 use a cell phone to take photos, according to a 2010 Pew Research Center report." Are the given proportions (half and two of three) population values, or were they calculated from a sample?

1.4 Based on a study of 2121 children between the ages of 1 and 4, researchers at the Medical College of Wisconsin concluded that there was an association between iron deficiency and the length of time that a child is bottle-fed (*Milwaukee Journal Sentinel*, November 26, 2005). Describe the sample and the population of interest for this study.

1.5 The student senate at a university with 15,000 students is interested in the proportion of students who favor a change in the grading system to allow for plus and minus grades (for example, B+, B, B−, rather than just B). Two hundred students are interviewed to determine their attitude toward this proposed change.

a. What is the population of interest?

b. What group of students constitutes the sample in this problem?

1.6 The National Retail Federation used data from a survey of 7439 adult Americans to estimate the percent who planned to spend more on holiday shopping in 2017 than they spent in 2016. They estimated that while 24% of adult Americans planned to spend more, for those age 16 to 24, the percentage was 46% ("Almost Half of Younger Consumers Plan to Spend More During the Holidays," nrf.com/media/press-releases/almost-half-of-younger-consumers-plan-spend-more-during-the-holidays, retrieved February 5, 2018). Are the estimates given calculated using data from a sample or for the entire population?

1.7 The supervisors of a rural county are interested in the proportion of property owners who support the construction of a sewer system. Because it is too costly to contact all 7000 property owners, a survey of 500 owners is undertaken. Describe the population and sample for this problem.

1.8 A consumer group conducts crash tests of new model cars. To determine the severity of damage to 2019 Toyota Camrys resulting from a 10-mph crash into a concrete wall, the research group tests six cars of this type and assesses the amount of damage. Describe the population and the sample for this problem.

1.9 A building contractor has a chance to buy an odd lot of 5000 used bricks at an auction. She is interested in determining the proportion of bricks in the lot that are cracked and therefore unusable for her current project, but she does not have enough time to inspect all 5000 bricks. Instead, she checks 100 bricks to determine which ones are cracked. Describe the population and the sample for this problem.

1.10 The article "Brain Shunt Tested to Treat Alzheimer's" (*San Francisco Chronicle*, October 23, 2002) summarizes the findings of a study that appeared in the journal *Neurology*. Doctors at Stanford Medical Center were interested in determining whether a new surgical approach to treating Alzheimer's disease results in improved memory functioning. The surgical procedure involves implanting a thin tube, called a shunt, which is designed to drain toxins from the fluid-filled space that cushions the brain. Eleven patients had shunts implanted and were followed for a year, receiving quarterly tests of memory function. Another sample of Alzheimer's patients was used as a comparison group. Those in the comparison group received the standard care for Alzheimer's disease. After analyzing the data from this study, the investigators concluded that the "results suggested the treated patients essentially held their own in the cognitive tests while the patients in the control group steadily declined. However, the study was too small to produce conclusive statistical evidence."

a. What were the researchers trying to learn? What questions motivated their research?

b. Do you think that the study was conducted in a reasonable way? What additional information would you want in order to evaluate this study?

1.11 In a study of whether taking a garlic supplement reduces the risk of getting a cold, participants were assigned to either a garlic supplement group or to a group that did not take a garlic supplement ("Garlic for the Common Cold," *Cochrane Database of Systematic Reviews*, 2009). Based on the study, it was concluded that the proportion of people taking a garlic supplement who get a cold is lower than the proportion of those not taking a garlic supplement who get a cold.

a. What were the researchers trying to learn? What questions motivated their research?

b. Do you think that the study was conducted in a reasonable way? What additional information would you want in order to evaluate this study?

SECTION 1.4 Types of Data and Some Simple Graphical Displays

Every discipline has its own particular way of using common words, and statistics is no exception. You will recognize some of the terminology from previous math and science courses, but much of the language of statistics will be new to you. In this section, you will learn some of the terminology used to describe data.

Types of Data

The individuals or objects in any particular population might possess many characteristics that could be studied. Consider a group of students currently enrolled in a statistics class. One characteristic of the students in the population is the brand of calculator they use (Casio, Hewlett-Packard, Sharp, Texas Instruments, and so on). Another characteristic is the number of textbooks purchased that semester, and yet another is the distance from the college to each student's home. A **variable** is any characteristic whose value may change from one individual or object to another. For example, *calculator brand* is a variable, and so are *number of textbooks purchased* and *distance to the college*. **Data** result from making observations either on a single variable or on two or more variables at the same time.

DEFINITIONS

Variable: A characteristic whose value may change from one observation to another.

Data: A collection of observations on one or more variables.

A **univariate data set** consists of observations on a single variable made on individuals in a sample or population. There are two types of univariate data sets: **categorical** (sometimes also called qualitative) and **numerical** (sometimes also called quantitative). In the previous example, *calculator brand* is a categorical variable, because each student's response to the query, "What brand of calculator do you use?" is a category. The collection of responses from all these students forms a categorical data set. The other two variables, *number of textbooks purchased* and *distance to the college*, are both numerical in nature. Determining the values of a numerical variable (by counting or measuring) results in a numerical data set.

DEFINITIONS

Univariate data set: A data set consisting of observations on a single characteristic.

Categorical data set: A univariate data set is **categorical** (or **qualitative**) if the individual observations are categorical responses.

Numerical data set: A univariate data set is **numerical** (or **quantitative**) if each observation is a number.

Example 1.4 College Choice Do-Over?

Understand the context ➤ The Higher Education Research Institute at UCLA surveys over 20,000 college seniors each year. One question on the survey asks seniors the following question: If you could make your college choice over, would you still choose to enroll at your current college? Possible responses are definitely yes (DY), probably yes (PY), probably no (PN), and definitely no (DN). Responses for 20 students were:

DY PN DN DY PY PY PN PY PY DY
DY PY DY DY PY PY DY DY PN DY

Consider the data ➤ (These data are just a small subset of the data from the survey.) Because the response to the question about college choice is categorical, this is a univariate categorical data set.