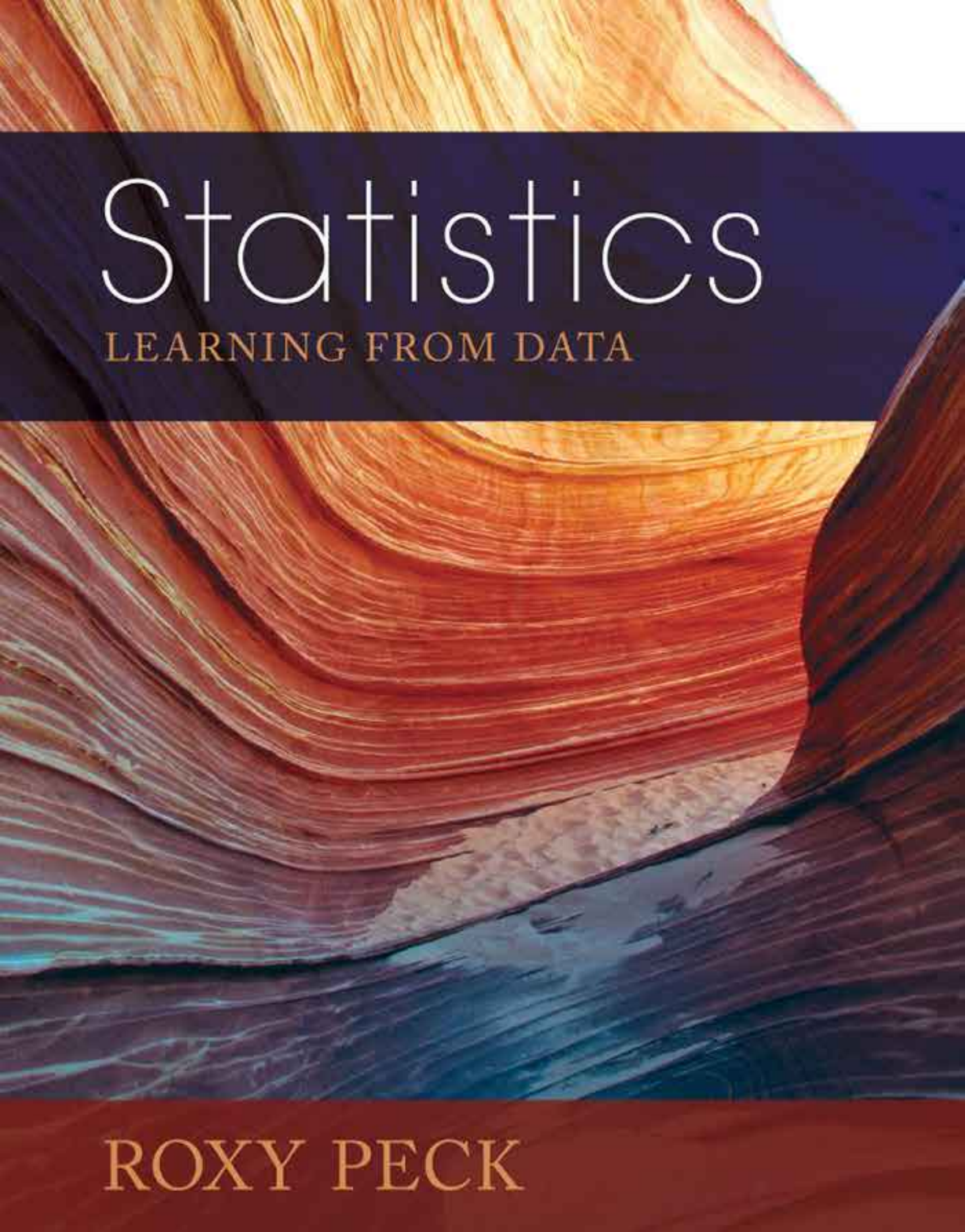

Statistics
LEARNING FROM DATA
ROXY PECK

STATISTICS: LEARNING FROM DATA

Roxy Peck
California Polytechnic State University, San Luis Obispo

CENGAGE Learning®

Australia • Brazil • Japan • Korea • Mexico • Singapore • Spain • United Kingdom • United States

Statistics: Learning from Data
Roxy Peck

Senior Product Team Manager: Richard Stratton

Senior Product Manager: Molly Taylor

Senior Content Developer: Jay Campbell

Product Assistant: Danielle Hallock

Media Developer: Andrew Coppola

Senior Marketing Manager: Gordon Lee

Marketing Development Manager: Ryan Ahern

Content Project Manager: Alison Eigel Zade

Senior Art Director: Linda May

Manufacturing Planner: Sandee Milewski

Rights Acquisition Specialist: Shalice Shah-Caldwell

Production Service: MPS Limited

Interior Designer: Diane Beasley

Cover Designer: Rokusek Design

Cover Image: Shutterstock

Compositor: MPS Limited

Library of Congress Control Number: 2013947061

Student Edition:

ISBN-13: 978-0-495-55326-7

ISBN-10: 0-495-55326-3

Cengage Learning
200 First Stamford Place, 4th Floor
Stamford, CT 06902
USA

Cengage Learning is a leading provider of customized learning solutions with office locations around the globe, including Singapore, the United Kingdom, Australia, Mexico, Brazil and Japan. Locate your local office at **international.cengage.com/region**

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

For your course and learning solutions, visit **www.cengage.com.**

Purchase any of our products at your local college store or at our preferred online store **www.cengagebrain.com.**

Instructors: Please visit **login.cengage.com** and log in to access instructor-specific resources.

Printed in the United States of America
1 2 3 4 5 6 7 17 16 15 14 13

Dedication

To my friends and colleagues in the Cal Poly Statistics Department

Author Bio

ROXY PECK is a professor emerita of statistics at California Polytechnic State University, San Luis Obispo. She was a faculty member in the Statistics Department for thirty years, serving for six years as Chair of the Statistics Department and thirteen years as Associate Dean of the College of Science and Mathematics. Nationally known in the area of statistics education, Roxy was made a Fellow of the American Statistical Association in 1998, and in 2003 she received the American Statistical Association's Founders Award in recognition of her contributions to K-12 and undergraduate statistics education. In 2009, she received the USCOTS Lifetime Achievement Award in Statistics Education. In addition to coauthoring the textbooks *Statistics: Learning from Data, Introduction to Statistics and Data Analysis*, and *Statistics: The Exploration and Analysis of Data*, she is also editor of *Statistics: A Guide to the Unknown*, a collection of expository papers that showcases applications of statistical methods. Roxy served from 1999 to 2003 as the Chief Faculty Consultant for the Advanced Placement Statistics exam and she is a past chair of the joint ASA/NCTM Committee on Curriculum in Statistics and Probability for Grades K-12 and of the ASA Section on Statistics Education. Outside the classroom, Roxy enjoys travel and has visited all seven continents. She collects Navajo rugs and heads to Arizona and New Mexico whenever she can find the time.

Brief Contents

Contents

Preface

Statistics is about learning from data and the role that variability plays in drawing conclusions from data. To be successful, it is not enough for students to master the computational aspects of descriptive and inferential statistics—they must also develop an understanding of the data analysis process at a conceptual level. *Statistics: Learning from Data* is informed by careful and intentional thought about how the conceptual and the mechanical should be integrated in order to promote three key types of learning objectives for students:

- conceptual understanding
- mastery of the mechanics
- the ability to demonstrate these by "putting it into practice"

A Unique Approach

A number of innovative features distinguish this text from other introductory statistics books:

- **A New Approach to Probability**
  There is now quite a bit of research on how students develop an understanding of probability and chance. Using *natural frequencies* to reason about probability, especially conditional probability, is much easier for students to understand. The treatment of probability in this text is complete, including conditional probability and Bayes' Rule type probability calculations, but is done in a way that eliminates the need for the symbolism and formulas that are a roadblock for so many students.
- **Chapter on Overview of Statistical Inference (Chapter 7)**
  This short chapter focuses on the things students need to think about in order to select an appropriate method of analysis. In most texts, these considerations are "hidden" in the discussion that occurs when a new method is introduced. Discussing these considerations up front in the form of four key questions that need to be answered before choosing an inference method makes it easier for students to make correct choices.
- **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. The ordering of topics in the text book mirrors this process: data collection, then data description, then statistical inference.
- **Inference for Proportions Before Inference for Means**
  Inference for proportions is covered before inference for means for the following reasons:
  - This makes it possible to develop the concept of a sampling distribution via simulation, an approach that is more accessible to students than a more formal, theoretical approach. Simulation is simpler in the context of proportions, where it is easy to construct a hypothetical population from which to sample (it is more complicated to create a hypothetical population in the context of means because this requires making assumptions about shape and spread).
  - Large-sample inferential procedures for proportions are based on the normal distribution and don't require the introduction of a new distribution (the t distribution). Students can focus on the new concepts of estimation and hypothesis testing without having to grapple at the same time with the introduction of a new probability distribution.

- **Separate Treatment of Inference Based on Experiment Data (Chapter 14)**
  Many statistical studies involve collecting data via experimentation. The same inference procedures used to estimate or test hypotheses about population parameters also are used to estimate or test hypotheses about treatment effects. However, the necessary assumptions are slightly different (for example, random assignment replaces the assumption of random selection), and the wording of hypotheses and conclusions is also different. Trying to treat both cases together tends to confuse students. This text makes the distinction clear.

Features That Support Student Engagement and Success

The text also includes a number of features that support conceptual understanding, mastery of mechanics, and putting ideas into practice.

- **Simple Design**
  There is now research showing that many of the "features" in current textbooks are not really helpful to students. In fact, cartoons, sidebars, historical notes, fake Post-it notes in the margins, and the like, actually distract students and interfere with learning. *Statistics: Learning from Data* has a simple, clean design in order to minimize clutter and maximize student understanding.
- **Chapter Learning Objectives—Keeping Students Informed About Expectations**
  Chapter learning objectives explicitly state the expected student outcomes. Learning objectives fall under three headings: Conceptual Understanding, Mastery of Mechanics, and Putting It Into Practice.
- **Preview—Motivation for Learning**
  Each chapter opens with a *Preview* and *Preview Example* that provide motivation for studying the concepts and methods introduced in the chapter. They address why the material is worth learning, provide the conceptual foundation for the methods covered in the chapter, and connect to what the student already knows. A relevant and current example provides a context in which one or more questions are proposed for further investigation. This context is revisited in the chapter once students have the necessary understanding to more fully address the questions posed.
- **Real Data**
  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. The exercises and examples are a particular strength of this text, and I 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. Examples and exercises that involve data extracted from journal articles, newspapers, and other published sources and that are of interest to today's students are used to motivate and engage students. 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, environmental research, law and criminal justice, and entertainment.
- **Exercises Organized Into a Developmental Structure—Structuring the Out-of-Class Experience**
  End-of-section exercises are organized into developmental sets. At the end of each section, there are two grouped problem sets. The exercises in each set work together to assess all of the learning objectives for that section. In addition to the two exercise sets, each section also has additional exercises for those who want more practice.

  Answers for the exercises of Exercise Set 1 in each section are included at the end of the book. In addition, many of the exercises in Exercise Set 1 include hints directing the student to a particular example or a relevant discussion that appears in the text. This feature provides direction for students who might need help getting started on a particular exercise. Instructors who prefer that students be more self-directed can

assign Exercise Set 2. Answers and hints are not provided for the exercises in Exercise Set 2.

- **Are You Ready to Move On?—Students Test Their Understanding**
  Prior to moving to the next chapter, "Are You Ready to Move On?" exercises allow students to confirm that they have achieved the chapter learning objectives. Like the developmental problem sets of the individual sections, this collection of exercises is developmental in nature. These exercises assess all of the chapter learning objectives and serve as a comprehensive end-of-chapter review.
- **Exploring the Big Ideas—Real Data Algorithmic Sampling Exercises**
  Most chapters contain extended sampling-based, real-data exercises at the end of the chapter. These exercises appear in CourseMate and Aplia, where each student gets a different random sample for the same exercise. These unique exercises
  - address the tension between the desire to use real data and the desire to have algorithmically generated exercises.
  - address the tension between the role of interpretation and communication in data analysis and the desire for exercises that can be machine scored.
  - are designed to teach about sampling variability.
- **Data Analysis Software**
  Each new textbook comes with free JMP data analysis software. See *Student Resources* for more information.
- **Technology Notes**
  Technology Notes appear at the end of most chapters and give students helpful hints and guidance on completing tasks associated with a particular chapter. The following technologies are included in the notes: JMP, Minitab, SPSS, Microsoft Excel 2007, TI-83/84, and TI-nspire. They include display screens to help students visualize and better understand the steps. More complete technology manuals are also available on the text web site.
- **Chapter Activities—Engaging Students in Hands-On Activities**
  There is a growing body of evidence that students learn best when they are actively engaged. Chapter activities guide students' thinking about important ideas and concepts.

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

In 2005, the American Statistical Association endorsed the report "College Guidelines in Assessment and Instruction for Statistics Education (GAISE Guidelines)," which included the following six recommendations for the introductory statistics course:

1. Emphasize statistical literacy and develop statistical thinking.
2. Use real data.
3. Stress conceptual understanding rather than mere knowledge of procedures.
4. Foster active learning in the classroom.
5. Use technology for developing conceptual understanding and analyzing data.
6. Use assessments to improve and evaluate student learning.

Statistic: Learning from Data is consistent with these recommendations and supports the GAISE guidelines in the following ways:

1. **Emphasize Statistical Literacy and Develop Statistical Thinking.**
   Statistical literacy is 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. **Use Real Data.**
   The examples and exercises are context driven, and the reference sources include the popular press as well as journal articles.
3. **Stress Conceptual Understanding Rather Than Mere Knowledge of Procedures.**
   Nearly all exercises in the text are multipart and ask students to go beyond just calculation, with a focus on interpretation and communication. 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 . . .
4. **Foster Active Learning in the Classroom.**
   While this recommendation speaks more to pedagogy and classroom practice, S*tatistics: Learning from Data* 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 for Developing Conceptual Understanding and Analyzing Data.**
   The computer has brought incredible statistical power to the desktop of every investigator. The wide availability of statistical computer packages, such as JMP, Minitab, 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, sample output is included throughout the book. In addition, numerous exercises contain data that can easily be analyzed using statistical software. JMP data analysis software is provided with the text, and technology manuals for JMP and for other software packages, such as Minitab and SPSS, and for the graphing calculator are available in the online materials that accompany this text.
6. **Use Assessments to Improve and Evaluate Student Learning.**
   Comprehensive chapter review exercises that are specifically linked to chapter learning objectives are included at the end of each chapter. In addition, 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.

Instructor and Student Resources

MindTap™

The Personal Learning Experience

MindTap for Peck **Statistics: Learning from Data** is a personalized, fully online digital learning platform of authoritative textbook content, Aplia's online learning and homework system, and services that engages your students with interactivity while also offering choices in the configuration of coursework and enhancement of the curriculum via complimentary Web apps known as MindApps. MindApps range from Aplia, ReadSpeaker (which reads the text out loud to students), to Kaltura (allowing you to insert inline video and audio into your curriculum), to ConnectYard (allowing you to create digital "yards" through social media—all without "friending" your students). MindTap is well beyond an eBook, a homework solution or digital supplement, a resource center Web site, a course delivery platform, or a Learning Management System. It is the first in a new category—the Personal Learning Experience.

Aplia™

Content

Aplia™ is an online interactive learning solution that improves comprehension and outcomes by increasing student effort and engagement. Founded by a professor to enhance his own courses, Aplia provides automatically graded assignments with detailed, immediate explanations for every question, along with innovative teaching materials. Our easy-to-use system has been used by more than 1,000,000 students at over 1,800 institutions. Exercises are taken directly from text.

Aplia homework engages students in critical thinking, requiring them to synthesize and apply knowledge, not simply recall it. The diverse types of questions reflect the types of exercises that help students learn. All homework is written by subject matter experts in the field who have taught the course before.

Aplia contains a robust course management system with powerful analytics, enabling professors to track student performance easily.

Service

Your adoption of Aplia® includes CourseCare, Cengage Learning's industry leading service and training program designed to ensure that you have everything that you need to make the most of your use of Aplia. CourseCare provides one-on-one service, from finding the right solutions for your course to training and support. A team of Cengage representatives, including Digital Solutions Managers and Coordinators as well as Service and Training Consultants assist you every step of the way. For additional information about CourseCare, please visit http:// www.cengage.com/coursecare.

Our Aplia training program provides a comprehensive curriculum of beginner, intermediate, and advanced sessions, designed to get you started and effectively integrate Aplia into your course. We offer a flexible online and recorded training program designed to accommodate your busy schedule. Whether you are using Aplia for the first time or are an experienced user, there is a training option to meet your needs.

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 CengageBrain.com or http://www.jmp.com/getse. Find out more at www.jmp.com.

Student Resources

Digital

To access additional course materials and companion resources, please visit www.cengagebrain.com. At the CengageBrain.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.

If your text includes a printed access card, you will have instant access to the following resources:

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

Print

Student Solutions Manual (ISBN: 9781285089836): Contains fully worked-out solutions to all of the Exercise Set 1 and odd-numbered additional 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

Teacher's Resource Binder (ISBN: 9781285094632): The Teacher's Resource Binder is full of wonderful resources for both college professors and AP Statistics teachers. These include:

- Recommendations for instructors on how to teach the course, including sample syllabi, pacing guides, and teaching tips.
- Recommendations for what students should read and review for a particular class period or set of class periods.
- Extensive notes on preparing students to take the AP exam.
- 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 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.
- Activity worksheets that can be duplicated and used in class.
- A test bank that includes assessment items, quizzes, and chapter exams.

Digital

- Solution Builder: This online instructor database offers complete worked-out solutions to all exercises in the text, allowing you to create customized, secure solutions printouts (in PDF format) matched exactly to the problems you assign in class. Sign up for access at www.cengage.com/solutionbuilder.
- E-book: This new premium eBook has highlighting, note-taking, and search features as well as links to multimedia resources.

Acknowledgments

I would like to express my thanks and gratitude to the following people who made this book possible:

Molly Taylor, my editor at Cengage, for her support and encouragement.

Jay Campbell, my development editor, for his unfailing good humor and his ability to field just about any curve ball I threw at him.

Alison Eigel Zade, the content project manager.

Ed Dionne, Lindsay Schmonsees and Liah Rose, project managers at MPS Limited.

Elena Montillo, the production manager.

Andrew Coppola, associate media editor, for managing the media content and for creating the implementation of the "Exploring the Bid Ideas" activities.

Cameron Troxell, for his careful review of the manuscript and many helpful suggestions for improving the readability of the book.

Stephen Miller, for his great work on the huge task of creating the student and instructor solutions manuals.

Michael Allwood, for his detailed work in checking the accuracy of examples and solutions.

Kathy Fritz, for creating the interactive PowerPoint presentations that accompany the text.

Melissa Sovak, for creating the Technology Notes sections.

Nicole Mollica, the developmental editor for marketing.

Ryan Ahern, the Market Development Manager.

Beth Chance and Francisco Garcia for producing the applet used in the confidence interval activities.

Gary McClelland for producing the applets from *Seeing Statistics* used in the regression activities.

Chris Sabooni, the copy editor for the book.

MPS, for producing the artwork used in the book.

I would also like to give a special thanks to those who served on the Editorial Board for the book and those who class tested some of the chapters with their students:

Editorial Board

John Climent, Cecil College
Ginger Holmes Rowell, Middle Tennessee State University
Karen Kinard, Tallahassee Community College
Glenn Miller, Borough of Manhattan Community College
Ron Palcic, Johnson County Community College

Class Testers

Rob Eby, Blinn College, Bryan Campus
Sumona Mondal, Clarkson University
Cameron Troxell, Mount San Antonio College
Cathy Zucco-Teveloff, Rowan University

Many people provided invaluable comments and suggestion as this text was being developed. I would like to thank the following people for their contributions:

Reviewers

Gregory Bloxom, Pensacola State College
Denise Brown, Collin College
Jerry Chen, Suffolk County Community College
Monte Cheney, Central Oregon Community College
Ivette Chuca, El Paso Community College
Mary Ann Connors, Westfield State University
George Davis, Georgia State University
Rob Eby, Blinn College, Bryan Campus
Karen Estes, St. Petersburg College
Larry Feldman, Indiana University, Pennsylvania
Kevin Fox, Shasta College
Melinda Holt, Sam Houston State University
Kelly Jackson, Camden County College
Clarence Johnson, Cuyahoga Community College
Hoon Kim, Calfornia State Polytechnic University, Pomona
Jackie MacLaughlin, Central Piedmont
Nola McDaniel, McNeese State
Glenn Miller, Borough of Manhattan Community College
Philip Miller, Indiana University SE
Kathy Mowers, Owensboro Community and Technical College
Linda Myers, Harrisburg Area Community College
Ron Palcic, Johnson County Community College
Nishant Patel, Northwest Florida State College
Maureen Petkewich, Univiversity of South Carolina
Nancy Pevey, Pellissippi State Community College
Chandler Pike, University of Georgia
Blanche Presley, Macon State College
Daniel Rowe, Heartland Community College
Fary Sami, Harford Community College
Laura Sather, St. Cloud State University
Sean Simpson, Westchester Community College
Sam Soleymani, Santa Monica College
Cameron Troxell, Mount San Antonio College
Diane Van Deusen, Napa Valley College
Richard Watkins, Tidewater Community College
Jane-Marie Wright, Suffolk Community College
Cathy Zucco-Teveloff, Rider University

Focus Group Participants (JSM and AMATYC)

Raoul Amstelveen, Johnson & Wales University
Melissa Bingham, University of Wisconsin-La Crosse
Norma Biscula, University of Maine at Augusta
Patricia Blus, National Louis University
Joe Castillo, Broward College
Andy Chang, Youngstown State University
Adrienne Chu, Suffolk County Community College
Julie Clark, Hollins University
Julie DePree, University of New Mexico
Elizabeth Donovan, Bristol Community College
Debbie Garrison, Valencia Community College
Brian Gill, Seattle Pacific University
Kathy Gray, California State University, Chico
David Gurney, Southeastern Louisiana University
Queen Harris, Georgia Perimeter College
Patricia Humphrey, Georgia Southern University
Debbie Hydron, University of Mary Washington
Maryann Justinger, Eric Community College-South
Jack Keating, Massasoit Community College
Alex Kolesnik, Ventura College
Doug Mace, Kirtland Community College
Erum Marfani, Frederick Community College
Eric Matsuoka, Leeward Community College
Glenn Miller, Borough of Manhattan Community College
Philip Miller, Indiana University SE
Megan Mocko, University of Florida
Adam Molnar, Bellarmine University
Sumona Mondal, Clarkson University
Calandra Moore, CUNY – College of Staten Island
Penny Morris, Polk State College
Kathy Mowers, Owensboro Community and Technical College
Linda Myers, Harrisburg Area Community College
Blanche Presley, Macon State College
Elaine A. Previte, Bristol Community College
Ahmed Rashed, Richland College
Patricia Rhodes, Treasure Valley Community College
Ned Schillow, Lehigh-Carbon Community College
Gail St. Jacques, Johnson & Wales University
Jenting Wang, SUNY-Oneonta
Jane-Marie Wright, Suffolk Community College
Rebecca Wong, West Valley College

Webinar Participants

Bilal Abu Bakr, Alcorn State University
Cheryl Adams, York College, CUNY
Josiah Alamu, University of Illinois-Springfield
Raoul Amstelveen, Johnson & Wales University
Gokarna Aryal, Purdue University, Calumet
Tatiana Arzivian, Edison State College
Christopher Barat, Stevenson University
J.B. Boren, Wayland Baptist University-Amarillo
Monica Brown, St. Catherine University
Brenda Burger, University of South Florida-St. Petersburg
Rose Elaine Carbone, Clarion University
Julie Clark, Hollins University
Vera Dauffenbach, Bellin College
Boyan Dimitrov, Kettering University
William Dougherty, SCCC

Robert Ellis, New York University
Michelle Everson, University of Minnesota
Mark Goldstein, West Virginia Northern Community College
David Gurney, Southeastern Louisiana University
Julie Hanson, Clinton Community College
Debra Hydorn, University of Mary Washington
Marc Isaacson, Augsburg College
Grazyna Kamburowska, SUNY – Oneonta
Alex Kolesnik, Ventura College
Jeff Kollath, Oregon State University
Philip Larson, Greenville Technical College
Pam Lowry, Bellvue College
John McKenzie, Babson College
Michael Moran, Shoreline Community College
Jerry Moreno, John Carroll Univeristy
Kathy Mowers, Owensboro Community and Technical College
Rebecca Nichols, American Statistical Association
Leesa Phol, Donnelly College
Rebecca Pierce, Ball State University
Gautam Pillay, Rowan University
Wendy Pogoda, Hillsborough Community College
Gyaneswor Pokharel, Mountain State University
Krish Revuluri, Harper College
Helen Roberts, Montclair State University
Dave Saha, Florida Gulf Coast University
Jennifer Sinclair, Foothill College
Thomas Strommer, Fayetteville Technical Community College
Lori Thomas, Midland College
Elizabeth Uptegrove, Felician College
Jennifer Ward, Clark College
Joan Weinstein, Harvard Extension School
Pamela Weston, Hiwassee College
Igor Woiciechowski, Alderson-Broaddus College
Barry Woods, Unity College

Content Survey Participants

Gregg Bell, University of Alabama – Tuscaloosa
Gulhan Bourget, California State University Fullerton
Pete Bouzar, Golden West College
Ralph Burr, Mesa Community College
Joe Castille, Broward County Community College
Lyle Cook, Eastern Kentucky University
Amy Curry, College of Lake County
Charles De Vogelaere, De Anza College
Ola Disu, Tarrant County College, NE
James Dulgeroff, San Bernardino Valley College
Debbie Garrison, Valencia Community College
Donna Gorton, Butler Community College
Fred Gruhl, Washtenaw Community College
Greg Henderson, Hillsborough Community College
Leslie Hendrix, University of South Carolina
Robert Indik, University of Arizona
Jay Jahangiri, Kent State University at Geauga
Jack Keating, Massasoit Community College
JoAnn Kump, West Chester University
Rudy Maglio, Northeastern Illinois University
Richard McGowan, Boston College
Wendy Miao, El Camino College
Jackie Miller, The Ohio State University
Kathy Mowers, Owensboro Community and Technical College
Alfredo Rodriguez, El Paso Community College
Thomas Roe, South Dakota State University
Rosa Seyfried, Harrisburg Area Community College
Aileen Solomon, Trident Technical College
T. C. Sun, Wayne State University
Mahbobeh Vezvaei, Kent State University
Dennis Walsh, Middle Tennessee State University
Sheila Weaver, University of Vermont
Dave Wilson, SUNY Buffalo State
Zhiwei Zhu, University of Louisiana at Lafayette

Marketing Survey Participants

Jerry Chen, Suffolk County Community College
Ivette Chuca, El Paso Community College
Rob Eby, Blinn College, Bryan Campus
Larry Feldman, Indiana University, Pennsylvania
Patricia Humphrey, Georgia Southern University
Kelly Jackson, Camden County College
Glenn Miller, Borough of Manhattan Community College
Megan Mocko, University of Florida
Sumona Mondal, Clarkson University
Ron Palcic, Johnson County Community College
Aileen Solomon, Trident Technical College
Jenting Wang, SUNY-Oneonta
Rebecca Wong, West Valley College
Cathy Zucco-Teveloff, Rider University

And last, but certainly not least, I thank my family, friends, and colleagues for their continued support.

Roxy Peck

STATISTICS: LEARNING FROM DATA

Overview

Learning From Data

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 on the effectiveness of apricots as a treatment for arthritis.

It is difficult to function in today's world without a basic understanding of statistics. For example, here are a few headlines from articles in a single issue of ***USA Today*** **(June 29, 2009)** that draw conclusions based on data:

- **"Poll Finds Generation Gap Biggest Since Vietnam War"** summarized a study that explored opinions regarding social values and political views. Not surprisingly, large behavioral differences between young and old were noted in the use of the Internet, cell phones, and text messaging.
- **"Few See Themselves as 'Old' No Matter What Their Age"** described results from a survey of 2,969 adults. Those surveyed were asked at what age a person would be considered old. The resulting data revealed that there were notable differences in the answer to this question depending on the age of the responder. The average age identified as old by young adults (ages 18 to 29) was 60, while the average was 69 for those who were ages 30 to 49, 72 for those ages 50 to 64, and 74 for those ages 65 and older.
- **"If You Were Given $1,000, What Would You Do?"** reported on one aspect of a study of consumer purchasing and saving behavior. Something was definitely amiss in this report, however—the percentages for the response categories (such as save it, pay off credit card debt, and use it for a vacation) added up to 107%!
- **"Many Adults Can't Name a Scientist"** summarized the results of a survey of 1,000 adults. Of those surveyed, 23% were unable to name a single famous scientist. Of those who did come up with a name, Albert Einstein was the scientist of choice, named by 47% of those surveyed.

To be an informed consumer of reports like these, you must be able to:

- Extract information from tables, charts, and graphs.
- Follow numerical arguments.
- Understand how data should be gathered, summarized, and analyzed to draw valid conclusions.

In addition to preparing you to critically evaluate the work of others, studying statistics will enable you to plan statistical studies, collect data in a sensible way, and use data to answer questions of interest.

Throughout your personal and professional life, you will need to use data to make informed decisions. Should you go out for a sport that involves the risk of injury? Will your college club do better by trying to raise funds with a benefit concert or with a direct appeal for donations? If you choose a particular major, what are your chances of finding a job when you graduate? How should you select a graduate program based on guidebook

ratings that include information on percentage of applicants accepted, time to obtain a degree, and so on? Your statistics course will provide you with a systematic approach to making decisions based on data.

Statistics—It's All About Variability

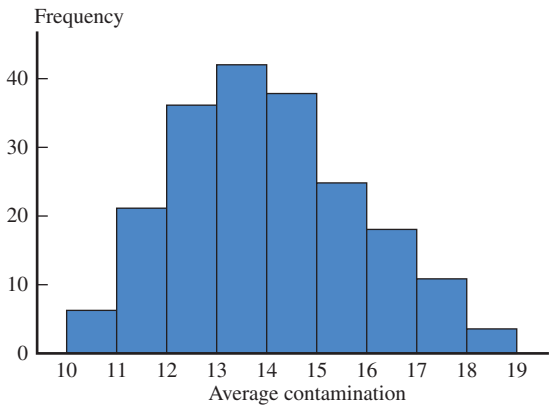
Statistical methods allow you to collect, describe, analyze, and draw conclusions from data. If you lived in a world where all measurements were identical for every individual, these tasks would be simple. For example, consider a population consisting of all of the students at your college. Suppose that *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 person studying students from this population to draw conclusions about any of these three variables would have a particularly easy task. It would not matter how many students were studied or how the students were selected. In fact, you 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 the library. Because there is no variability in the population, this one individual would provide complete and accurate information about the population, and you could draw conclusions with no risk of error.

The situation just described is obviously unrealistic. Populations with no variability are exceedingly rare, and they are of little statistical interest because they present no challenge. In fact, variability is almost universal. It is variability that makes life interesting. You need to develop an understanding of variability to be able to collect, describe, analyze, and draw conclusions from data in a sensible way.

The following example illustrates how describing and understanding variability provide the foundation for learning from data.

Overview Example: Monitoring Water Quality

As part of its regular water quality monitoring efforts, an environmental control board selects five water specimens from a particular well each day. The concentration of contaminants in parts per million (ppm) is measured for each of the five specimens, and then the average of the five measurements is calculated. The following graph is an example of a histogram. (You will learn how to construct and interpret histograms in Chapter 2.) This histogram summarizes the average contamination values for 200 days.



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.

One month after the spill, five water specimens are collected from the well, and the average contamination is 15.5 ppm. Considering the variation before the spill shown in

the histogram, would you interpret this as evidence that the well water was affected by the spill? What if the calculated average was 17.4 ppm? How about 22.0 ppm?

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, you would probably conclude that the well contamination level has increased.

Reaching a conclusion in the water quality example required an understanding of variability. Understanding variability allows you to distinguish between usual and unusual values. The ability to recognize unusual values in the presence of variability is an important aspect of most statistical methods and is also what enables you to quantify the chance of being incorrect when a conclusion is based on available data.

The Data Analysis Process

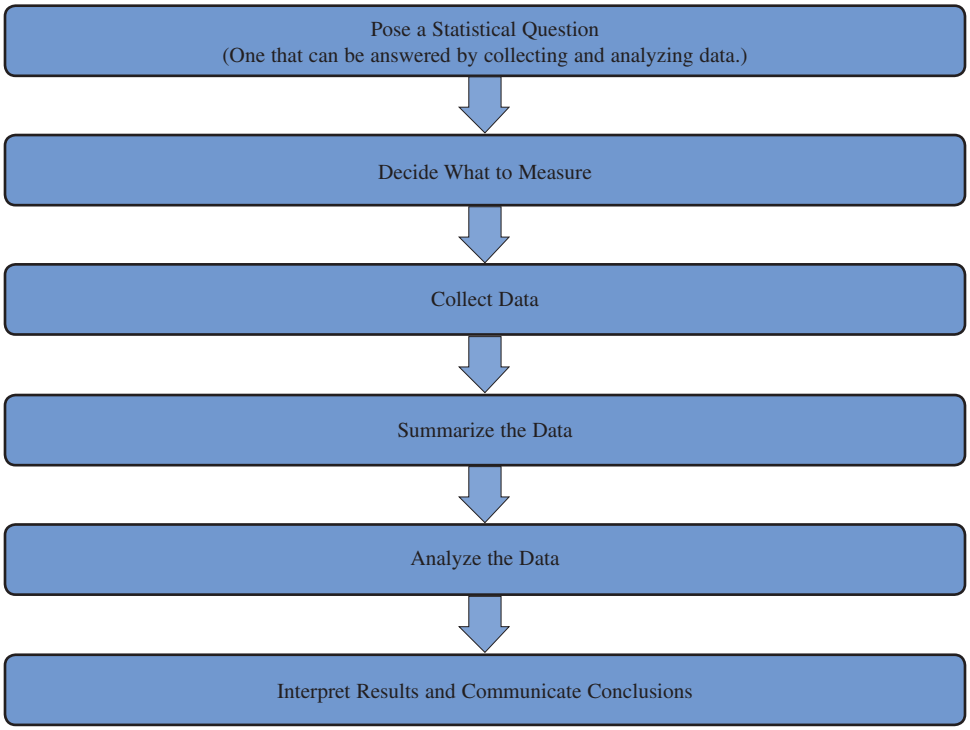
Statistics involves collecting, summarizing, and analyzing data. All three tasks are critical. Without summarization and analysis, raw data are of little value, and even sophisticated analyses can't produce meaningful information from data that were not collected in a sensible way.

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? How many credit cards do college students have? Data collection and analysis allow you to answer questions like these.

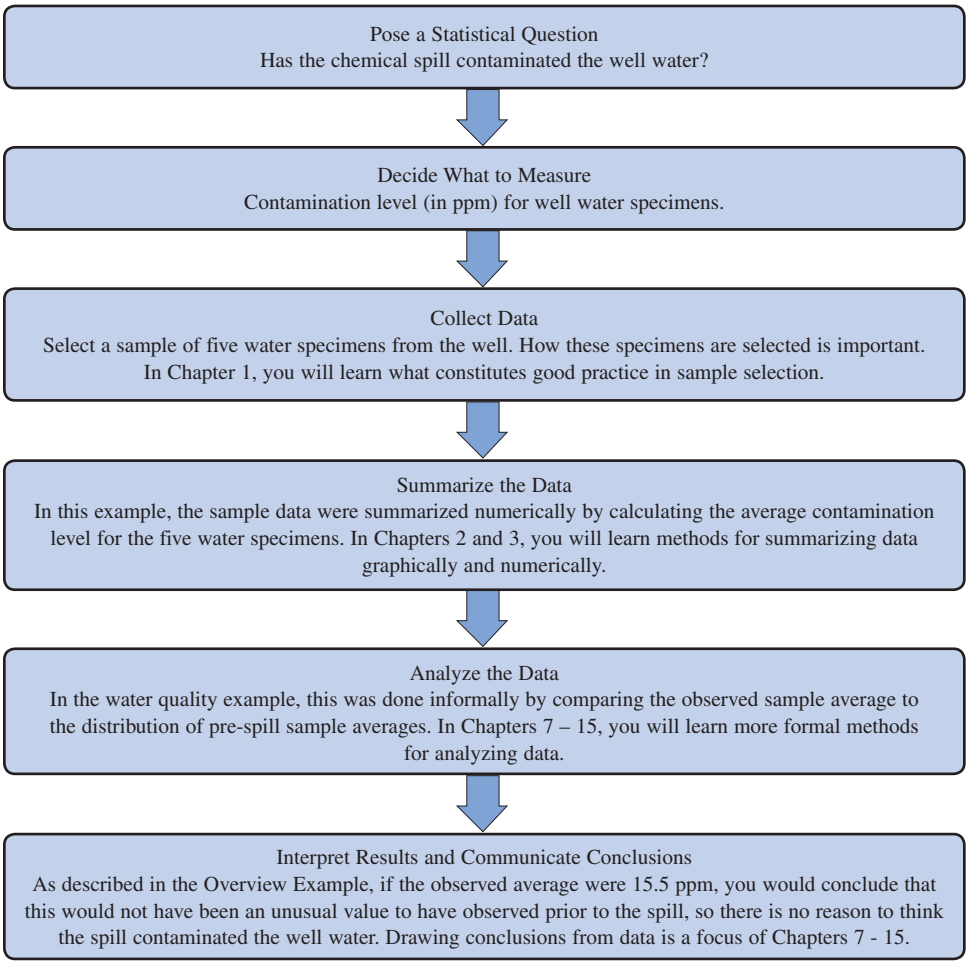
Data analysis is a process that can be viewed as a sequence of steps leading from planning to data collection to making informed conclusions based on the resulting data. The process can be organized into the following six steps:

1. **Understand the questions of interest.** Effective data analysis requires an understanding of the research problem. You must know the goal of the research and what questions you hope to answer. It is important to have a clear direction before gathering data to ensure that you will be able to answer the questions of interest using the data collected.
2. **Decide 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, you would need to collect data on player weight and position). In other cases, the choice is not as straightforward (for example, in a study of the relationship between preferred learning style and intelligence, how would you define learning style and measure it, and what measure of intelligence would you use?). It is important to carefully define the variables to be studied and to develop appropriate methods for determining their values.
3. **Collect data.** The data collection step is crucial. You must first decide whether an existing data source is adequate or whether new data must be collected. Even 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 and judged to be acceptable. 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. **Summarize the data.** After data are collected, the next step usually involves summarizing the data graphically and numerically. This summarization provides insight into important characteristics of the data and can also provide guidance in selecting appropriate methods for further analysis.
5. **Analyze the data.** This step involves selecting and applying appropriate statistical methods. Much of this textbook is devoted to methods that can be used to carry out this step.
6. **Interpret results.** The final step in the data analysis process involves communicating what you have learned from the data. This step often leads to the formulation of new research questions, which, in turn, leads back to the first step. In this way, good data analysis is often an iterative process.

The data analysis process is summarized in the following figure.



For the water quality setting of the Overview Example, these steps might be



Goals for Student Learning

Statistics is about learning from data and the role that variability plays in drawing conclusions from data. To be successful, you need to develop an understanding of the important concepts, master the computational aspects, and be able to combine conceptual understanding and mastery of the mechanics in a way that allows you to learn from data. This textbook has been written with these three broad goals in mind. At the beginning of each chapter, you will find chapter learning objectives organized under the following headings:

- Conceptual Understanding
- Mastering the Mechanics
- Putting It into Practice

As you begin each chapter, it is a good idea to take a few minutes to familiarize yourself with the learning objectives for that chapter.

The Structure of the Chapters That Follow

The examples and exercises in each chapter have been carefully selected with the stated chapter learning objectives in mind. At the end of each section, you will find two exercise sets. Each of these sets is designed to assess all of the learning objectives addressed in that section.

In addition, at the end of each chapter, you will find a comprehensive set of review exercises titled "Are You Ready to Move On?" This collection of exercises assesses all of the chapter learning objectives, and the learning objectives assessed are identified for each exercise. As you complete your work in any given chapter, it is a good idea to test yourself by working the "Are You Ready to Move On?" exercises. This will help you to solidify your knowledge and ensure that you are ready to move on to the next chapter.

1

Collecting Data in Reasonable Ways

Phil Date/Shutterstock.com

PREVIEW

Statistical methods help you to make sense of data and gain insight into the world around you. The ability to learn from data is critical for success in your personal and professional life. Data and conclusions based on data are everywhere—in newspapers, magazines, online resources, and professional publications. But should you believe what you read? For example, should you eat garlic to prevent a cold? Will doing tai chi (exercises designed for relaxation and balance) one hour per week increase the effectiveness of your flu shot? Will eating cheese before going to bed help you sleep better? These are just three recommendations out of many that appeared in one issue of ***Woman's World*** **(September 27, 2010)**, *a magazine with 1.6 million readers. In fact, if you followed all of the recommendations in that issue, you would also be loading up on sweet potatoes, black tea, yogurt, walnuts, grape juice, olive oil, peanuts, and strawberries! Some of these recommendations are supported by evidence (data) from research studies, but how reliable is this evidence? Are the conclusions drawn reasonable, and do they apply to you? These are important questions that will be explored in this chapter.*

CHAPTER LEARNING OBJECTIVES

Conceptual Understanding

After completing this chapter, you should be able to

C1 Understand the difference between an observational study and an experiment.
C2 Understand that the conclusions that can be drawn from a statistical study depend on the way in which the data are collected.
C3 Explain the difference between a census and a sample.
C4 Explain the difference between a statistic and a population characteristic.
C5 Understand why random selection is an important component of a sampling plan.
C6 Understand why random assignment is important when collecting data in an experiment.
C7 Understand the difference between random selection and random assignment.
C8 Explain why volunteer response samples and convenience samples are unlikely to produce reliable information about a population.
C9 Understand the limitations of using volunteers as subjects in an experiment.
C10 Explain the purpose of a control group in an experiment.
C11 Explain the purpose of blinding in an experiment.

Mastering the Mechanics

After completing this chapter, you should be able to

M1 Create a sampling plan that would result in a simple random sample from a given population.
M2 Describe a procedure for randomly assigning experimental units to experimental conditions (for example, subjects to treatments) given a description of an experiment, the experimental conditions, and the experimental units.

Putting It into Practice

After completing this chapter, you should be able to

P1 Distinguish between an observational study and an experiment.
P2 Evaluate the design of an observational study.
P3 Evaluate the design of a simple comparative experiment.
P4 Evaluate whether conclusions drawn from the study are appropriate, given a description of a statistical study.

SECTION 1.1 Statistical Studies: Observation and Experimentation

If the goal is to make good decisions based on data, it should come as no surprise that the way you obtain the data is very important. It is also important to know what questions you hope to answer with the data. Depending on what you want to learn, two types of statistical studies are common—*observational studies* and *experiments*.

Sometimes you are interested in answering questions about characteristics of a single **population** or in comparing two or more well-defined populations. To accomplish this, you select a **sample** from each population and use information from the samples to learn about characteristics of the populations.

DEFINITION

Population: The population is the entire collection of individuals or objects that you want to learn about.

Sample: A sample is a part of the population that is selected for study.

For example, many people, including the author of **"The 'CSI Effect': Does It Really Exist?" (*National Institute of Justice* [2008]: 1–7)**, have speculated that watching crime