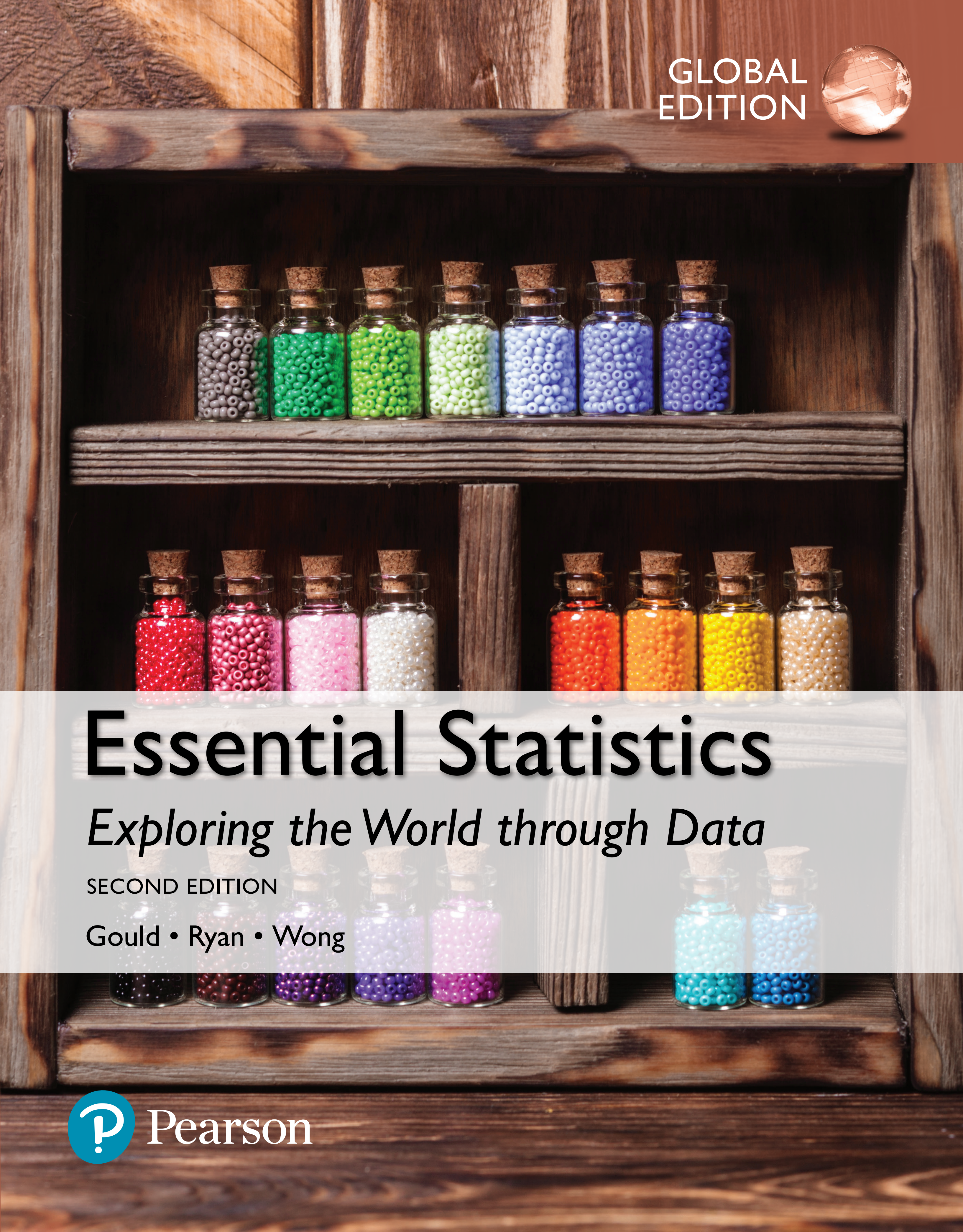


GLOBAL EDITION

Essential Statistics

Exploring the World through Data

SECOND EDITION

Gould • Ryan • Wong

Pearson

This page intentionally left blank

Essential Statistics: Exploring the World through Data

Second Edition

Global Edition

Robert Gould

University of California, Los Angeles

Colleen Ryan

California Lutheran University

Rebecca Wong

West Valley College

Pearson

Boston Columbus Indianapolis New York San Francisco Amsterdam
Cape Town Dubai London Madrid Milan Munich Paris Montreal Toronto Delhi
Mexico City Sao Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

Editor in Chief: Deirdre Lynch
Senior Acquisitions Editor: Suzanna Bainbridge
Editorial Assistant: Justin Billing
Field Marketing Manager: Andrew Noble
Product Marketing Manager: Tiffany Bitzel
Marketing Assistant: Jennifer Myers
Program Team Lead: Karen Wernholm
Program Manager: Chere Bemelmans
Project Team Lead: Peter Silvia
Project Manager: Peggy McMahon
Senior Author Support/Technology Specialist: Joe Vetere
Manager, Rights Management, Higher Education: Gina M. Cheselka
Media Producer: Aimee Thorne
Acquisitions Editor, Global Edition: Sourabh Maheshwari
Assistant Project Editor, Global Edition: Sulagna Dasgupta
Media Production Manager, Global Edition: Vikram Kumar
Senior Manufacturing Controller, Production, Global Edition: Trudy Kimber
Associate Director of Design, USHE EMSS/HSC/EDU: Andrea Nix
Program Design Lead: Beth Paquin
Design, Full-Service Project Management, Composition, and Illustration: Cenveo® Publisher Services
Senior Project Manager, MyStatLab: Robert Carroll
QA Manager, Assessment Content: Marty Wright
Procurement Manager: Carol Melville

Pearson Education Limited
Edinburgh Gate
Harlow
Essex CM20 2JE
England

and Associated Companies throughout the world

Visit us on the World Wide Web at:
www.pearsonglobaleditions.com

Authorized adaptation from the United States edition, entitled Essential Statistics, 2nd edition, *ISBN 978-0-134-13440-6, by Robert Gould, Colleen Ryan, and Rebecca Wong, published by Pearson Education © 2017.*

ISBN 10: 1-292-16122-1
ISBN 13: 978-1-292-16122-8

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

10 9 8 7 6 5 4 3 2 1

Printed and bound by Vivar in Malaysia

Dedication

To my parents and family, my friends, and my colleagues who are also friends. Without their patience and support, this would not have been possible.

—Rob

To my teachers and students, and to my family who have helped me in many different ways.

—Colleen

To my students, colleagues, family, and friends who have helped me be a better teacher and a better person.

—Rebecca

About the Authors

Robert Gould

Robert L. Gould (Ph.D., University of California, San Diego) is a leader in the statistics education community. He has served as chair of the American Statistical Association's Committee on Teacher Enhancement, has served as chair of the ASA's Statistics Education Section, and served on a panel of co-authors for the *Guidelines for Assessment in Instruction on Statistics Education (GAISE) College Report*. While serving as the associate director of professional development for CAUSE (Consortium for the Advancement of Undergraduate Statistics Education), Rob worked closely with the American Mathematical Association of Two-Year Colleges (AMATYC) to provide traveling workshops and summer institutes in statistics. For over ten years, he has served as Vice-Chair of Undergraduate Studies at the UCLA Department of Statistics, and he is director of the UCLA Center for the Teaching of Statistics. In 2012, Rob was elected Fellow of the American Statistical Association.

In his free time, Rob plays the cello and is an ardent reader of fiction.

Colleen Ryan

Colleen N. Ryan has taught statistics, chemistry, and physics to diverse community college students for decades. She taught at Oxnard College from 1975 to 2006, where she earned the Teacher of the Year Award. Colleen currently teaches statistics part-time at California Lutheran University. She often designs her own lab activities. Her passion is to discover new ways to make statistical theory practical, easy to understand, and sometimes even fun.

Colleen earned a B.A. in physics from Wellesley College, an M.A.T. in physics from Harvard University, and an M.A. in chemistry from Wellesley College. Her first exposure to statistics was with Frederick Mosteller at Harvard.

In her spare time, Colleen sings, has been an avid skier, and enjoys time with her family.

Rebecca K. Wong

Rebecca K. Wong has taught mathematics and statistics at West Valley College for more than twenty years. She enjoys designing activities to help students actively explore statistical concepts and encouraging students to apply those concepts to areas of personal interest.

Rebecca earned a B.A. in mathematics and psychology from the University of California, Santa Barbara, an M.S.T. in mathematics from Santa Clara University, and an Ed.D. in Educational Leadership from San Francisco State University. She has been recognized for outstanding teaching by the National Institute of Staff and Organizational Development and the California Mathematics Council of Community Colleges.

When not teaching, Rebecca is an avid reader and enjoys hiking trails with friends.

Contents

This page intentionally left blank

Preface

About This Text

The primary focus of this text is still, as in the first edition, data. We live in a data-driven economy and, more and more, in a data-centered culture. We don't choose whether we interact with data; the choice is made for us by websites that track our browsing patterns, membership cards that track our spending habits, cars that transmit our driving patterns, and smart phones that record our most personal moments.

The silver lining of what some have called the Data Deluge is that we all have access to rich and valuable data relevant in many important fields: environment, civics, social sciences, economics, health care, entertainment. This text teaches students to learn from such data and, we hope, to become cognizant of the role of the data that appear all around them. We want students to develop a data habit of mind in which, when faced with decisions, claims, or just plain curiosity, they know to reach for an appropriate data set to answer their questions. More important, we want them to have the skills to access these data and the understanding to analyze the data critically. Clearly, we've come a long way from the "mean median mode" days of rote calculation. To survive in the modern economy requires much more than knowing how to plug numbers into a formula. Today's students must know which questions can be answered by applying which statistic, and how to get technology to compute these statistics from within complex data sets.

What's New in the Second Edition

The second edition remains true to the goals of the first edition: to provide students with the tools they need to make sense of the world by teaching them to collect, visualize, analyze, and interpret data. With the help of several wise and careful readers and class testers, we have fine-tuned the second edition to better achieve this vision. In some sections, we have rewritten explanations or added new ones. In others, we have more substantially reordered content.

More precisely, in this new edition you will find

- Coverage of two-proportion confidence intervals in Chapters 7 and 8.
- An increase of more than 150 homework exercises in this edition, with more than 400 total new, revised, and updated exercises. We've added larger data sets to Chapters 2, 3, 4, and 9. We've also added exercises to Section 2.5 and more Chapter Review exercises throughout.
- New or updated examples in each chapter, with current topics such as views of stem cell research (Chapter 7) and online classes (Chapter 10).
- A more careful and thorough integration of technology in many examples.
- Two new case studies: Student-to-Teacher Ratios in Chapter 2 and Dodging the Question in Chapter 8.
- A more straightforward implementation of simulations to understand probability in Chapter 5.
- A more unified presentation of hypothesis testing in Chapter 8 that better joins conceptual understanding with application.
- A greater number of "Looking Back" and "Caution" marginal boxes to help direct students' studies.
- Updated technology guides to match current hardware and software.

Approach

Our text is concept-based, as opposed to method-based. We teach useful statistical methods, but we emphasize that applying the method is secondary to understanding the concept.

In the real world, computers do most of the heavy lifting for statisticians. We therefore adopt an approach that frees the instructor from having to teach tedious procedures and leaves more time for teaching deeper understanding of concepts. Accordingly, we present formulas as an aid to understanding the concepts, rather than as the focus of study.

We believe students need to learn how to

- Determine which statistical procedures are appropriate.
- Instruct the software to carry out the procedures.
- Interpret the output.

We understand that students will probably see only one type of statistical software in class. But we believe it is useful for students to compare output from several different sources, so in some examples we ask them to read output from two or more software packages.

One of the authors (Rob Gould) served on a panel of co-authors for the first edition of the collegiate version of the American Statistical Association–endorsed *Guidelines for Assessment and Instruction in Statistics Education* (*GAISE*). We firmly believe in its main goals and have adopted them in the preparation of this book.

- We emphasize understanding over rote performance of procedures.
- We use real data whenever possible.
- We encourage the use of technology both to develop conceptual understanding and to analyze data.
- We believe strongly that students learn by doing. For this reason, the homework problems offer students both practice in basic procedures and challenges to build conceptual understanding.

Coverage

The first few chapters of this book are concept-driven and cover exploratory data analysis and inferential statistics—fundamental concepts that every introductory statistics student should learn. The last part of the book builds on that strong conceptual foundation and is more methods-based. It presents several popular statistical methods and more fully explores methods presented earlier.

Our ordering of topics is guided by the process through which students should analyze data. First, they explore and describe data, possibly deciding that graphics and numerical summaries provide sufficient insight. Then they make generalizations (inferences) about the larger world.

Chapters 1–4: Exploratory Data Analysis. The first four chapters cover data collection and summary. Chapter 1 introduces the important topic of data collection and compares and contrasts observational studies with controlled experiments. This chapter also teaches students how to handle raw data so that the data can be uploaded to their statistical software. Chapters 2 and 3 discuss graphical and numerical summaries of single variables based on samples. We emphasize that the purpose is not just to produce a graph or a number but, instead, to explain what those graphs and numbers say about the world. Chapter 4 introduces simple linear regression and presents it as a technique for providing graphical and numerical summaries of relationships between two numerical variables.

We feel strongly that introducing regression early in the text is beneficial in building student understanding of the applicability of statistics to real-world scenarios. After completing the chapters covering data collection and summary, students have acquired the skills and sophistication they need to describe two-variable associations and to generate informal hypotheses. Two-variable associations provide a rich context for class discussion and allow the course to move from fabricated problems (because one-variable analyses are relatively rare in the real world) to real problems that appear frequently in everyday life.

Chapters 5–8: Inference. These chapters teach the fundamental concepts of statistical inference. The main idea is that our data mirror the real world, but imperfectly; although our estimates are uncertain, under the right conditions we can quantify our uncertainty. Verifying that these conditions exist and understanding what happens if they are not satisfied are important themes of these chapters.

Chapters 9–10: Methods. Here we return to the themes covered earlier in the text and present them in a new context by introducing additional statistical methods, such as estimating population means and analyzing categorical variables. We also provide (in Section 10.3) guidance for reading scientific literature, to offer students the experience of critically examining real scientific papers.

Organization

Our preferred order of progressing through the text is reflected in the Contents, but there are some alternative pathways as well.

10-week Quarter. The first eight chapters provide a full, one-quarter course in introductory statistics. If time remains, cover Sections 9.1 and 9.2 as well, so that students can solidify their understanding of confidence intervals and hypothesis tests by revisiting the topic with a new parameter.

Proportions First. Ask two statisticians, and you will get three opinions on whether it is best to teach means or proportions first. We have come down on the side of proportions for a variety of reasons. Proportions are much easier to find in popular news media (particularly around election time), so they can more readily be tied to students' everyday lives. Also, the mathematics and statistical theory are simpler; because there's no need to provide a separate estimate for the population standard deviation, inference is based on the Normal distribution, and no further approximations (that is, the t-distribution) are required. Hence, we can quickly get to the heart of the matter with fewer technical diversions.

The basic problem here is how to quantify the uncertainty involved in estimating a parameter and how to quantify the probability of making incorrect decisions when posing hypotheses. We cover these ideas in detail in the context of proportions. Students can then more easily learn how these same concepts are applied in the new context of means (and any other parameter they may need to estimate).

Means First. Conversely, many people feel that there is time for only one parameter and that this parameter should be the mean. For this alternative presentation, cover Chapters 6, 7, and 9, in that order. On this path, students learn about survey sampling and the terminology of inference (population vs. sample, parameter vs. statistic) and then tackle inference for the mean, including hypothesis testing.

To minimize the coverage of proportions, you might choose to cover Chapter 6, Section 7.1 (which treats the language and framework of statistical inference in detail), and then Chapter 9. Chapters 7 and 8 develop the concepts of statistical inference more slowly than Chapter 9, but essentially, Chapter 9 develops the same ideas in the context of the mean.

If you present Chapter 9 before Chapters 7 and 8, we recommend that you devote roughly twice as much time to Chapter 9 as you have devoted to previous chapters, because many challenging ideas are explored in this chapter. If you have already covered Chapters 7 and 8 thoroughly, Chapter 9 can be covered more quickly.

Features

We've incorporated into this text a variety of features to aid student learning and to facilitate its use in any classroom.

Integrating Technology

Modern statistics is inseparable from computers. We have worked to make this text accessible for any classroom, regardless of the level of in-class exposure to technology, while still remaining true to the demands of the analysis. We know that students sometimes do not have access to technology when doing homework, so many exercises provide output from software and ask students to interpret and critically evaluate that given output.

Using technology is important because it enables students to handle real data, and real data sets are often large and messy. The following features are designed to guide students.

- **TechTips** outline steps for performing calculations using TI-84® (including TI-84 + C®) graphing calculators, Excel®, Minitab®, and StatCrunch®. We do not want students to get stuck because they don't know how to reproduce the results we show in the text, so whenever a new method or procedure is introduced, an icon, Tech, refers students to the TechTips section at the end of the chapter. Each set of TechTips contains at least one mini-example, so that students are not only learning to use the technology but also practicing data analysis and reinforcing ideas discussed in the text. Most of the provided TI-84 steps apply to all TI-84 calculators, but some are unique to the TI-84 + C calculator.
- **Check Your Tech** examples help students understand that statistical calculations done by technology do not happen in a vacuum and assure them that they can get the same numerical values by hand. Although we place a higher value on interpreting results and verifying conditions required to apply statistical models, the numerical values are important, too.
- All **data sets** used in the exposition and exercises are available on the companion website at www.pearsonglobaleditions.com/gould.

Guiding Students

- Each chapter opens with a **Theme**. Beginners have difficulty seeing the forest for the trees, so we use a theme to give an overview of the chapter content.
- Each chapter begins by posing a real-world **Case Study**. At the end of the chapter, we show how techniques covered in the chapter helped solve the problem presented in the Case Study.
- **Margin Notes** draw attention to details that enhance student learning and reading comprehension.

  The data icon appears alongside examples or discussions to indicate that the original data are available on the companion website.

  Caution notes provide warnings about common mistakes or misconceptions.

Looking Back reminders refer students to earlier coverage of a topic.

Details clarify or expand on a concept.

- KEY POINT **Key Points** highlight essential concepts to draw special attention to them. Understanding these concepts is essential for progress.
- **Snapshots** break down key statistical concepts introduced in the chapter, quickly summarizing each concept or procedure and indicating when and how it should be used.
- An abundance of worked-out **examples** model solutions to real-world problems relevant to students' lives. Each example is tied to an end-of-chapter exercise so that students can practice solving a similar problem and test their understanding. Within the exercise sets, the icon TRY indicates which problems are tied to worked-out examples in that chapter, and the numbers of those examples are indicated.
- The **Chapter Review** that concludes each chapter provides a list of important new terms, student learning objectives, a summary of the concepts and methods discussed, and sources for data, articles, and graphics referred to in the chapter.

Active Learning

- For each chapter we've included an activity, **Exploring Statistics**, that students are intended to do in class as a group. We have used these activities ourselves, and we have found that they greatly increase student understanding and keep students engaged in class.
- All exercises are located at the end of the chapter. **Section Exercises** are designed to begin with a few basic problems that strengthen recall and assess basic knowledge, followed by mid-level exercises that ask more complex, open-ended questions. **Chapter Review Exercises** provide a comprehensive review of material covered throughout the chapter.

The exercises emphasize good statistical practice by requiring students to verify conditions, make suitable use of graphics, find numerical values, and interpret their findings in writing. All exercises are paired so that students can check their work on the odd-numbered exercise and then tackle the corresponding even-numbered exercise. The answers to all odd-numbered exercises appear in the back of the text.

Challenging exercises, identified with an asterisk (*), ask open-ended questions and sometimes require students to perform a complete statistical analysis. For exercises marked with a , accompanying data sets are available in MyStatLab and on the companion website.

- Most chapters include select exercises marked with a g within the exercise set, to indicate that problem-solving help is available in the **Guided Exercises** section. If students need support while doing homework, they can turn to the Guided Exercises to see a step-by-step approach to solving the problem.

Acknowledgments

We are grateful for the attention and energy that a large number of people devoted to making this a better book. We extend our gratitude to Elaine Newman (Sonoma State University) and Ann Cannon (Cornell College), who checked the accuracy of this text and its many exercises. Thanks also to David Chelton, our developmental editor, to Carol Merrigan, who handled production, to Peggy McMahon, project manager, and to Connie Day, our copyeditor. Many thanks to John Norbutas for his technical advice and help with the TechTips. We thank Suzanna Bainbridge, our acquisitions editor, Justin Billing, editorial assistant, and Deirdre Lynch, editor-in-chief, for signing us up and sticking with us, and we are grateful to Dona Kenly and Erin Kelly for their market development efforts.

We extend our sincere thanks for the suggestions and contributions made by the following reviewers of this edition:

Lloyd Best, *Pacific Union College*
Mario Borha, *Loyola University of Chicago*
David Bosworth, *Hutchinson Community College*
Beth Burns, *Bowling Green State University*
Jim Johnston, *Concord University*
Manuel Lopez, *Cerritos College*
Ralph Padgett Jr., *University of California – Riverside*
Ron Palcic, *Johnson County Community College*
Patrick Perry, *Hawaii Pacific University*
Victor I. Piercey, *Ferris State University*
Danielle Rivard, *Post University*
Alex Rolon, *North Hampton Community College*
Ali Saadat, *University of California – Riverside*
Carol Saltsgaver, *University of Illinois – Springfield*
Kelly Sakkinen, *Lake Land College*
Sharon l. Sullivan, *Catawba College*
Manuel Uy, *College of Alameda*
Mahbobeh Vezvaei, *Kent State University*
Jerimi Ann Walker, *Moraine Valley Community College*
Dottie Walton, *Cuyahoga Community College*
Judy H. Williams, *Tidewater Community College*
Jane Marie Wright, *Suffolk County Community College*
Cathleen Zucco-Teveloff, *Rider University*

We would also like to extend our sincere thanks for the suggestions and contributions that the following reviewers, class testers, and focus group attendees made to the previous edition.

Arun Agarwal, *Grambling State University*
Anne Albert, *University of Findlay*
Michael Allen, *Glendale Community College*
Eugene Allevato, *Woodbury University*
Dr. Jerry Allison, *Trident Technical College*
Polly Amstutz, *University of Nebraska*
Patricia Anderson, *Southern Adventist University*
MaryAnne Anthony-Smith, *Santa Ana College*
David C. Ashley, *Florida State College at Jacksonville*
Diana Asmus, *Greenville Technical College*
Kathy Autrey, *Northwestern State University of Louisiana*
Wayne Barber, *Chemeketa Community College*
Roxane Barrows, *Hocking College*
Jennifer Beineke, *Western New England College*
Diane Benner, *Harrisburg Area Community College*
Norma Biscula, *University of Maine, Augusta*
K.B. Boomer, *Bucknell University*
David Bosworth, *Hutchinson Community College*
Diana Boyette, *Seminole Community College*
Elizabeth Paulus Brown, *Waukesha County Technical College*
Leslie Buck, *Suffolk Community College*
R.B. Campbell, *University of Northern Iowa*
Stephanie Campbell, *Mineral Area College*
Ann Cannon, *Cornell College*
Rao Chaganty, *Old Dominion University*
Carolyn Chapel, *Western Technical College*
Christine Cole, *Moorpark College*
Linda Brant Collins, *University of Chicago*
James A. Condor, *Manatee Community College*
Carolyn Cuff, *Westminster College*
Phyllis Curtiss, *Grand Valley State University*
Monica Dabos, *University of California, Santa Barbara*
Greg Davis, *University of Wisconsin, Green Bay*
Bob Denton, *Orange Coast College*
Julie DePree, *University of New Mexico–Valencia*
Jill DeWitt, *Baker Community College of Muskegon*
Paul Drelles, *West Shore Community College*
Keith Driscoll, *Clayton State University*
Rob Eby, *Blinn College*
Nancy Eschen, *Florida Community College at Jacksonville*
Karen Estes, *St. Petersburg College*
Mariah Evans, *University of Nevada, Reno*
Harshini Fernando, *Purdue University North Central*
Stephanie Fitchett, *University of Northern Colorado*
Elaine B. Fitt, *Bucks County Community College*
Michael Flesch, *Metropolitan Community College*
Melinda Fox, *Ivy Tech Community College, Fairbanks*
Joshua Francis, *Defiance College*
Michael Frankel, *Kennesaw State University*
Heather Gamber, *Lone Star College*
Debbie Garrison, *Valencia Community College, East Campus*
Kim Gilbert, *University of Georgia*
Stephen Gold, *Cypress College*
Nick Gomersall, *Luther College*
Mary Elizabeth Gore, *Community College of Baltimore County–Essex*
Ken Grace, *Anoka Ramsey Community College*
Larry Green, *Lake Tahoe Community College*
Jeffrey Grell, *Baltimore City Community College*

Albert Groccia, *Valencia Community College, Osceola Campus*
David Gurney, *Southeastern Louisiana University*
Chris Hakenkamp, *University of Maryland, College Park*
Melodie Hallet, *San Diego State University*
Donnie Hallstone, *Green River Community College*
Cecil Hallum, *Sam Houston State University*
Josephine Hamer, *Western Connecticut State University*
Mark Harbison, *Sacramento City College*
Beverly J. Hartter, *Oklahoma Wesleyan University*
Laura Heath, *Palm Beach State College*
Greg Henderson, *Hillsborough Community College*
Susan Herring, *Sonoma State University*
Carla Hill, *Marist College*
Michael Huber, *Muhlenberg College*
Kelly Jackson, *Camden County College*
Bridgette Jacob, *Onondaga Community College*
Robert Jernigan, *American University*
Chun Jin, *Central Connecticut State University*
Maryann Justinger, Ed.D., *Erie Community College*
Joseph Karnowski, *Norwalk Community College*
Susitha Karunaratne, *Purdue University North Central*
Mohammed Kazemi, *University of North Carolina–Charlotte*
Robert Keller, *Loras College*
Omar Keshk, *Ohio State University*
Raja Khoury, *Collin County Community College*
Brianna Killian, *Daytona State College*
Yoon G. Kim, *Humboldt State University*
Greg Knofczynski, *Armstrong Atlantic University*
Jeffrey Kollath, *Oregon State University*
Erica Kwiatkowski-Egizio, *Joliet Junior College*
Sister Jean A. *Lanahan, OP, Molloy College*
Katie Larkin, *Lake Tahoe Community College*
Michael LaValle, *Rochester Community College*
Deann Leoni, *Edmonds Community College*
Lenore Lerer, *Bergen Community College*
Quan Li, *Texas A&M University*
Doug Mace, *Kirtland Community College*
Walter H. Mackey, *Owens Community College*
Keith McCoy, *Wilbur Wright College*
Elaine McDonald-Newman, *Sonoma State University*
William McGregor, *Rockland Community College*
Bill Meisel, *Florida State College at Jacksonville*
Bruno Mendes, *University of California, Santa Cruz*
Wendy Miao, *El Camino College*
Robert Mignone, *College of Charleston*
Ashod Minasian, *El Camino College*
Megan Mocko, *University of Florida*
Sumona Mondal, *Clarkson University*
Kathy Mowers, *Owensboro Community and Technical College*
Mary Moyinhan, *Cape Cod Community College*
Junalyn Navarra-Madsen, *Texas Woman's University*
Azarnia Nazanin, *Santa Fe College*
Stacey O. Nicholls, *Anne Arundel Community College*
Helen Noble, *San Diego State University*
Lyn Noble, *Florida State College at Jacksonville*
Keith Oberlander, *Pasadena City College*
Pamela Omer, *Western New England College*
Nabendu Pal, *University of Louisiana at Lafayette*
Irene Palacios, *Grossmont College*
Adam Pennell, *Greensboro College*
Joseph Pick, *Palm Beach State College*
Philip Pickering, *Genesee Community College*
Robin Powell, *Greenville Technical College*
Nicholas Pritchard, *Coastal Carolina University*
Linda Quinn, *Cleveland State University*
William Radulovich, *Florida State College at Jacksonville*
Mumunur Rashid, *Indiana University of Pennsylvania*
Fred J. Rispoli, *Dowling College*
Nancy Rivers, *Wake Technical Community College*
Corlis Robe, *East Tennesee State University*
Thomas Roe, *South Dakota State University*
Dan Rowe, *Heartland Community College*
Carol Saltsgaver, *University of Illinois–Springfield*
Radha Sankaran, *Passaic County Community College*
Delray Schultz, *Millersville University*
Jenny Shook, *Pennsylvania State University*
Danya Smithers, *Northeast State Technical Community College*
Larry Southard, *Florida Gulf Coast University*
Dianna J. Spence, *North Georgia College & State University*
René Sporer, *Diablo Valley College*
Jeganathan Sriskandarajah, *Madison Area Technical College–Traux*
David Stewart, *Community College of Baltimore County–Cantonsville*
Linda Strauss, *Penn State University*
John Stroyls, *Georgia Southwestern State University*
Joseph Sukta, *Moraine Valley Community College*
Lori Thomas, *Midland College*
Malissa Trent, *Northeast State Technical Community College*
Ruth Trygstad, *Salt Lake Community College*
Gail Tudor, *Husson University*
Manuel T. Uy, *College of Alameda*
Lewis Van Brackle, *Kennesaw State University*
Mahbobeh Vezvaei, *Kent State University*
Joseph Villalobos, *El Camino College*
Barbara Wainwright, *Sailsbury University*
Henry Wakhungu, *Indiana University*
Dottie Walton, *Cuyahoga Community College*
Jen-ting Wang, *SUNY, Oneonta*
Jane West, *Trident Technical College*
Michelle White, *Terra Community College*
Bonnie-Lou Wicklund, *Mount Wachusett Community College*
Sandra Williams, *Front Range Community College*
Rebecca Wong, *West Valley College*
Alan Worley, *South Plains College*
Jane-Marie Wright, *Suffolk Community College*
Haishen Yao, *CUNY, Queensborough Community College*
Lynda Zenati, *Robert Morris Community College*
Yan Zheng-Araujo, *Springfield Community Technical College*
Cathleen Zucco-Teveloff, *Rider University*
Mark A. Zuiker, *Minnesota State University,*

Acknowledgments for the Global Edition

Contributors
Vikas Arora, Statistician
Kiran Paul, Statistician

Reviewers
Santhosh Kumar, *Christ University*
Kiran Paul, Statistician
Chirag Trivedi, *R. J. Tibrewal Commerce College*

BREAKTHROUGH
To improving results

Resources for Success

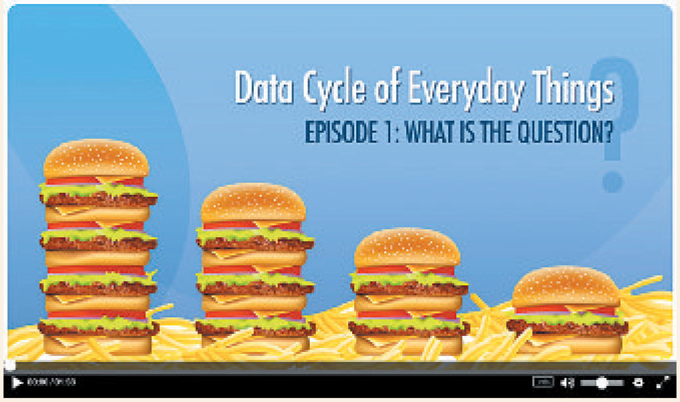
MyStatLab® Online Course for *Essential Statistics: Exploring the World through Data*, Second Edition, Gould/Ryan/Wong

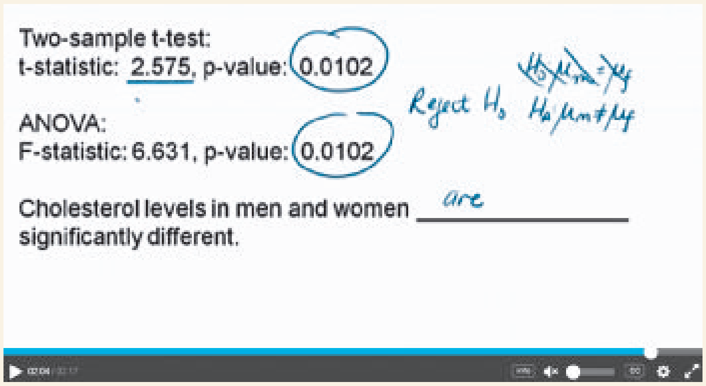
(access code required)

MyStatLab is available to accompany Pearson's market-leading text offerings. To give students a consistent tone, voice, and teaching method, each text's flavor and approach is tightly integrated throughout the accompanying MyStatLab course, making learning the material as seamless as possible.

Data Cycle of Everyday Things Videos

Data Cycle of Everyday Things Videos demonstrate for students that data can be, and is, a part of everyday life! Through a series of fun and engaging episodes, students learn to collect, analyze, and apply data to answer any range of real-world statistical questions.



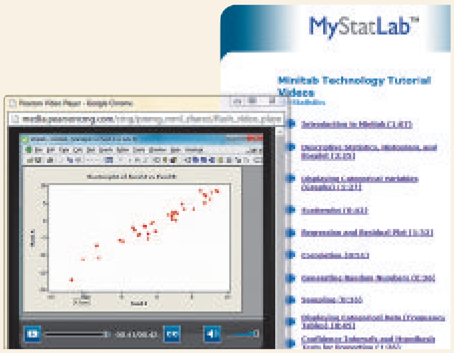


Chapter Review Videos

Chapter Review Videos walk students through solving a selection of the more complex problems posed in each chapter, providing a review of the chapter's key concepts and methods and offering students support where they most need it.

Technology Tutorial Videos and Study Cards

Technology Tutorials provide brief video walkthroughs and step-by-step instructional study cards on common statistical procedures for Minitab®, Excel®, and the TI-83/84 graphing calculator.



www.mystatlab.com

Resources for Success

Instructor Resources

Additional resources can be downloaded from **www.pearsonglobaleditions.com/gould** or MyStatLab.

Instructor's Solutions Manual

Instructor's Solutions Manual contains worked-out solutions to all the text exercises.

TestGen

TestGen® (**www.pearsoned.com/testgen**)
Updated to more closely mirror the 2nd Edition book, *TestGen* enables instructors to build, edit, print, and administer tests using a computerized bank of questions developed to cover all the objectives of the text.

Online Test Bank

Online Test Bank (download only) includes tests for each chapter with questions aimed to reinforce the text's learning objectives.

PowerPoint® Lecture Slides

PowerPoint Lecture Slides, aligned with the text, provide an overview of each chapter, stressing important definitions and offering additional examples. Multiple-choice questions are included for class assessment.

Learning Catalytics

Learning Catalytics is a web-based engagement and assessment tool. As a "bring-your-own-device" direct response system, Learning Catalytics offers a diverse library of dynamic question types that allow students to interact with and think critically about statistical concepts. As a real-time resource, instructors can take advantage of critical teaching moments both in the classroom or through assignable and gradeable homework.

Student Resources

Additional resources for student success via MyStatLab.

Study Cards for Statistics Software

Study Cards for Statistics Software This series of study cards, available for Excel, Minitab, JMP®, SPSS®, R, StatCrunch, and the TI-84 graphing calculators, provides students with easy, step-by-step guides to the most common statistics software.

PowerPoint Slides

PowerPoint Slides provide an overview of each chapter, stressing important definitions and offering additional examples. These slides are an excellent resource for both traditional and online students.

Chapter Review Videos

Chapter Review Videos walk students through solving a selection of the more complex problems posed in each chapter, thereby offering a review of the chapter's key concepts and providing students support where they most need it.

The Data Cycle of Everyday Things Videos

The Data Cycle of Everyday Things Videos demonstrate for students that data collection and data analysis can be applied to answer questions about everyday life. Through a series of fun and engaging episodes, students learn to collect, analyze, and apply data to answer any range of real-world statistical questions.

Index of Applications

LAW

POLITICS

PSYCHOLOGY

SOCIAL ISSUES

SPORTS

SURVEYS AND OPINION POLLS

TECHNOLOGY

TRANSPORTATION

Essential Statistics: Exploring the World through Data

Second Edition

1 Introduction to Data

THEME

Statistics is the science of data, so we must learn the types of data we will encounter and the methods for collecting data. The method used to collect data is very important because it determines what types of conclusions we can reach and, as you'll learn in later chapters, what types of analyses we can do. By organizing the data we've collected, we can often spot patterns that are not otherwise obvious.

This text will teach you to examine data to better understand the world around you. If you know how to sift data to find patterns, can communicate the results clearly, and understand whether you can generalize your results to other groups and contexts, you will be able to make better decisions, offer more convincing arguments, and learn things you did not know before. Data are everywhere, and making effective use of them is such a crucial task that one prominent economist has proclaimed statistics one of the most important professions of the decade (*McKinsley Quarterly* 2009).

The use of statistics to make decisions and convince others to take action is not new. Some statisticians date the current practice of statistics back to the mid-nineteenth century. One famous example occurred in 1854, when the British were fighting the Russians in the brutal Crimean War. A British newspaper had criticized the military medical facilities, and a young but well-connected nurse, Florence Nightingale, was appointed to study the situation and, if possible to improve it.

Nightingale carefully recorded the numbers of deaths, the causes of the deaths, and the times and dates of the deaths. She organized these data graphically, and these graphs enabled her to see a very important pattern: A large percentage of deaths were due to contagious disease, and many deaths could be prevented by improving sanitary conditions. Within six months, Nightingale had reduced the death rate by half. Eventually she convinced Parliament and military authorities to completely reorganize the medical care they provided. Accordingly, she is credited with inventing modern hospital management.

In modern times, we have equally important questions to answer. Do cell phones cause brain tumors? Are alcoholic drinks healthful in moderation? Which diet works best for losing weight? What percentage of the public is concerned about job security? **Statistics**—the science (and art!) of collecting and analyzing observations to learn about ourselves, our surroundings, and our universe—helps answer questions such as these.

Data are the building blocks of statistics. This chapter introduces some of the basic types of data and explains how we collect them, store them, and organize them. These ideas and skills will provide a basic foundation for your study of the rest of the text.

CASE STUDY

Deadly Cell Phones?

In September 2002, Dr. Christopher Newman, a resident of Maryland, sued Motorola, Verizon, and other wireless carriers, accusing them of causing a cancerous brain tumor behind his right ear. As evidence, his lawyers cited a study by Dr. Lennart Hardell. Hardell had studied a large number of people with brain tumors and had found that a greater percentage of them used cell phones than of those who did not have brain tumors (CNN 2002; Brody 2002).

Speculation that cell phones might cause brain cancer began as early as 1993, when (as CNN reports) the interview show *Larry King Live* featured a man who claimed that his wife died because of cancer caused by her heavy cell phone use. However, more recent studies have contradicted Hardell's results, as well as earlier reports about the health risks of heavy cell phone use.

The judge in Dr. Newman's trial was asked to determine whether Hardell's study was compelling enough to support allowing the trial to proceed. Part of this determination involved evaluating the method that Hardell used to collect data. If you were the judge, how would you rule? You will learn the judge's ruling at the end of the chapter. You will also see how the methods used to collect data about important *cause-and-effect relationships*—such as that which Dr. Newman alleged to exist between cell phone use and brain cancer—can affect the conclusions we can draw.

SECTION 1.1

What Are Data?

The study of statistics rests on two major concepts: variation and data. **Variation** is the more fundamental of these concepts. To illustrate this idea, draw a circle on a piece of paper. Now draw another one, and try to make it look just the same. Now another. Are all three exactly the same? We bet they're not. They might be slightly different sizes, for instance, or slightly different versions of round. This is an example of variation. How can you reduce this variation? Maybe you can get a penny and outline the penny. Try this three times. Does variation still appear? Probably it does, even if you need a magnifying glass to see, say, slight variations in the thickness of the penciled line.

Data are observations that you or someone else records. The drawings in Figure 1.1 are data that record our attempts to draw three circles that look the same. Analyzing pictorial data such as these is not easy, so we often try to quantify such observations—that is, to turn them into numbers. How would you measure whether these three circles are the same? Perhaps you would compare diameters or circumferences, or somehow try to measure how and where these circles depart from being perfect circles. Whatever technique you chose, these measurements could also be considered data.

Data are more than just numbers, though. David Moore, a well-known statistician, defined data as "numbers in context." By this he meant that data consist not only of the numbers we record, but also of the story behind the numbers. For example,

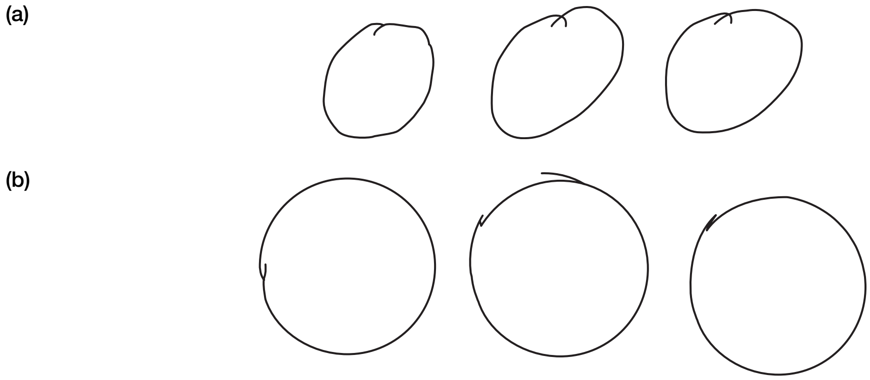
10.00, 9.88, 9.81, 9.81, 9.75, 9.69, 9.5, 9.44, 9.31

are just numbers. But in fact these numbers represent "Weight in pounds of the ten heaviest babies in a sample of babies born in North Carolina in 2004." Now these numbers have a context and have been elevated into data. See how much more interesting data are than numbers?

Details

Data Are What Data Is

If you want to be "old school" grammatically correct, then the word *data* is plural. So we say "data *are*" and not "data *is*." The singular form is *datum*. However, this usage is changing over time, and some dictionaries now say that *data* can be used as both a singular and a plural noun.



▶ **FIGURE 1.1** **(a)** Three circles drawn by hand. **(b)** Three circles drawn using a coin. It is clear that the circles drawn by hand show more variability than the circles drawn with the aid of a coin.

These data were collected by the state of North Carolina in part to help researchers understand the factors that contribute to low-weight and premature births. If doctors understand the causes of premature birth, they can work to prevent it—perhaps by helping expectant mothers change their behavior, perhaps by medical intervention, and perhaps by a combination of both.

KEY POINT Data are "numbers in context."

In the last few years, our culture and economy have been inundated with data. The magazine *The Economist* has called this surge of data the "data deluge." One reason for the rising tide of data is the application of automated data collection devices. These range from automatic sensors that simply record everything they see and store the data on a computer, to websites and smart phone apps that record every transaction their users make. Google, for example, saves every search you make and combines this with data on which links you click in order to improve the way it presents information (and also, of course, to determine which advertisements will appear on your search results).

Thanks to small, portable sensors, you can now join the "Personal Data Movement." Members of this movement record data about their daily lives and analyze it in order to improve their health, to run faster, or just to make keepsakes—a modern-day scrapbook. Maybe you or a friend uses a Nike Fuel Band to keep track of regular runs. One of the authors of this text carries a FitBit in his pocket to record his daily activity. From this he learned that on days he lectures, he typically takes 7600 steps, and on days that he does not lecture, he typically only takes 4900 steps. Some websites, such as your.flowingdata.com, make use of Twitter to help users collect, organize, and understand whatever personal data they choose to record.

Of course, it is not only machines that collect data. Humans still actively collect data with the intent of better understanding some phenomenon or making a discovery. Marketers prepare focus groups and surveys to describe the market for a new product. Sports analysts collect data to help their teams' coaches win games, or to help fantasy football league players. Scientists perform experiments to test theories and to measure changes in the economy or the climate. In this text you'll learn about the many ways in which data are used.

The point is that we have reached a historical moment where almost everything can be thought of as data. And once you find a way of capturing data about something in your world, you can organize, sort, visualize, and analyze those data to gain deeper understanding about the world around you.

What Is Data Analysis?

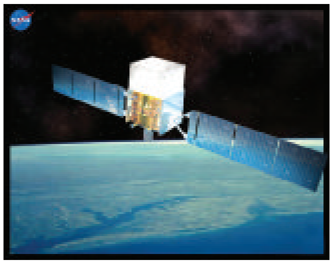
In this text you will study the science of data. Most important, you will learn to analyze data. What does this mean? You are analyzing data when you examine data of some sort and explain what they tell us about the real world. In order to do this, you must first learn about the different types of data, how data are stored and structured, and how they are summarized. The process of summarizing data takes up a big part of this text; indeed, we could argue that the entire text is about summarizing data, either through creating a visualization of the data or distilling them down to a few numbers that we hope capture their essence.

KEY POINT Data analysis involves creating summaries of data and explaining what these summaries tell us about the real world.

SECTION 1.2

Classifying and Storing Data

▲ **FIGURE 1.2** A photo of Carhenge, Nebraska.



▲ **FIGURE 1.3** Satellites in NASA's Earth Observing Mission record ultraviolet reflections and transmit these data back to Earth. Such data are used to construct images of our planet. Earth Observer (http://eos.gsfc.nasa.gov/).

The first step in understanding data is to understand the different types of data you will encounter. As you've seen, data are numbers in context. But that's only part of the story; data are also recorded observations. Your photo from your vacation to Carhenge in Nebraska is data (Figure 1.2). The ultraviolet images streaming from the Earth Observer Satellite system are data (Figure 1.3). These are just two examples of data that are not numbers. Statisticians work hard to help us analyze complex data, such as images and sound files, just as easily as we study numbers. Most of the methods involve recoding the data into numbers. For example, your photos can be digitized in a scanner, converted into a very large set of numbers, and then analyzed. You might have a digital camera that gives you feedback about the quality of a photo you've taken. If so, your camera is not only collecting data but also analyzing it!

Almost always, our data sets will consist of characteristics of people or things (such as gender and weight). These characteristics are called **variables**. Variables are not "unknowns" like those you studied in algebra. We call these characteristics variables because they have variability: The values of the variable can be different from person to person.

KEY POINT Variables in statistics are different from variables in algebra. In statistics, variables record characteristics of people or things.

When we work with data, they are grouped into a collection, which we call either a **data set** or a **sample**. The word *sample* is important, because it implies that the data we see are just one part of a bigger picture. This "bigger picture" is called a **population**. Think of a population as the Data Set of Everything—it is the data set that contains all of the information about everyone or everything with respect to whatever variable we are studying. Quite often, the population is really what we want to learn about, and we learn about it by studying the data in our sample. However, many times it is enough just to understand and describe the sample. For example, you might collect data from students in your class simply because you want to know about the students in your class, and not because you wish to use this information to learn about all students at your school. Sometimes, data sets are so large that they effectively *are* the population, as you'll soon see in the data reflecting births in North Carolina.

Details

More Grammar

We're using the word *sample* as a noun—it is an object, a collection of data that we study. Later we'll also use the word *sample* as a verb—that is, to describe an action. For example, we'll sample ice cream cones to measure their weight.

Two Types of Variables

The variables you'll find in your data set come in two basic types, which can themselves be broken into smaller divisions, as we'll discuss later.

Numerical variables describe quantities of the objects of interest. The values will be numbers. The weight of an infant is an example of a numerical variable.

Categorical variables describe qualities of the objects of interest. These values will be categories. The sex of an infant is an example of a categorical variable. The possible values are the categories "male" and "female." Eye color of an infant is another example; the categories might be brown, blue, black, and so on. You can often identify categorical variables because their values are *usually* words, phrases, or letters. (We say "usually" because we sometimes use numbers to represent a word or phrase. Stay tuned.)

Details

Quantitative and Qualitative Data

Some statisticians use the word *quantitative* to refer to numerical variables (think "quantity") and *qualitative* to refer to categorical variables (think "quality"). We prefer *numerical* and *categorical*. Both sets of terms are commonly used, and you should be prepared to hear and see both.

EXAMPLE 1 Crash-Test Results

The data in Table 1.1 are an excerpt from crash-test dummy studies in which cars are crashed into a wall at 35 miles per hour. Each row of the data set represents the observed characteristics of a single car. This is a small sample of the database, which is available from the National Transportation Safety Administration. The *head injury* variable reflects the risk to the passengers' heads. The higher the number, the greater the risk.

Make	Model	Doors	Weight	Head Injury
Acura	Integra	2	2350	599
Chevrolet	Camaro	2	3070	733
Chevrolet	S-10 Blazer 4X4	2	3518	834
Ford	Escort	2	2280	551
Ford	Taurus	4	2390	480
Hyundai	Excel	4	2200	757
Mazda	626	4	2590	846
Volkswagen	Passat	4	2990	1182
Toyota	Tercel	4	2120	1138

◄ **TABLE 1.1** Crash-test results for cars.

QUESTION For each variable, state whether it is numerical or categorical.

SOLUTION The variables *make* and *model* are categorical. Their values are descriptive names. The units of *doors* are, quite simply, the number of doors. The units of *weight* are pounds. The variables *doors* and *weight* are numerical because their values are measured quantities. The units for *head injury* are unclear; head injury is measured using some scale that the researchers developed.

TRY THIS! Exercise 1.3

Details

Categorical Doors
Some people might consider *doors* a categorical variable, because nearly all cars have either 2 doors or 4 doors, and for many people, the number of doors designates a certain type of car (small or larger). There's nothing wrong with that.

Coding Categorical Data with Numbers

Sometimes categorical variables are "disguised" as numerical. The smoke variable in the North Carolina data set (Table 1.2) has numbers for its values (0 and 1), but in fact those numbers simply indicate whether or not the mother smoked. Mothers were asked, "Did you smoke?" and if they answered "Yes," the researchers coded this categorical response with a 1. If they answered "No," the response was coded with a 0. These particular numbers represent categories, not quantities. *Smoke* is a categorical variable.

Coding is used to help both humans and computers understand what the values of a variable represent. For example, a human would understand that a "yes" under the "Smoke" column would mean that the person was a smoker, but to the computer, "yes" is just a string of symbols. If instead we follow a convention where a 1 means "yes" and a 0 means "no," then a human understands that the 1's represent smokers, and a computer can easily add the values together to determine, for example, how many smokers are in the sample.

Caution

Don't Just Look for Numbers!
You can't always tell whether a variable is categorical simply by looking at the data table. You must also consider what the variable represents. Sometimes, researchers code categorical variables with numerical values.

Weight	Female	Smoke
7.69	1	0
0.88	0	1
6.00	1	0
7.19	1	0
8.06	1	0
7.94	1	0

▲ **TABLE 1.2** Data for newborns with coded categorical variables.

This approach for coding categorical variables is quite common and useful. If a categorical variable has only two categories, as do *gender* and *smoke*, then it is almost always helpful to code the values with 0 and 1. To help readers know what a "1" means, rename the variable with either one of its category names. A "1" then means the person belongs to that category, and a 0 means the person belongs to the other category. For example, instead of calling a variable *gender*, we rename it *female*. And then if the baby is a boy we enter the code 0, and if it's a girl we enter the code 1.

Sometimes your computer does the coding for you without your needing to know anything about it. So even if you see the words *female* and *male* on your computer, the computer has probably coded these with values of 0 and 1 (or vice versa).

Details

Numerical Categories

Categories might be numbers. Sometimes, numerical variables are coded as categories, even though we wish to use them as numbers. For example, *number of siblings* might be coded as "none," "one," "two," "three," etc. Although words are used, this is really a numerical variable since it is counting something.

Storing Your Data

The format in which you record and store your data is very important. Computer programs will require particular formats, and by following a consistent convention, you can be confident that you'll better remember the qualities of your own data set if you need to revisit it months or even years later. Data are often stored in a spreadsheet-like format in which each row represents the object (or person) of interest. Each column represents a variable. In Table 1.2, each row represents a baby. The column heads are variables: *Weight*, *Female*, and *Smoke*. This format is sometimes referred to as the **stacked data** format.

When you collect your own data, the stacked format is almost always the best way to record and store your data. One reason is that it allows you to easily record several different variables for each subject. Another reason is that it is the format that most software packages will assume you are using for most analyses. (The exceptions are TI-84 and Excel.)

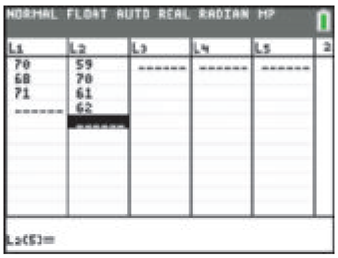
Some technologies, such as the TI calculators, require, or at least accommodate, data stored in a different format, called **unstacked data**. Unstacked data tables are also common in some books and media publications. In this format, each column represents a variable from a different group. For example, one column could represent men's heights, and another column could represent women's heights. The data set, then, is a single variable (*height*) broken into two groups. The groups are determined by a categorical variable. Table 1.3 shows an example of unstacked data, and Figure 1.4 shows the same data in TI-84 input format.

By way of contrast, Table 1.4 shows the same data in stacked format.

The great disadvantage of the unstacked format is that it can store only two variables at a time: the variable of interest (for example, height), and a categorical variable that tells us which group the observation belongs in (for example, gender). However, most of the time, we record many variables for each observation. For example, we record a baby's weight, gender, and whether or not the mother smoked. The stacked format enables us to display as many variables as we wish.

Men's Heights	Women's Heights
70	59
68	70
71	61
	62

▲ **TABLE 1.3** Data by groups (unstacked).



▲ **FIGURE 1.4** TI-84 data input screen (unstacked data).

EXAMPLE 2 Personal Data Collection

Using a sensor worn around her wrist, Safaa recorded the amount of sleep she got on several nights. She also recorded whether it was a weekend or a weeknight. For the weekends, she recorded (in hours): 8.1, 8.3. For the weeknights she recorded 7.9, 6.5, 8.2, 7.0, 7.3.

QUESTION Write these data in both the stacked format and the unstacked format.

SOLUTION In the stacked format, each row represents a unit of observation, and each column measures a characteristic of that observation. For Safaa, the unit of observation was a night of sleep, and she measured two characteristics: time and whether or not it was a weekend. In stacked format, her data would look like this:

Time	Weekend
8.1	Yes
8.3	Yes
7.9	No
6.5	No
8.2	No
7.0	No
7.3	No

Height	Gender
70	male
68	male
71	male
59	female
70	female
61	female
62	female

▲ **TABLE 1.4** The same data as in Table 1.3, shown here in stacked format.

(Note that you might have coded the "Weekend" variable differently. For example, instead of entering "Yes" or "No," you might have written either "Weekend" or "Weeknight" in each row.)

In the unstacked format, the numerical observations appear in separate columns, depending on the value of the categorical variable:

Weekend	Weeknight
8.1	7.9
8.3	6.5
	8.2
	7.0
	7.3

See the Tech Tips to review how to enter data like these using your technology.

Caution

Look at the Data Set!
The fact that different people use different formats to store data means that the first step in any data investigation is to look at the data set. In most real-life situations, stacked data are the more useful format, because this format permits you to work with several variables at the same time.

TRY THIS! Exercise 1.11

Context Is Key

The context is the most important aspect of data, although it is frequently overlooked. Table 1.5 shows a few lines from the data set of births in 2004 in North Carolina (Holcomb 2006).

To understand these data, we need to ask and try to answer some questions in order to better understand the context: Who, or what, was observed? What variables were measured? How were they measured? What are the units of measurement? Who collected the data? How did they collect the data? Where were the data collected? Why were the data collected? When were the data collected?

Many, but not all, of these questions can be answered for these data by reading the information provided on the website that hosts the data. Other times we are not so lucky and must rely on very flimsy supporting documentation. If you collect the data yourself, you should be careful to record this extra supporting information. Or, if you get a chance to talk with the people who collected the data, then you should ask them these questions.

Weight	Gender	Smoke
7.69	F	0
0.88	M	1
6.00	F	0
7.19	F	0
8.06	F	0
7.94	F	0

▲ **TABLE 1.5** Birth data from North Carolina in 2004.

- ***Who, or what, was observed?*** In this data set, we observed babies. Each line in the table represents a newborn baby born in North Carolina in 2004. If we were to see the whole table, we would see a record of every baby born in 2004 in North Carolina.

- ***What variables were measured?*** For each baby, the state records the weight, the gender, and whether the mother smoked.
- ***How were the variables measured?*** Unknown. Presumably, most measurements on the baby were taken from a medical caregiver at the time of the birth, but we don't know how or when information about the mother was collected.
- ***What are the units of measurement?*** Units of measurement are important. The same variable can have different units of measurement. For example, weight could be measured in pounds, in ounces, or in kilograms. For Table 1.5,

  Weight: reported in pounds.
  Gender: reported as M for boys and F for girls.
  Smoke: reported as a 1 if the mother smoked during the pregnancy, as a 0 if she did not.
- ***Who collected the data?*** The government of the state of North Carolina.
- ***How did they collect the data?*** Data were recorded for *all* births that occurred in hospitals in North Carolina. Later in the chapter you'll see that data can be collected by drawing a random sample of subjects, or by assigning subjects to receive different treatments, as well as through other methods. The exact method used for Table 1.5 is not clear, but the data were probably compiled from publicly available medical records and from reports by the physicians and caregivers.
- ***Where were the data collected?*** The location where the data were collected often gives us information about who (or what) the study is about. These data were collected in North Carolina and consist of babies born in that state. We should therefore be very wary about generalizing our findings to other states or other countries.
- ***Why were the data collected?*** Sometimes, data are collected to learn about a larger population. At other times, the goals are limited to learning more about the sample itself. In this case the data consist of all births in North Carolina, and it is most likely that researchers wanted to learn how the health of infants was related to the smoking habits of mothers within this sample.
- ***When were the data collected?*** The world is always changing, and so conclusions based on a data set from 1980 might be different from conclusions based on data collected for a similar study in 2015. These data were collected in 2004.

KEY POINT The first time you see a data set, ask yourself these questions:

- Who, or what, was observed?
- What variables were measured?
- How were the variables measured?
- What are the units of measurement?
- Who collected the data?
- How did they collect the data?
- Where were the data collected?
- Why did they collect the data?
- When were the data collected?

SECTION 1.3

Organizing Categorical Data

Once we have a data set, we next need to organize and display the data in a way that helps us see patterns. This task of organization and display is not easy, and we discuss it throughout the entire text. In this section we introduce the topic for the first time, in the context of categorical variables.

With categorical variables, we are usually concerned with knowing how often a particular category occurs in our sample. We then (usually) want to compare how often a category occurs for one group with how often it occurs for another (liberal/conservative, man/woman). To do these comparisons, you need to understand how to calculate percentages and other rates.

A common method for summarizing two potentially related categorical variables is to use a two-way table. **Two-way tables** show how many times each combination of categories occurs. For example, Table 1.6 is a two-way table from the Youth Behavior Risk Survey that shows gender and whether or not the respondent always (or almost always) wears a seat belt when riding in or driving a car. The actual Youth Behavior Risk Survey has over 10,000 respondents, but we are practicing on a small sample from this much larger data set.

The table tells us that 2 people were male and did not always wear a seat belt. Three people were female and did not always wear a seat belt. These counts are also called frequencies. A **frequency** is simply the number of times a value is observed in a data set.

Some books and publications discuss two-way tables as if they displayed the original data collected by the investigators. However, two-way tables do not consist of "raw" data but, rather, are summaries of data sets. For example, the data set that produced Table 1.6 is shown in Table 1.7.

To summarize this table, we simply count how many of the males (a 1 in the Male column) also do not always wear seat belts (a 1 in the Not Always column). We then count how many both are male and always wear seat belts (a 1 in the Male column, a 0 in the Not Always column); how many both are female and don't always wear seat belts (a 0 in the Male column, a 1 in the Not Always column); and finally, how many both are female and always wear a seat belt (a 0 in the Male column, a 0 in the Not Always column).

Example 3 illustrates that summarizing the data in a two-way table can make it easy to compare groups.

	Male	Female
Not Always	2	3
Always	3	7

▲ **TABLE 1.6** This two-way table shows counts for 15 youths who responded to a survey about wearing seat belts.

Male	Not Always
1	1
1	1
1	0
1	0
1	0
0	1
0	1
0	1
0	0
0	0
0	0
0	0
0	0
0	0
0	0

▲ **TABLE 1.7** This data set is equivalent to the two-way summary shown in Table 1.6. We highlighted in red those who did not always wear a seat belt (the risk takers).

EXAMPLE 3 Percentages of Seat Belt Wearers

The 2011 Youth Behavior Risk Survey is a national study that asks American youths about potentially risky behaviors. We show the two-way summary again. All of the people in the table were between 14 and 17 years old. The participants were asked whether they wear a seat belt while driving or riding in a car. The people who said always or almost always were put in the Always group. The people who said sometimes or rarely were put in the Not Always group.

	Male	Female
Not Always	2	3
Always	3	7

QUESTIONS

a. How many men are in this sample? How many women? How many people do not always wear seat belts? How many always wear seat belts?

b. What percent of the sample are men? What percent are women? What percent don't always wear seat belts? What percent always wear seat belts?

c. Are the men in the sample more likely than the women in the sample to take the risk of not wearing a seat belt?

! Caution

Two-way Tables Summarize Categorical Variables
It is tempting to look at a two-way table like Table 1.6 and think that you are looking at numerical variables, because you see numbers. But the values of the variables are actually categories (gender and whether or not the subject always wears a seat belt). The numbers you see are summaries of the data.

SOLUTIONS

a. We can count the men by adding the first column: $2 + 3 = 5$ men. Adding the second column gives us the number of women: $3 + 7 = 10$.

   We get the number who do not always wear seat belts by adding the first row: $2 + 3 = 5$ people don't always wear seat belts. Adding the second row gives us the number who always wear seat belts: $3 + 7 = 10$.

b. This question asks us to convert the numbers we found in part (a) to percentages. To do this, we divide the numbers by 15, because there were 15 people in the sample. To convert to percentages, we multiply this proportion by 100%.

   The proportion of men is $5/15 = 0.333$. The percentage is $0.333 \times 100\% = 33.3\%$. The proportion of women must be $100\% - 33.3\% = 66.7\%$ ($10/15 \times 100\% = 66.7\%$).

   The proportion who do not always wear seat belts is $5/15 = 0.333$, or 33.3%. The proportion who always wear seat belts is $100\% - 33.3\% = 66.7\%$.

c. You might be tempted to answer this question by counting the number of males who don't always wear seat belts (2 people) and comparing that to the number of females who don't always wear seat belts (3 people). However, this is not a fair comparison because there are more females than males in the sample. Instead, we should look at the percentage of those who don't always wear seat belts in each group. This question should be reworded as follows:

   Is the percentage of males who don't always wear seat belts greater than the percentage of females who don't always wear seat belts?

   Because 2 out of 5 males don't always wear seat belts, the percent of males who don't always wear seat belts is $(2/5) \times 100\% = 40\%$.

   Because 3 out of 10 females don't always wear seat belts, the percent of females who don't always wear seat belts is $(3/10) \times 100\% = 30\%$.

   In fact, females in this sample engage in this risky behavior less often than males. Among all U.S. youth, it is estimated that about 28% of males do not always wear their seat belt, compared to 23% of females.

TRY THIS! Exercise 1.15

The calculations in Example 3 took us from frequencies to percentages. Sometimes, we want to go in the other direction. If you know the total number of people in a group, and are given the percentage that meets some qualification, you can figure out *how many* people in the group meet that qualification.

EXAMPLE 4 Numbers of Seat Belt Wearers

A statistics class has 300 students, and they are asked whether they always ride or drive with a seat belt.

QUESTIONS

a. Suppose that 30% of the students do not always wear a seat belt. How many students is this?

b. Suppose we know that in another class, 20% of the students do not always wear seat belts, and this is equal to 43 students. How many students are in the class?

SOLUTIONS

a. We need to find 30% of 300. When working with percentages, first convert the percentage to its decimal equivalent:

$$30\% \text{ of } 300 = 0.30 \times 300 = 90$$

Therefore, 90 students don't always wear seat belts.

b. The question tells us that 20% of some unknown larger number (call it y) must be equal to 43.

$$0.20y = 43$$

Divide both sides by 0.20 and you get

$$y = 215$$

There are 215 total students in the class, and 43 of them don't always wear seat belts.

TRY THIS! Exercise 1.17

Sometimes, you may come across data summaries that are missing crucial information. Suppose we wanted to know which team sports are the most dangerous to play. Table 1.8 shows the number of sports-related injuries that were treated in U.S. emergency rooms in 2009 (National Safety Council 2011). (Note that this table is not the table of original data but is, instead, a summary of the original data.)

Wow! It's a dangerous world out there. Which would you conclude is the most dangerous sport? Which is the least dangerous?

Did you answer that basketball was the most dangerous sport? It did have the most injuries (501,251)—in fact, 50,000 more injuries than in football (451,961). Ice hockey is known for its violence (you've heard the old joke, "I went to a fight and suddenly a hockey match broke out"), but here, it seems to have caused relatively few injuries and looks safe.

The problem with comparing the numbers of injuries for these sports is that the sports have different numbers of participants. Injuries might be more common in basketball simply because more people play basketball. Also, there might be relatively few injuries in ice hockey merely because fewer people play. One important component is missing in Table 1.8, and the lack of this component makes our analysis impossible.

Table 1.9 includes the component missing from Table 1.8: the number of participants in each sport. We can't directly compare the number of injuries from sport to sport, because the numbers of members of the various groups are not the same. This improved table shows us the total membership of each group.

Sport	Injuries
Baseball	165,842
Basketball	501,251
Bowling	20,878
Football	451,961
Ice hockey	19,035
Soccer	208,214
Softball	121,175
Tennis	23,611
Volleyball	60,159

▲ **TABLE 1.8** Summary of counts of sports injuries.

Sport	Participants	Injuries
Baseball	11,500,000	165,842
Basketball	24,400,000	501,251
Bowling	45,000,000	20,878
Football	8,900,000	451,961
Ice hockey	3,100,000	19,035
Soccer	13,600,000	208,214
Softball	11,800,000	121,175
Tennis	10,800,000	23,611
Volleyball	10,700,000	60,159

◀ **TABLE 1.9** Summary of counts of sports injuries and numbers of participants.

Which sport is the most dangerous? We now have the information we need to answer this question. Specifically, we can find the percentage of participants injured in each sport. For example, what percent of basketball players were injured? There were 24,400,000 participants and 501,251 were injured, so the percent injured is $(501{,}251/24{,}400{,}000) \times 100\% = 2.05\%$.

Sometimes, with percentages as small as this, we understand the numbers more easily if we report not a percentage, but "number of events per 1000 objects" or maybe even "per 10,000 objects." We call such numbers **rates**. To get the injury rate per 1000 people, instead of multiplying $(501{,}251/24{,}400{,}000)$ by 100 we multiply by 1000: $(501{,}251/24{,}400{,}000) \times 1000 = 20.54$ injuries per 1000 people.

These results are shown in Table 1.10.

▶ **TABLE 1.10** Summary of rates of sports injuries.

Sport	Participants	Injuries	Rate of Injury per Participant	Rate of Injury per Thousand Participants
Baseball	11,500,000	165,842	0.01442	14.42
Basketball	24,400,000	501,251	0.02054	20.54
Bowling	45,000,000	20,878	0.00046	0.46
Football	8,900,000	451,961	0.05078	50.78
Ice hockey	3,100,000	19,035	0.00614	6.14
Soccer	13,600,000	208,214	0.01531	15.31
Softball	11,800,000	121,175	0.01027	10.27
Tennis	10,800,000	23,611	0.00219	2.19
Volleyball	10,700,000	60,159	0.00562	5.62

We see now that football is the most dangerous sport: 50.78 players are injured out of every 1000 players. Basketball is less risky, with 20.54 injuries per 1000 players.

EXAMPLE 5 Comparing Rates of Stolen Cars

Which model of car has the greatest risk of being stolen? The Highway Loss Data Institute reports that the Ford F-250 pickup truck is the most stolen car; 7 F-250's are reported stolen out of every 1000 that are insured. By way of contrast, the Jeep Compass is the least stolen; only 0.5 Jeep Compass is reported stolen for every 1000 insured (Insurance Institute for Highway Safety 2013).

QUESTION Why does the Highway Loss Data Institute report theft rates rather than the number of each type of car stolen?

SOLUTION We need to take into account the fact that some cars are more popular than others. Suppose there were many more Jeep Compasses than Ford F-250's. In that case, we might see a greater number of stolen Jeeps, simply because there are more of them to steal. By looking at the *theft rate*, we adjust for the total number of cars of that particular kind on the road.

TRY THIS! Exercise 1.29

KEY POINT In order for us to compare groups, the groups need to be similar. When the data consist of counts, then percentages or rates are often better for comparisons because they take into account possible differences among the sizes of the groups.

SECTION 1.4

Collecting Data to Understand Causality

Often, the most important questions in science, business, and everyday life are questions about **causality**. These are usually phrased in the form of "what if" questions. What if I take this medicine; will I get better? What if I change my Facebook profile; will my profile get more hits?

Questions about causality are often in the news. The *Los Angeles Times* reported that many people believe a drink called peanut milk can cure gum disease and slow the onslaught of baldness. The BBC News (2010) reported that "Happiness wards off heart disease." Statements such as these are everywhere we turn these days. How do we know whether to believe these claims?

The methods we use to collect data determine what types of conclusions we can make. Only one method of data collection is suitable for making conclusions about causal relationships, but as you'll see, that doesn't stop people from making such conclusions anyway. In this section we talk about three methods commonly used to collect data in an effort to answer questions about causality: anecdotes, observational studies, and controlled experiments.

Most questions about causality can be understood in terms of two variables: the **treatment variable** and the **outcome variable**. (The outcome variable is also sometimes called the **response variable**, because it responds to changes in the treatment.) We are essentially asking whether the treatment variable causes changes in the outcome variable. For example, the treatment variable might record whether or not a person drinks Peanut Milk, and the outcome variable might record whether or not that person's gum disease improved. Or the treatment variable might record whether or not a person is generally happy, and the outcome variable might record whether or not that person suffered from heart disease in a ten-year period.

People who receive the treatment of interest (or have the characteristic of interest) are said to be in the **treatment group**. Those who do not receive that treatment (or do not have that characteristic) are in the **comparison group**, which is also called the **control group.**

Anecdotes

Peanut milk is a drink invented by Jack Chang, an entrepreneur in San Francisco, California. He noticed that after he drank peanut milk for a few months, he stopped losing hair and his gum disease went away. According to the *Los Angeles Times* (Glionna 2006), another regular drinker of peanut milk says that the beverage caused his cancer to go into remission. Others have reported that drinking the beverage has reduced the severity of their colds, has helped them sleep, and has helped them wake up.

This is exciting stuff! Peanut milk could very well be something we should all be drinking. But can peanut milk really solve such a wide variety of problems? On the face of it, it seems that there's evidence that peanut milk has cured people of illness. The *Los Angeles Times* reports the names of people who claim that it has. However, the truth is that this is simply not enough evidence to justify any conclusion about whether the beverage is helpful, harmful, or without any effect at all.

These testimonials are examples of anecdotes. An **anecdote** is essentially a story that someone tells about her or his own (or a friend's or relative's) experience. Anecdotes are an important type of evidence in criminal justice because eyewitness testimony can carry a great deal of weight in a criminal investigation. However, for answering questions about groups of people with great variability or diversity, anecdotes are essentially worthless.

The primary reason why anecdotes are not useful for reaching conclusions about cause-and-effect relationships is that the most interesting things that we study have so much variety that a single report can't capture the variety of experience. For example, have you ever bought something because a friend recommended it, only to find that after a few weeks it fell apart? If the object was expensive, such as a car, you might have been angry at your friend for recommending such a bad product. But how do you know whose experience was more typical, yours or your friend's? Perhaps the car is in fact a very reliable model, and you just got a lemon.

A very important question to ask when someone claims that a product brings about some kind of change is to ask, "Compared to what?" Here the claim is that drinking peanut milk will make you healthier. The question to ask is "Healthier compared to what?" Compared to people who don't drink peanut milk? Compared to people who take medicine for their particular ailment? To answer these questions, we need to examine the health of these other groups of people who do not drink peanut milk.

Anecdotes do not give us a comparison group. We might know that a group of people believe that peanut milk made them feel better, but we don't know how the milk drinkers' experiences compare to those of people who did not drink peanut milk.

KEY POINT When someone makes a claim about causality, ask, "Compared to what?"

Another reason for not trusting anecdotal evidence is a psychological phenomenon called the placebo effect. People often react to the idea of a treatment, rather than to the treatment itself. A **placebo** is a harmless pill (or sham procedure) that a patient believes is actually an effective treatment. Often, the patient taking the pill feels better, even though the pill actually has no effect whatsoever. In fact, a survey of U.S. physicians published in the *British Medical Journal* (Britt 2008) found that up to half of physicians prescribe sugar pills—placebos—to manage chronic pain. This psychological wish fulfillment—we feel better because we think we *should* be feeling better—is called the **placebo effect**.

Observational Studies

The identifying mark of an **observational study** is that the subjects in the study are put into the treatment group or the control group either by their own actions or by the decision of someone else who is not involved in the research study. For example, if we wished to study the effects on health of smoking cigarettes (as many researchers have), then our treatment group would consist of people who had chosen to smoke, and the control group would consist of those who had chosen not to smoke.

Observational studies compare the outcome variable in the treatment group with the outcome variable in the control group. Thus, if many more people are cured of gum disease in the group that drinks peanut milk (treatment) than in the group that does not (control), then we would say that drinking peanut milk is associated with improvement in gum disease; that is, there is an **association** between the two variables. If the group of happy people tend to have less heart disease than the not-happy people, we would say that happiness is associated with improved heart health.

Note that we do not conclude that peanut milk *caused* the improvement in gum disease. In order for us to draw this conclusion, the treatment group and the control group must be very similar in every way except that one group gets the treatment and the other doesn't. For example, if we knew that the group of people who started drinking peanut milk and the group that did not drink peanut milk were alike in every way—both groups had the same overall health, were roughly the same ages, included the same mix of genders and races, education levels, and so on—then if the peanut milk group is healthier after a year, we would be fairly confident in concluding that peanut milk is the reason for their better health.

Unfortunately, in observational studies this goal of having very similar groups is *extremely* difficult to achieve. *Some* characteristic is nearly always different in one group than in the other. This means that the groups may experience different outcomes because of this different characteristic, not because of the treatment. A difference between the two groups that could explain why the outcomes were different is called a **confounding variable** or **confounding factor**.

For example, early observational studies on the effects of smoking found that a greater percent of smokers than of nonsmokers had lung cancer. However, some scientists argued that genetics was a confounding variable (Fisher 1959). They maintained that the smokers differed genetically from the nonsmokers. This genetic difference made some people more likely to smoke and also more susceptible to lung cancer.

This was a convincing argument for many years. It not only proposed a specific difference between the groups (genetics) but also explained how that difference might come about (genetics makes some people smoke more, perhaps because it tastes better to them or because they have addictive personalities). And the argument also explained why this difference might affect the outcome (the same genetics cause lung cancer). Therefore, the skeptics said, genetics—and not smoking—might be the cause of lung cancer.

Later studies established that the skeptics were wrong about genetics. Some studies compared pairs of identical twins in which one twin smoked and the other did not. These pairs had the same genetic makeup, and still a higher percentage of the smoking twins had cancer than of the nonsmoking twins. Because the treatment and control groups had the same genetics, genetics could not explain why the groups had different cancer rates. When we compare groups in which we force one of the variables to be the same, we say that we are *controlling for* that variable. In these twin studies, the researchers controlled for genetics by comparing people with the same genetic makeup (Kaprio and Koskenvuo 1989).

A drawback of observational studies is that we can never know whether there exists a confounding variable. We can search very hard for it, but the mere fact that we don't find a confounding variable does not mean it isn't there. For this reason, we can never make cause-and-effect conclusions from observational studies.

KEY POINT We can never draw cause-and-effect conclusions from observational studies because of potential confounding variables. A single observational study can show only that there is an *association* between the treatment variable and the outcome variable.

EXAMPLE 6 Does Poverty Lower IQ?

"Chronic Poverty Can Lower Your IQ, Study Shows" is a headline from the online magazine *Daily Finance*. The article (Nisen 2013) reported on a study published in the journal *Science* (Mani et al. 2013) that examined the effects of poverty on problem-solving skills from several different angles. In one part of the study, researchers observed sugar cane farmers in rural India both before and after harvest. Before the harvest, the farmers typically have very little money and are often quite poor. Researchers gave the farmers IQ exams before the harvest and then after the harvest, when they had more money, and found that the farmers scored much higher after the harvest.

QUESTION Based on this evidence alone, can we conclude that poverty lowers people's IQ scores? If yes, explain why. If no, suggest a possible confounding factor.

SOLUTION No, we cannot. This is an observational study. The participants are in or out of the treatment group ("poverty") because of a situation beyond the researchers' control. A possible confounding variable is nutrition; before harvest, without much money, the farmers are perhaps not eating well, and this could lower their IQ scores. (In fact, the researchers considered this confounding variable. They determined that nutrition was relatively constant both before and after the harvest, so it was ruled out as a confounding variable. But other confounding variables may still exist.)

TRY THIS! Exercise 1.41

Controlled Experiments

In order to answer cause-and-effect questions, we need to create a treatment group and a control group that are alike in every way possible, except that one group gets a treatment and the other does not. As you've seen, this cannot be done with observational studies because of confounding variables. In a **controlled experiment**, researchers take control by assigning subjects to the control or treatment group. If this assignment is done correctly, it ensures that the two groups can be nearly alike in every relevant way except whether or not they receive the treatment under investigation.

Well-designed and well-executed controlled experiments are the only means we have for definitively answering questions about causality. However, controlled experiments are difficult to carry out (this is one reason why observational studies are often done instead). Let's look at some of the attributes of a well-designed controlled experiment.

A well-designed controlled experiment has four key features:

- The sample size must be large so that we have opportunities to observe the full range of variability in the humans (or animals or objects) we are studying.
- The subjects of the study must be assigned to the treatment and control groups at random.
- Ideally, the study should be "double-blind," as explained below.
- The study should use a placebo if possible.

These features are all essential in order to ensure that the treatment group and the control group are as similar as possible.

To understand these key design features, imagine that a friend has taken up a new exercise routine in order to lose weight. By exercising more, he hopes to burn more calories and so lose weight. But he notices a strange thing: the more he exercises, the hungrier he gets, and so the more he eats. If he is eating more, can he lose weight? This is a complex issue, because different people respond differently both to exercise and to food. How can we know whether exercise actually leads to weight loss? (See Rosenkilde et al. 2012 for a study related to the hypothetical one presented here.) To think about how you might answer this question, suppose you select a group of slightly overweight young men to participate in your study. For a comparison group, you might ask some of them not to exercise at all during the study. The men in the treatment group, however, you will ask to exercise for about 30 minutes each day at a moderate level.

Sample Size

A good controlled experiment designed to determine whether exercise leads to weight loss should have a large number of people participate in the study. People react to changes in their activity level in a variety of ways, so the effects of exercise can vary from person to person. To observe the full range of variability, you therefore need a large number of people. How many people? This is a hard question to answer, but in general, the more the better. You should be critical of studies with very few participants.

Random Assignment

The next step is to assign people to the treatment group and the comparison group such that the two groups are similar in every way possible. As we saw when we discussed observational studies, letting the participants choose their own group doesn't work, because people who like to exercise might differ in important ways (such as level of motivation) that would affect the outcome.

Instead, a good controlled experiment uses **random assignment** to assign participants to groups. One way of doing this is to flip a coin. Heads means the participant goes into the treatment group, and tails means she or he goes into the comparison group (or the other way around—as long as you're consistent). In practice, the randomizing might instead be done with a computer or even with the random number generator on a calculator, but the idea is always the same: No human determines group assignment. Rather, assignment is left to chance.

If both groups have enough members, random assignment will "balance" the groups. The variation in weights, the mix of metabolisms and daily calorie intake, and the mix of most variables will be similar in both groups. Note that by "similar" we don't mean exactly the same. We don't expect both groups to have exactly the same percentage of people who like to exercise, for example. Except in rare cases, random variation results in slight differences in the mixes of the groups. But these differences should be small.

Whenever you read about a controlled experiment that does not use random assignment, remember that there is a very real possibility that the results of the study are invalid. The technical term for what happens with nonrandomized assignment is **bias**. We say that a study exhibits bias when the results are influenced in one particular direction. A researcher who puts the heaviest people in the exercise group, for example, is biasing the outcome. It's not always easy, or even possible, to predict what the effects of the bias will be, but the important point is that the bias creates a confounding variable and makes it difficult, or impossible, to determine whether the treatment we are investigating really affects the outcome we're observing.

KEY POINT Random assignment (assignment to treatment groups by a randomization procedure) helps balance the groups to minimize bias. This helps make the groups comparable.

Blinding

So far, we've recruited a large number of men and randomly assigned half to exercise and half to remain sedentary. In principle, these two groups will be very similar. However, there are still two potential differences.

First, we might know who is in which group. This means that when we interact with a participant, we might consciously or unconsciously treat that person differently, depending on which group he or she belongs to. For example, if we believe strongly that exercise helps with weight loss, we might give special encouragement or nutrition advice to people who are in the exercise group that we don't give to those in the comparison group. If we do so, then we have biased the study.

To prevent this from happening, researchers should be **blind** to assignment. This means that an independent party—someone who does not regularly see the participants and who does not participate in determining the results of the study—handles the assignment to groups. The researchers who measure the participants' weight loss do not know who is in which group until the study has ended; this ensures that their measurements will not be influenced by their prejudices about the treatment.

Second, we must consider the participants themselves. If they know they are in the treatment group, they may behave differently than they would if they knew nothing about their group assignment. Perhaps they will work harder at losing weight. Or perhaps they will eat much more, because they believe that the extra exercise allows them to eat anything they want.

To prevent this from happening, the participants should also not know whether they are in the treatment group or the comparison group. In some cases, this can be accomplished by not even telling the participants the intent of the study; for example, the participants might not know whether the goal is to examine the effect of a sedentary lifestyle on weight, or whether it is about the effects of exercise. (However, ethical considerations often forbid the researchers from engaging in deception.)

When neither the researchers nor the participants know whether the participants are in the treatment or the comparison group, we say that the study is **double-blind**. The double-blind format helps prevent the bias that can result if one group acts differently from the other because they know they are being treated differently, or because the researchers treat the groups differently or evaluate them differently because of what the researchers hope or expect.

Details

The Real Deal
A study similar to the one described here was carried out in Denmark in 2012. The researchers found that young, overweight men who exercised 30 minutes per day lost more body fat than those who exercised 60 minutes per day. Both groups of exercisers lost more body fat than a group that did not exercise at all. One conclusion is that more intense exercise for overweight people leads to an increase in appetite, and so moderate levels of exercise are best for weight loss.

Placebos

The treatment and comparison groups might differ in still another way. People often react not just to a particular medical treatment, but also to the very idea that they are getting medical treatment. This means that patients who receive a pill, a vaccine, or some other form of treatment often feel better even when the treatment actually does absolutely nothing. Interestingly, this placebo effect also works in the other direction: If they are told that a certain pill might cause side effects (for example, a rash), some patients experience the side effects even though the pill they were given is just a sugar pill.

To neutralize the placebo effect, it is important that the comparison group receive attention similar to what the treatment group receives, so that both groups feel they are being treated the same by the researchers. In our exercise study, the groups behave very differently. However, the sedentary group might receive weekly counseling about lifestyle change, or might be weighed and measured just as frequently as the treatment group. If, for instance, we were studying whether peanut milk improves baldness, we would require the comparison group to take a placebo drink so that we could rule out any placebo effect and thus perform a valid comparison between treatment and control.

KEY POINT

The following qualities are the "gold standard" for experiments.

Large sample size. This ensures that the study captures the full range of variation among the population and allows small differences to be noticed.

Controlled and randomized. Random assignment of subjects to treatment or comparison groups to minimize bias.

Double-blind. Neither subjects nor researchers know who is in which group.

Placebo (if appropriate). This format controls for possible differences between groups that occur simply because some subjects are more likely than others to expect their treatment to be effective.

EXAMPLE 7 Brain Games

Brain-training video games, such as Nintendo's Brain Age, claim to improve basic intelligence skills, such as memory. A study published in the journal *Nature* investigated whether playing such games can actually boost intelligence (Owen et al. 2010). The researchers explain that 11,430 people logged onto a webpage and were randomly assigned to one of three groups. Group 1 completed six training tasks that emphasized "reasoning, planning and problem-solving." Group 2 completed games that emphasized a broader range of cognitive skills. Group 3 was a control group and didn't play any of these games; instead, members were prompted to answer "obscure" questions. At the end of six weeks, the participants were compared on several different measures of thinking skills. The results? The control group did just as well as the treatment groups.

QUESTION Which features of a well-designed controlled experiment does this study have? Which features are missing?

SOLUTION Sample size: The sample size of 11,430 is quite large. Each of the three groups will have about 3800 people.

Randomization: The authors state that patients were randomly assigned to one of the three groups.

Double-blind format: Judging on the basis of this description, there was no double-blind format. It's possible (indeed, it is likely) that the researchers did not know, while analyzing the outcome, to which treatment group individuals had been assigned. But we do not know whether *participants* were aware of the existence of the three different groups and how they differed.

Placebo: The control group participated in a "null" game, in which they simply answered questions. This activity is a type of placebo, because the participants could have thought that this null game was a brain game.

TRY THIS! Exercise 1.43

Extending the Results

In both observational studies and controlled experiments, researchers are often interested in knowing whether their findings, which are based on a single collection of people or objects, will extend to the world at large.

The researchers in Example 7 concluded that brain games are not effective, but might it just be that the games weren't effective for those people who decided to participate? Maybe if the researchers tested people in another country, for example, the findings would be different.

It is usually not possible to make generalizations to a larger group of people unless the subjects for the study are representative of the larger group. The only way to collect a sample that is representative is to collect the objects we study at random. We will discuss how to collect a random sample, and why we can then make generalizations about people or objects who were not in the sample, in Chapter 7.

Selecting subjects using a random method is quite common in polls and surveys (which you'll also study in Chapter 7), but it is much less common in other types of studies. Most medical studies, for example, are not conducted on people selected randomly, so even when a cause-and-effect relationship emerges between the treatment and the response, it is impossible to say whether this relationship will hold for a larger (or different) group of people. For this reason, medical researchers often work hard to replicate their findings in diverse groups of people.

> **! Caution**
>
> **At Random**
> The concept of randomness is used in two different ways in this section. *Random assignment* is used in a controlled experiment. Subjects are randomly assigned to treatment and control groups in order to achieve a balance between groups. This ensures that the groups are comparable to each other and that the only difference between the groups is whether or not they receive the treatment under investigation. ***Random selection*** occurs when researchers select subjects from some larger group via a random method. We must employ random selection if we wish to extend our results to a larger group.

Statistics in the News

When reading in a newspaper or blog about a research study that relies on statistical analysis, you should ask yourself several questions to evaluate how much faith you can put in the conclusions reached in the study.

1. *Is this an observational study or a controlled experiment?*

   If it's an observational study, then you can't conclude that the treatment caused the observed outcome.

2. *If the study is a controlled experiment, was there a large sample size? Was randomization used to assign participants to treatment groups? Was the study double-blind? Was there a placebo?*

   See the relevant section of this chapter for a review of the importance of these attributes.

3. *Was the paper published in a peer-reviewed journal? What is the journal's reputation?*

   "Peer-reviewed" means that each paper published in the journal is rigorously evaluated by at least two anonymous researchers familiar with the field. The best journals are very careful about the quality of the research they report on. They have many checkpoints to make sure that the science is as good as it can be. (But remember, this doesn't mean the science is perfect. If you read a medical journal regularly, you'll see much debate from issue to issue about certain results.) Other journals, by contrast, sometimes allow sloppy research results, and you should be very wary of these journals.

4. *Did the study follow people for a long enough time?*

   Some treatments take a long time to work, and some illnesses take a long time to show themselves. For example, many cost-conscious people like to refill water bottles again and again with tap water. Some fear that drinking from the same plastic bottle again and again might lead to cancer. If this is true, it might take a very long time for a person to get cancer from drinking out of the same bottle day after day. So researchers who wish to determine whether drinking water from the same bottle causes cancer should watch people for a very long time.

Often it is hard to get answers to all of these questions from a newspaper article. Fortunately, the Internet has made it much easier to find the original papers, and your college library or local public library will probably have access to many of the most popular journals.

Even when a controlled experiment is well designed, things can still go wrong. One common way in which medical studies go astray is that people don't always do what their doctor tells them to do. Thus, people randomized to the treatment group might not actually take their treatments. Or people randomized to the Atkins diet might switch to Weight Watchers because they don't like the food on the Atkins diet. A good research paper will report on these difficulties and will be honest about the effect on their conclusions.

EXAMPLE 8 Does City Living Raise Blood Pressure?

The *New York Times* provided a link to an article that claims higher blood pressure occurs in people who live in urban areas (www.scienceblog.com 2010).

QUESTION Is this more likely to be an observational study or a controlled experiment? Why? Does this mean that moving to a more urban area will result in an increase in your blood pressure?

SOLUTION This is most likely an observational study. The treatment variable is whether or not a person lives in an urban area. We are not told that participants were randomly assigned to live in urban or nonurban areas for some period of time, and in fact it is pretty unlikely that such a study could be done. Researchers probably measured blood pressure in people who chose to live in urban or nonurban settings. Because the participants themselves chose where to live, this is an observational study.

Because this is an observational study, we cannot conclude that the *treatment* (living in urban areas) causes the *outcome* (increased blood pressure). This means that simply moving to (or away from) an urban area may not change your blood pressure. One potential confounding variable could be personality, because people who prefer the fast pace of city life might have other personality traits that increase their blood pressure.

TRY THIS! Exercise 1.49

EXAMPLE 9 Crohn's Disease

Crohn's disease is a bowel disease that causes cramping, abdominal pain, fever, and fatigue. A study reported in the *New England Journal of Medicine* (Columbel et al. 2010) tested two medicines for the disease: injections of infliximab (Inflix) and oral azathioprine (Azath). The participants were randomized into three groups. All groups received an injection and a pill (some were placebos, but still a pill and an injection). One group received Inflix injections alone (with placebo pills), one received Azath pills alone (with placebo injections), and one group received both injections and pills. A good outcome was defined as the disease being in remission after 26 weeks. The accompanying table shows the summary of the data that this study yielded.

	Combination	Inflix Alone	Azath Alone
Remission	96	75	51
Not in Remission	73	94	119

QUESTIONS

a. Compare the percentages in remission for the three treatments. Which treatment was the most effective and which was the least effective for this sample?

b. Can we conclude that the combination treatment causes a better outcome? Why or why not?

SOLUTIONS

a. For the combination: $96/169$, or 56.8%, success
For the Inflix alone: $75/169$, or 44.4%, success
For the Azath alone: $51/170$, or 30%, success

The combination treatment was the most effective for this sample, and Azath alone was the least effective.

b. Yes, we can conclude that the combination of drugs causes a better outcome than the single drugs. The study was placebo-controlled and randomized. The sample size was reasonably large. Blinding was not mentioned, but at least, thanks to the placebos, the patients did not know what treatment they were getting.

TRY THIS! Exercise 1.51

CASE STUDY REVISITED

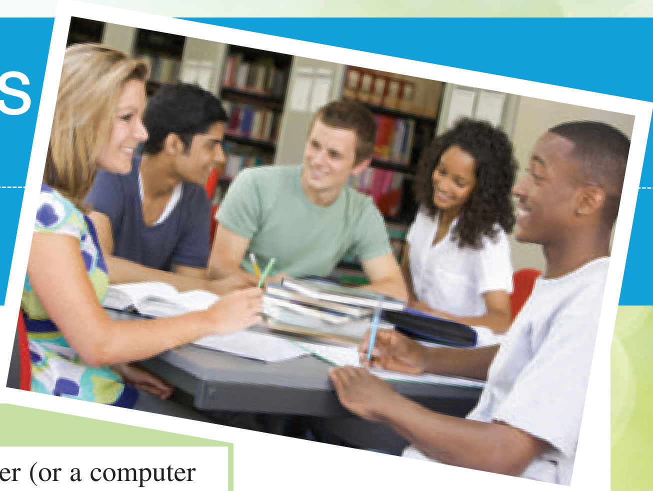
Do cell phones cause cancer? The evidence presented at the beginning of the chapter seems to suggest that they might. But let's consider the nature of the evidence. The man who appeared on *Larry King Live* said that his wife, who had died of brain cancer, talked on her cell phone a great deal. But this is just an anecdote. We can conclude nothing about the effects of cell phones on brain cancer, because we don't know how many other people who talked the same amount of time on their cell phones did not get brain cancer, and we don't know how many people who *did* get brain cancer did not use cell phones at all.

The Hardell study was an observational study. The two groups compared were a group that had brain cancer and a group that did not. Clearly, the researchers could not assign subjects to one of these two groups, so all they could do was observe that one group had a different brain cancer rate than the other. However, many potential confounders might explain why the two groups had different cell phone usage. For example, critics of the study point out that Hardell examined only brain cancer survivors, which means that he excluded a large percentage of those with brain cancer. Survivors are different from others, and whether or not people survived cancer determined whether they were in this study or not and affected their responses. (Clearly, it's easier to determine how much time people spend talking on the phone if they are alive and able to answer questions.) Researchers in Australia actually did a controlled experiment with mice. They found that the brain cancer risk for mice that received cell phone levels of radiation was no different from that for mice that did not receive the radiation.

Ruling that the evidence was insufficient to suggest a causal relationship between cell phone use and brain cancer, the judge threw out the lawsuit.

EXPLORING STATISTICS
CLASS ACTIVITY

Collecting a Table of Different Kinds of Data



GOALS	MATERIALS
In this activity, you will learn about different types of data and discuss how to summarize data derived from members of the class.	A board and a marker (or a computer spreadsheet).

ACTIVITY

As a class, you will suggest possible variables that would help you describe your class to others. You will then collect data and summarize and describe these variables. How would you describe the people in your class as a group? How does your class compare to another class, maybe the class next door? Suggest variables you might collect on individuals that would help you describe your class and make comparisons. (Examples of such variables might include gender, number of credits currently taken, distance the individual lives from campus, and the like.) Be prepared to give your own values for these variables.

Some variables (such as family income, for example) might be too private; inquiring about others might be offensive to some students. You should not feel compelled to provide information that might embarrass you; try to avoid suggesting variables that you think might embarrass others. Table A shows an example of the format for the input.

Gender	Age	Favorite Class	Units	____	____

▲ TABLE A

BEFORE THE ACTIVITY

If someone asked you to describe the students who make up your class, which characteristics would you focus on?

AFTER THE ACTIVITY

1. Which variables that were used are numerical? Which are categorical?
2. Write a one-paragraph summary of the data you collected that you think would help someone understand the composition of your class.

CHAPTER REVIEW

KEY TERMS

statistics, *27*
variation, *28*
data, *28*
variables, *30*
data set, *30*
sample, *30*
population, *30*
numerical variable, *30*
categorical variable, *30*
stacked data, *32*
unstacked data, *32*
two-way table, *35*
frequency, *35*
rate, *38*
causality, *39*
treatment variable, *39*
outcome variable (or response variable), *39*
treatment group, *39*
comparison group (or control group), *39*
anecdotes, *39*
placebo, *40*
placebo effect, *40*
observational study, *40*
association, *40*
confounding variable, *41*
controlled experiment, *42*
random assignment, *43*
bias, *43*
blind, *43*
double-blind, *44*
random selection, *45*

LEARNING OBJECTIVES

After reading this chapter and doing the assigned homework problems, you should

- Be able to distinguish between numerical and categorical variables and understand methods for coding categorical variables.
- Know how to find and use rates (including percentages) and understand when and why they are more useful than counts for describing and comparing groups.
- Understand when it is possible to infer a cause-and-effect relationship from a research study and when it is not.
- Be able to explain how confounding variables prevent us from inferring causation, and suggest confounding variables that are likely to occur in some situations.
- Be able to distinguish between observational studies and controlled experiments.

SUMMARY

Statistics is the science (and art) of collecting and analyzing observations (called data) and communicating your discoveries to others. Often, we are interested in learning about a population on the basis of a sample taken from that population. Here are some questions to consider when first examining a data set:

Who, or what, was observed?
What variables were measured?
How were they measured?
What are the units of measurement?
Who collected the data?
How did they collect the data?
Where were the data collected?
Why were the data collected?
When were the data collected?

With categorical variables, we are often concerned with comparing rates or frequencies between groups. A two-way table is sometimes a useful summary. Always be sure that you are making valid comparisons by comparing proportions or percentages of groups, or that you are comparing the appropriate rates.

Many studies are focused on questions of causality: If we make a change to one variable, will we see a change in the other? Anecdotes are not useful for answering such questions. Observational studies can be used to determine whether associations exist between treatment and outcome variables, but because of the possibility of confounding variables, observational studies cannot support conclusions about causality. Controlled experiments, if they are well designed, do allow us to draw conclusions about causality.

A well-designed controlled experiment should have the following attributes:

A large sample size
Random assignment of subjects to a treatment group and to a control group
A double-blind format
A placebo

SOURCES

BBC News, February 18, 2010. http://news.bbc.co.uk (accessed May 13, 2010).

Britt, C. October 23, 2008. U.S. doctors prescribe drugs for placebo effects, survey shows. *British Medical Journal*. Bloomberg.com (accessed October 23, 2008).

Brody, J. 2002. Cellphone: A convenience, a hazard or both? *New York Times*, October 1.

CNN. 2002. Cancer study may be used in Motorola suit. cnn.com (accessed September 23, 2002).

Colombel, Sandborn, Reinisch, Mantzaris, Kornbluth, Rachmilewitz, Lichtiger, D'Haens, Diamond, Broussard, Tang, van der Woude, and Rutgeerts. 2010. Infliximab, azathioprine, or combination therapy for Crohn's disease. *New England Journal of Medicine* 362, 1383–1395.

Fisher, R. 1959. Smoking: The cancer controversy. Edinburgh, UK: Oliver and Boyd.

Glionna, J. 2006. Word of mouth spreading about peanut milk. *Los Angeles Times,* May 17, 2006.

Holcomb, J. 2006. http://www2.irss.unc.edu/ncvital/index.html (accessed May 17, 2006).

Insurance Institute for Highway Safety. 2013. http://www.iihs.org/news/ (accessed August 20, 2013).

Kaprio, J., and M. Koskenvuo. 1989. Twins, smoking and mortality: A 12-year prospective study of smoking-discordant twin pairs. *Social Science and Medicine* 29(9), 1083–1089.

McKinsley Quarterly. 2009. Hal Varian on how the Web challenges managers. Business Technology Office, January.

Mani, A., et al. 2013. Poverty impedes cognitive function. *Science* 341, 976.

National Safety Council. 2011. *Injury facts*. Itasca, Illinois.

Nisen, M. 2013. Chronic poverty can lower your IQ, study shows. *Daily Finance*, August 31, 2013. http://www.dailyfinance.com/ (accessed September 1, 2013).

Owen, A., et al. 2010. Letter: Putting brain training to the test. *Nature* advance online publication, April 20, www.nature.com (accessed May 15, 2010).

Rosenkilde, M., et al. 2012. Body fat loss and compensatory mechanisms in response to different doses of aerobic exercise in a randomized controlled trial in overweight sedentary males. *Am J Physiol Regul Integr Comp Physiol* 303, R571–R579.

www.scienceblog.com. 2010. Higher blood pressure found in people living in urban areas (accessed May 2010).

SECTION EXERCISES

SECTION 1.2

The data in Table 1A were collected from a gym class. The column heads give the variable, and each of the other rows represents an individual in the class. Refer to this table for Exercises 1.1–1.10, 1.19, and 1.20.

Male	Age	Eye Color	Shoe Size	Height (inches)	Weight (pounds)	Number of Siblings	College Units This Term	Handedness
1	20	Brown	9.5	71	170	1	16	Right
0	19	Blue	8	66	135	1	13	Right
0	42	Brown	7.5	63	130	3	5	Right
0	19	Brown	8.5	65	150	0	15	Left
1	21	Brown	11	70	185	5	19.5	Right
0	20	Hazel	5.5	60	105	2	11.5	Right
1	21	Blue	12	76	210	2	9.5	Right
0	21	Brown	10	70	140	0	8	Left
0	32	Brown	8	64	165	1	13.5	Right
1	23	Brown	7.5	63	145	6	12	Right
0	21	Brown	6.5	61.5	110	4	14	Right

▲ TABLE 1A

1.1 Variables In Table 1A, how many variables are there?

1.2 People In Table 1A, there are observations on how many people?

TRY **1.3 (Example 1)** Are the following variables, from Table 1A, numerical or categorical? Explain.

a. Handedness
b. Age

1.4 Are the following variables, from Table 1A, numerical or categorical? Explain.

a. Shoe size
b. Eye color

1.5 Give an example of another numerical variable we might have recorded for the students whose data are in Table 1A?

1.6 Give an example of another categorical variable we might have recorded for the students whose data are in Table 1A?

1.7 Coding Suppose you decided to code eye color using 1 for Brown Eyes and 0 for Not Brown Eyes. What would be the label at the top of the column, and how many ones and zeros would there be?

1.8 Coding Suppose you decided to code handedness using Right-handed as the label for the column. How many ones and how many zeros would there be?

1.9 Coding Explain why the variable Male, in Table 1A, is categorical, even though its values are numbers. Often, it does not make sense, or is not even possible, to add the values of a categorical variable. Does it make sense for Male? If so, what does the sum represent?

1.10 Coding Students with fewer than 12 units in the current term are considered part-time. Create a new categorical variable that classifies each student in Table 1A as full-time (12 or more units) or part-time. Call this variable Full. Report the values in a column in the same order as those in the table. Use codes (1 and 0) in your column.

TRY **1.11 Brain Size (Example 2)** In 1991, researchers conducted a study on brain size as measured by pixels in a magnetic resonance imagery (MRI) scan. The numbers are in hundreds of thousands of pixels. The data table provides the sizes of the brains and the gender. (Source: www.lib.stat.cmu.edu/DASL)

a. Is the format of the data set stacked or unstacked?

b. Explain the coding. What do 1 and 0 represent?

c. If you answered "stacked" in part a, then unstack the data into two columns labeled Male and Female. If you answered "unstacked," then stack the data into one column; choose an appropriate name for the stacked variable.

Brain	Male
9.4	1
9.5	0
9.5	1
9.5	1
9.5	0
9.7	1
9.9	0

1.12 Students' Heights The accompanying table gives heights (in centimeters) of students in two statistics classes. One class is labeled "Class A" and the other is labeled "Class B".

a. Is the format for the data set stacked or unstacked?

b. If you answered "stacked," then unstack the data set. If you answered "unstacked," then stack and code the data set.

Class A	Class B
176	182
156	146
178	155
165	171
172	168

1.13 Meetings Philip, an employee at an IT firm in Mexico, kept track of the duration of his daily meetings (in minutes) for one week. He also took note of whether the meetings took place in the morning or in the afternoon.

Morning meetings: 45, 90, 75, 35, 60, 80

Afternoon meetings: 50, 65, 45, 70

Write these data as they might appear in (a) stacked format with codes and (b) unstacked format.

1.14 A day at the Spa A sample of working professionals in the age group of 25–30 were questioned to determine how much they would be willing to pay for a day at the spa. The males responded (in dollars): 200, 250, 330, 275, and 300. The female students responded: 300, 350, 250, and 215. Write these data as they might appear in (a) stacked format with codes and (b) unstacked format.

SECTION 1.3

TRY **1.15 Older Siblings (Example 3)** At a small four-year college, some psychology students were asked whether or not they had at least one older sibling. The table shows the results for men and women and shows some of the totals.

	Men	Women	Total
Yes, Older S	12	55	?
No Older S	11	39	50
	23	?	117

a. Calculate the totals that are not shown, and report them in the table.

b. What percentage of the men had an older sibling?

c. What percentage of the men did not have an older sibling?

d. What percentage of the women had an older sibling?

e. What percentage of the people had an older sibling?

f. What percentage of the people with an older sibling were women?

g. Suppose that in a group of 600 women, the percentage who have an older sibling is the same as in the sample here. How many of the 600 women would have an older sibling?

1.16 Favorite Subject At a high school, a few eighth grade students were asked which subject they liked more: mathematics or science. The table shows the results for boys and girls.

	Boys	Girls	Total
Mathematics	55	52	?
Science	67	75	142
Total	?	127	?

a. Figure out the missing totals, and report them in your table.

b. What percentage of the boys liked mathematics?

c. What percentage of the boys liked science?

d. What percentage of the girls liked mathematics?

e. What percentage of the students liked science?

f. What percentage of the students who liked science were girls?

g. What percentage of the students who liked mathematics were boys?

h. Suppose that in a group of 800 girls, the percentage who like science is the same as in the sample here. How many of the 800 girls would like science?

TRY **1.17 Finding and Using Percentages (Example 4)**

a. A glass jar contains 17 blue marbles and 25 red marbles. What percentage of the marbles are blue?

b. A different glass jar has 430 marbles, and 63% of them are blue. How many blue marbles are in the jar?

c. A different glass jar contains 45% red marbles and has 90 red marbles in it. What is the total number of marbles in the jar?

1.18 Finding and Using Percentages

a. In a class of 465 students, 54% of the students were right-handed. How many right-handed students are there?

b. A pharmaceutical company recently hired 182 pharmacy graduates and 94 of them are male. What percentage of these graduates are female?

c. A publishing house comprises 56% male editors, or 280 male editors. What is the total number of editors at the firm?

1.19 Women Find the frequency, proportion, and percentage of women in Table 1A on page 51.

1.20 Right-Handed People Find the frequency, proportion, and percentage of right-handed people in Table 1A on page 51.

g **1.21 Two-way Table from Data** The first step in finding an association between gender and handedness from among the students in Table 1A (on page 51) is to create a two-way table with these two variables. Make a two-way table with the labels Male and Female across the top and the labels Right and Left on the side, and then answer the following questions about it. *See page 58 for guidance on parts a and b.*

a. Report how many students are in each of the four cells of the table by writing these values in your table.
b. Sum the numbers of students in each row and each column, and put these sums into your table. Also find the total number of students, and put it in the lower right corner.
c. What percentage of the females are right-handed?
d. What percentage of the right-handed students are female?
e. What percentage of the students are right-handed?
f. If the percentage of right-handed females remained roughly the same and there were 70 females, how many of them would be right-handed?

1.22 Two-way Table from Data Make a two-way table from Table 1A, for gender and eye color. Put the labels Male and Female across the top and the labels Brown, Blue, and Hazel on the side, and then tally the data.

a. Arrange the data as a two-way table and report how many students are in each cell.
b. Sum the numbers of students in each row and each column, and put these sums into your table. Also put the total number of students in the lower right corner.
c. What percentage of the females have brown eyes?
d. What percentage of the people who have brown eyes are female?
e. What percentage of the people have brown eyes?
f. If the percentage of brown-eyed females was roughly the same and there were 60 females, how many of them would have brown eyes?

1.23 Population Prediction The *2009 World Almanac and Book of Facts* predicted that the United States will have an elderly population (65 and older) of 88,547,000 in the year 2050 and that this will be 20.2% of the population. What is the total predicted U.S. population in 2050?

1.24 2007 Population The *2009 World Almanac and Book of Facts* reported that in 2007 there were 12,608,000 people age 16 or older who had a "go outside the home" disability and that this was 5.5% of the U.S. population (of this age group). These are people who cannot go outside the home without help. How large was the total population (of this age group) in 2007?

g **1.25 Living with AIDS** The table gives the number of people diagnosed with AIDS/HIV in 2010 in the five states with the largest number of cases, as well as the District of Columbia, as reported by the U.S. Centers for Disease Control and Prevention (CDC). It also shows the population of those regions at that time from the U.S. Census Bureau.

Find the number of people diagnosed with AIDS/HIV per thousand residents in each region, and rank the six regions from highest rate (rank 1) to lowest rate (rank 6). Compare these rankings (of rates) with the ranks of total number of cases. If you moved to one of these regions and met 50 random people, in which region would you be most likely to meet at least one person diagnosed with HIV? In which of these regions would you be least likely to meet at least one person with HIV? *See page 58 for guidance.*

State	HIV	Population
New York	192,753	19,421,005
California	160,293	37,341,989
Florida	117,612	18,900,773
Texas	77,070	25,258,418
New Jersey	54,557	8,807,501
District of Columbia	9,257	601,723

1.26 Population Density The accompanying table gives the population of the six U.S. states with the largest populations in 2008 and the area of these states. (Source: www.infoplease.com)

State	Population	Area (square miles)
Pennsylvania	12,448,279	44,817
Illinois	12,901,563	55,584
Florida	18,328,340	53,927
New York	19,490,297	47,214
Texas	24,326,974	261,797
California	36,756,666	155,959

a. Determine and report the rankings of the population density by dividing each population by the number of square miles to get the population density (in people per square mile). Use rank 1 for the highest density.
b. If you wanted to live in the state (of these six) with the lowest population density, which would you choose?
c. It you wanted to live in the state (of these six) with the highest population density, which would you choose?

1.27 Marriage Rates The number of married people in the United States and the total number of adults in the United States (in millions) are provided in the accompanying table for several years.

Find the percentage of people married in each of the given years, and describe the trend over time. (Source: *2009 World Almanac and Book of Facts*)

Year	Married	Total
1990	112.6	191.8
1997	116.8	207.2
2000	120.2	213.8
2007	129.9	235.8

1.28 Births and Deaths The following information about the number of births and the number of deaths (in thousands) for certain years is taken from the *2012 World Almanac and Book of Facts*. Report the death rate as a percentage of the birth rate, and comment on its trend over time. What is causing the trend?

Year	Births	Deaths
2006	4266	2426
2007	4316	2424
2008	4248	2473
2009	4131	2437
2010	4007	2452

TRY *** 1.29 Course Enrollment Rates (Example 5)** Two sections of statistics are offered, the first at 8 a.m. and the second at 10 a.m. The 8 a.m. section has 25 women, and the 10 a.m. section has 15 women. A student claims this is evidence that women prefer earlier statistics classes than men do. What information is missing that might contradict this claim?

*** 1.30 Pedestrian Fatalities** In 2008, the National Highway Traffic Safety Administration reported that the number of pedestrian fatalities in Miami-Dade County, Florida, was 65 and that the number in Hillsborough County, Florida, was 45. Can we conclude that pedestrians are safer in Hillsborough County? Why or why not?

SECTION 1.4

For Exercises 1.31–1.38, indicate whether the study is an observational study or a controlled experiment.

1.31 A scientist is interested in studying the effect of eating bananas on athletes' performance levels. She randomly divides a group of athletes into three groups: one group will eat a banana after 30 minutes of exercise, the second will drink water after exercise, and the third will consume nothing. Their efficiency after consuming banana, water, or nothing is measured.

1.32 Children with ADHD are randomly divided into two groups. One group is given fish oil supplements, and the other is given a placebo. After six months, their symptoms were assessed to see whether fish oil supplements helped reduce ADHD symptoms better than a placebo.

1.33 A dietician observed for a week whether employees in a random office added sugar or sugar substitute to their coffee.

1.34 Medical records of patients who smoked are examined to see whether those who quit smoking show improvement in lung function test results in comparison to those who did not.

1.35 A large IT firm provides the option of flexible work hours to its employees. The head of each department compares the productivity of employees who start working in the morning to those who start working in the afternoon to determine whether those who come to work early are more productive than those who start late.

1.36 A group of children is randomly divided into two groups. One group plays outdoor games in the evening for one hour, and the other group plays indoor games for one hour. After six months, the two groups are compared to see the effects of outdoor playing versus indoor playing on their mental growth.

1.37 A researcher was interested in the effect of physical education on the mental alertness in school children. She assigned students of one class to attend a physical education session in the morning while students in the other class attended a science class. The researcher then asked students from both classes to fill out a questionnaire that assessed their attentiveness.

1.38 A researcher was interested in the effect of chamomile tea on sleep. He compared a group of middle-aged women who drink chamomile tea each night to those who do not. He observed their moods in the morning to determine whether or not chamomile tea was effective in improving the quality of their sleep.

1.39 Effects of Audio-visual Education on Grades A group of educators wants to determine the effectiveness of multimedia tools in raising students' grades. So they gave the students an option to choose a class with audio-visual tools or one without any. Then they conducted a test to check how well the students retain their lessons and compared the results of both the groups. Let's assume the group that chose the audio-visual class received higher grades. Does that show that the audio-visual teaching aids work better? If not, explain why not and suggest a confounding variable.

1.40 Weight Loss A health instructor, who believes that weight training is better than cardio for weight loss, divides a bunch of obese people into two groups: one group is assigned to weight training while the other is assigned to cardio. After six months, the participants are assessed to measure the effectiveness of both the exercise regimens on their weight loss.

a. The health instructor is concerned that the health of the highly obese people will deteriorate if they do not start weight training and thus assigns them to the weight training group. Explain why this will affect his ability to determine which exercise regimen works best.

b. What suggestion would you give the health instructor to improve his experiment?

c. The health instructor asks the personal trainers whether it is advisable for him to know which exercise regimen each individual follows and to assess them himself at the end of the experiment to rate their improvement. Explain why this practice will affect his ability to determine which approach works best.

d. What improvements to the plan in part c do you recommend?

TRY **1.41 Early Tonsillectomy for Children (Example 6)** Marcus reported on a study that treated children who had sleep apnea. Sleep apnea interferes with breathing while the child is asleep. Read the excerpts from the abstract, and answer the questions that follow it. (Adenotonsillectomy is surgery to remove the tonsils and adenoids.) (Source: Marcus et al., A randomized trial of adenotonsillectomy for childhood sleep apnea, *New England Journal of Medicine*, vol. 368: 2366–2376, June 20, 2013)

> We randomly assigned 464 children, 5 to 9 years of age, with the obstructive sleep apnea syndrome to early adenotonsillectomy or a strategy of watchful waiting....
>
> There were significantly greater improvements in behavioral, quality-of-life, and polysomnographic [sleep study] findings and significantly greater reduction in symptoms in the early-adenotonsillectomy group than in the watchful-waiting group....

a. Was the study a controlled experiment or an observational study? Explain how you know.

b. Assuming that the study was properly conducted, can we conclude that the early surgery caused the improvements? Explain.

1.42 Pneumonia Vaccine for Young Children A study reported by Griffin et al. compared the rate of pneumonia in

1997–1999 before pneumonia vaccine (PCV7) was introduced and in 2007–2009 after pneumonia vaccine was introduced. Read the excerpts from the abstract, and answer the question that follows it. (Source: Griffin et al., U.S. hospitalizations for pneumonia after a decade of pneumococcal vaccination, *New England Journal of Medicine*, vol. 369: 155–163, July 11, 2013)

> We estimated annual rates of hospitalization for pneumonia from any cause using the Nationwide Inpatient Sample database…. Average annual rates of pneumonia-related hospitalizations from 1997 through 1999 (before the introduction of PCV7) and from 2007 through 2009 (well after its introduction) were used to estimate annual declines in hospitalizations due to pneumonia.
>
> The annual rate of hospitalization for pneumonia among children younger than 2 years of age declined by 551.1 per 100,000 children … which translates to 47,000 fewer hospitalizations annually than expected on the basis of the rates before PCV7 was introduced.

Results for other age groups were similar. Does this show that pneumonia vaccine caused the decrease in pneumonia that occurred? Explain.

TRY **1.43 Copper Bracelets (Example 7)** Some people believe that wearing copper bracelets is a good treatment for arthritis of the hand. To test this belief, suppose you recruit 100 people and supply them all with copper bracelets. After the patients wear the bracelets for a month, you ask them whether or not their pain is less than it was before they began wearing the bracelets. Explain how to improve this study.

1.44 Stress Management Study A group of working middle-aged men are asked to participate in a stress management study. Participants are allowed to choose whether they want to try daily meditation or follow a daily exercise routine. Half of the people choose meditation, and the other half choose to exercise every day. Let's assume that there is greater stress reduction in the exercise group.

a. Suggest a plausible confounding variable that would prevent us from concluding that the stress reduction was due to the exercise alone. Explain why it is a confounding variable.
b. Explain a better way to conduct the experiment that is likely to remove the influence of confounding variables.

1.45 Do Pesticides Cause Parkinson's Disease? A study by Pezzoli and Cereda was reported in *Neurology*, May 28, 2013. The report said that the use of pesticides is associated with the development of Parkinson's disease, which is a neurological disease that causes people to shake. The study reported that exposure to bug killers and weed killers is "associated with" an increase of 33% to 80% in the chances of getting Parkinson's. Does this study show that pesticides cause Parkinson's disease? Why or why not? (Source: Pezzoli and Cereda, Exposure to pesticides or solvents and risk of Parkinson disease, *Neurology*, vol. 80, no. 22: 2035–2041, May 28, 2013)

1.46 Breast Cancer Two drugs were tested to see whether they helped women who had breast cancer without lymph node involvement. The drugs are called TAC (docetaxel, doxorubicin, and cyclophosphamide) and FAC (fluorouracil, doxorubicin, and cyclophosphamide). About half of the 1060 women with breast cancer without lymph node involvement were randomly assigned to TAC, and the other half were assigned to FAC. After 77 months, 473 out of 539 of the women assigned to TAC were alive, and 426 out of 521 women assigned to FAC were alive. (Source: Martin et al., Adjuvant docetaxel for high-risk, node-negative breast cancer, *New England Journal of Medicine,* vol. 363, no. 23: 2200–10, December 2, 2010)

a. Find both sample percentages of survival, and compare them descriptively.
b. Was this a controlled experiment or an observational study? Explain why. From studies like these, can we conclude a cause-and-effect relationship between the drug type and the survival percentage? Why or why not?

1.47 Flu Vaccine In the fall of 2004, there was a shortage in flu vaccine in the United States after it was discovered that vaccines from one of the manufacturers were contaminated. The *New England Journal of Medicine* reported on a study that was done to see whether a smaller dose of the vaccine could be used successfully. If that were the case, then a small amount of vaccine could be divided into more flu shots. In this study, the usual amount of vaccine was injected into half the patients, and the other half of the patients had only a small amount of vaccine injected. The response was measured by looking at the production of antibodies (more antibodies generally result in less risk of getting the flu). In the end, the lower dose of vaccine was just as effective as a higher dose for those under 65 years old. What more do we need to know to be able to conclude that the lower dose of vaccine was equally effective at preventing the flu for those under 65? (Source: Beishe et al., Serum antibody responses after intradermal vaccination against influenza, *New England Journal of Medicine*, 2004)

1.48 Effect of Confederates on Compliance A study was conducted to see whether participants would ignore a sign that said, "Elevator may stick between floors. Use the stairs." The study was done at a university dorm on the ground floor of a three-level building. Those who used the stairs were said to be compliant, and those who used the elevator were said to be noncompliant. There were three possible situations, two of which involved confederates. A confederate is a person who is secretly working with the experimenter. In the first situation, there was no confederate. In the second situation, there was a compliant confederate (one who used the stairs), and in the third situation, there was a noncompliant confederate (one who used the elevator). The subjects tended to imitate the confederates. What more do you need to know about the study to determine whether the presence or absence of a confederate causes a change in the compliance of subjects? (Source: Wogalter, et al., (1987), reported in Shaffer & Merrens (2001), *Research Stories in Introductory Psychology,* Allyn and Bacon: Boston.)

TRY **1.49 Vitamin C and Allergies (Example 8)** Posted at the Mayo Clinic's website was information on the use of vitamin C for breast-feeding mothers. The children whose mothers had chosen to take high doses of vitamin C had a 30% lower risk of developing allergies. Can you conclude that the use of vitamin C caused the reduction in allergies? Why or why not?

1.50 Does Sleep Deprivation Make You Fat? The *American Journal of Clinical Nutrition* (June 2011) suggested that reduction in sleep increases energy and fat intakes. Is this likely to be a conclusion from observational studies or randomized experiments? Can we conclude that sleeping less causes one to gain weight? Why or why not?

TRY **1.51 Effects of Light Exposure (Example 9)** A study carried out by Baturin and colleagues looked at the effects of light on female mice. Fifty mice were randomly assigned to a regimen of 12 hours of light and 12 hours of dark (LD), while another fifty mice were assigned to 24 hours of light (LL). Researchers observed the mice for two years, beginning when the mice were two months old. Four of the LD mice and 14 of the LL mice developed tumors. The accompanying table summarizes the data. (Source: Baturin et al., The effect of light regimen and melatonin on the development of spontaneous mammary tumors in mice, *Neuroendocrinology Letters*, 2001)

	LD	LL
Tumors	4	14
No tumors	46	36

a. Determine the percentage of mice that developed tumors from each group (LL and LD). Compare them and comment.
b. Was this a controlled experiment or an observational study? How do you know?
c. Can we conclude that light for 24 hours a day causes an increase in tumors in mice? Why or why not?

1.52 Scared Straight The idea of sending delinquents to "Scared Straight" programs has appeared recently in several media programs (such as *Dr. Phil*) and on a program called *Beyond Scared Straight*. So it seems appropriate to look at a randomized experiment from the past. In 1983, Roy Lewis reported on a study in California. Each male delinquent in the study (all were aged 14–18) was randomly assigned to either Scared Straight or no treatment. The males who were assigned to Scared Straight went to a prison, where they heard prisoners talk about their bad experiences there. Then the males in both the experimental and the control group were observed for 12 months to see whether they were rearrested. The table shows the results. (Source: Lewis, Scared straight—California style: Evaluation of the San Quentin Squires program. *Criminal Justice and Behavior,* vol. 10: 209–226, 1983)

	Scared Straight	No Treatment
Rearrested	43	37
Not rearrested	10	18

a. Report the rearrest rate for the Scared Straight group and for the No Treatment group, and state which is higher.
b. This experiment was done in the hope of showing that Scared Straight would cause a lower arrest rate. Did the study show that? Explain.

CHAPTER REVIEW EXERCISES

1.53 Obesity and Marital Status A 2009 study analyzed data from the National Longitudinal Study of Adolescent Health. Participants were studied into adulthood. Each study participant was categorized as to whether they were obese or not and whether they were dating, cohabiting, or married. Obesity was defined as having a Body Mass Index of 30 or more. (Source: The and Larsen, Entry into romantic partnership is associated with obesity, *Obesity*, vol. 17, no. 7: 1441–1447, 2009)

	Dating	Cohabiting	Married
Obese	81	103	147
Not Obese	359	326	277
	440	429	424

a. What percentage of those who were dating were obese?
b. What percentage of those who were cohabiting were obese?
c. What percentage of those who were married were obese?
d. Which group had the highest rate of obesity? Does this imply that marital status causes obesity? Why or why not? If not, can you name a confounding variable?

1.54 Coffee and Prostate Cancer The September 2011 issue of the *Berkeley Wellness Letter* said that coffee reduces the chance of prostate cancer. A study of 48,000 male health care professionals showed that those consuming the most coffee (six or more cups per day) had a 60% reduced risk of developing advanced prostate cancer. Does this mean that a man can reduce his chance of developing prostate cancer by increasing the amount of coffee he drinks?

1.55 Probation A statistics student conducted a study on young male and female criminals 15 years of age and under who were on probation. The purpose of the study was to see whether there was an association between type of crime and gender. The subjects of the study lived in Ventura County, California. Violent crimes involve physical contact such as hitting or fighting. Nonviolent crimes are vandalism, robbery, or verbal assault. The raw data are shown in the accompanying table; v stands for violent, n for nonviolent, b for boy, and g for girl.

a. Make a two-way table that summarizes the data. Label the columns (across the top) Boy and Girl. Label the rows Violent and Nonviolent.
b. Find the percentage of girls on probation for violent crimes and the percentage of boys on probation for violent crimes, and compare them.

c. Are the boys or the girls more likely to be on probation for violent crimes?

Gen	Viol?
b	n
b	n
b	n
b	n
b	n
b	n
b	n
b	n
b	n
b	n
b	n
b	n
b	n
b	n
b	n

Gen	Viol?
b	n
b	n
b	n
b	n
b	v
b	v
b	v
b	v
b	v
b	v
b	v
b	v
b	v
b	v
g	n

Gen	Viol?
g	n
g	n
g	n
g	v
g	v
g	v
g	v
g	v
g	v
g	v
g	v
g	v
g	v
g	v

1.56 Scorpion Antivenom A study was done on children (6 months to 18 years of age) who had (nonlethal) scorpion stings. Each child was randomly assigned to receive an experimental anti-venom or a placebo. Good results were no symptoms after four hours. Make a summary of the data in the form of a two-way table. Label the columns Antivenom and Placebo. Label the rows Better and Not Better. Compare the percentage better for the antivenom group and the placebo group. (Source: Boyer, Leslie V. et al., Antivenom for critically ill children with neurotoxicity from scorpion stings, *New England Journal of Medicine*, vol. 360: 2090–2098, no. 20, May 14, 2009)

Antivenom	Better
1	1
1	1
1	1
1	0
1	1
1	1
1	1
0	0

Antivenom	Better
0	0
0	0
0	0
0	1
0	0
0	0
1	1

* **1.57 Writing: Vitamin D** Describe the design of a controlled experiment to determine whether the use of vitamin D supplements reduces the chance of broken bones in women with osteoporosis (weak bones). Assume you have 200 women with osteoporosis to work with. Your description should include all the features of a controlled experiment. Also decide how the results would be determined.

* **1.58 Writing: Strokes** People who have had strokes are often put on "blood thinners" such as aspirin or Coumadin to help prevent a second stroke. Describe the design of a controlled experiment to determine whether aspirin or Coumadin works better in preventing second strokes. Assume you have 300 people who have had a first stroke to work with. Include all the features of a good experiment. Also decide how the results would be determined.

1.59 Medicaid Expansion Medicaid is a program administered by the states that provides medical help to low-income residents. Read the extract given, and then answer the questions that follow it. (Source: Sommers et al., Mortality and access to care among adults after state Medicaid expansions, *New England Journal of Medicine*, vol. 367: 1025–1034, July 25, 2012)

> *Methods* We compared three states that had substantially expanded adult Medicaid eligibility since 2000 (New York, Maine, and Arizona) with neighboring states that had not enacted such expansions. The sample consisted of adults between the ages of 20 and 64 years who were observed 5 years before and 5 years after the expansions, from 1997 through 2007. The primary outcome was all-cause county-level mortality rates. . . .
>
> *Results* Medicaid expansions were associated with a significant reduction in adjusted all-cause mortality (by 19.6 deaths per 100,000 adults, for a relative reduction of 6.1%; $P = 0.001$). Mortality rate reductions were greatest among older adults, nonwhites, and residents of poorer counties. . . . Medicaid expansions increased rates of self-reported health status of "excellent" or "very good" (by 2.2 percentage points, for a relative increase of 3.4%; $P = 0.04$).

a. Identify the treatment variable and the response variable.

b. Was this a controlled experiment or an observational study? Explain

c. Can you conclude that the increase in Medicaid caused the good results? Why or why not?

* **1.60 HIV-1 and HIV-2** Does infection with HIV-2 slow the progression to AIDS for those with HIV-1? A blood test is used to determine whether a person is infected with HIV. If a person has AIDS, it means the person is experiencing symptoms that result from the HIV infection. (It is possible to be infected with HIV and not show symptoms.) Read the extract given, and then answer the questions that follow it. (Source: Esbjornsson et al., Inhibition of HIV-1 disease progression by contemporaneous HIV-2 infection, *New England Journal of Medicine*, vol. 367: 224–232, July 19, 2012)

> *Methods* We analyzed data from 223 participants who were infected with HIV-1 after enrollment (with either HIV-1 infection alone or HIV-1 and HIV-2 infection) in a cohort with a long follow-up duration (approximately 20 years), according to whether HIV-2 infection occurred first and the time to the development of AIDS (time to AIDS). . . .
>
> *Results* The median time to AIDS was 104 months in participants with dual infection and 68 months in participants infected with HIV-1 only. . . . Participants with dual infection with HIV-2 infection preceding HIV-1 infection had the longest time to AIDS. . . .

a. Identify the treatment variable and the response variable.

b. Was this a controlled experiment or an observational study? Explain.

c. What do the median times to AIDS show, descriptively?

d. Can you conclude that the presence of HIV-2 slows the progression of HIV to AIDS? Why not?

1.61 Death Row and Head Trauma A study conducted by Lewis et al. in 1986 looked at 14 juveniles awaiting execution. They found that 57% (8 of the 14) had had a serious brain injury. Can we conclude that head trauma causes bad behavior later in life? What primary factor is not present here that should be present in both observational studies and controlled experiments? (Source: Psychiatric, neurological, and psychoeducational characteristics of 15 death row inmates in the United States, *American Journal of Psychiatry*, vol. 143: 838–845. 1986)

1.62 Brief Exercise and Diabetes As part of a study, sixteen young men performed high-intensity exercise that totaled only 15 minutes in a two-week period. At the end of two weeks, several (but not all) tests for diabetes, such as an insulin sensitivity test, showed improvement. Do these results indicate that brief, high-intensity exercise causes an improvement in markers for diabetes? What essential component of both controlled experiments and observational studies is missing from this study? (Source: Babraj et al., Extremely short duration high intensity interval training substantially improves insulin action in young healthy males, *BMC Endocrine Disorders*, vol. 9: 3doi: 10.1186/1472-6823-9-3, January 2009)

GUIDED EXERCISES

g **1.21 Two-way Table from Data** The first step in finding an association between gender and handedness among the students in Table 1A (on page 51) is to create a two-way table with these two variables. Make a two-way table with the labels Male and Female across the top and the labels Right and Left on the side, and then answer the following questions about it. Guidance is given for parts a and b.

a. Report how many are in each cell by using STEP 1 and 2 of the guidance given below.

b. Report the totals by using STEP 3 of the guidance.

c.-g. Go back to the original question on page 53 to see the other parts.

Gender	Hand	Checked
Male	Right	✓
Female	Right	✓
Female	Right	✓
Female	Left	✓
Male	Right	
Female	Right	
Male	Right	
Female	Left	
Female	Right	
Male	Right	
Female	Right	

Guidance

Step 1 ▶ Refer to the part of the spreadsheet given. To make the tallying a bit easier, we have reported the gender without coding. For each cell, make a tally mark (|) for each person who has both of the characteristics belonging to that cell. After making the tally mark, cross off that cell (or put a check mark next to the table on the right) so you will know how far you have gotten. The first four tally marks are given (So far, we have counted one person who is both male and right-handed, two who are female and right-handed, and one who is female and left-handed.)

	Male	Female
Right	|	| |
Left		|

Step 2 ▶ When you have finished tallying, check to see that you have a total of 11 tally marks. Then summarize the table with numbers. Note that one number is given for you so that you can check to ensure that you get the same value.

	Male	Female
Right		5
Left		

Step 3 ▶ Put in the totals: Put the total number of males at the location shown as T_{Male} and put the total number of females at T_{Female}. Put the total number of subjects who are right-handed at T_{Right} and the total number of those who are left-handed at T_{Left}. Note that the grand total (total number of people) is 11, as shown.

	Male	Female	Total
Right		5	T_{Right}
Left			T_{Left}
Total	T_{Male}	T_{Female}	11

Go back to the original question on page 53 for the other parts.

g **1.25 Living with AIDS** The accompanying table gives the numbers of people diagnosed with AIDS/HIV in 2010 in the five states with the largest numbers of cases, as well as the District of Columbia, as reported by the U.S. Centers for Disease Control and Prevention (CDC). It also shows the population of those regions at that time, from the U.S. Census Bureau.

QUESTION Find the number of people diagnosed with AIDS/HIV per thousand in each region, and rank the six regions from highest rate (rank 1) to lowest rate (rank 6). Compare these rankings (of rates) with the ranks of total number of cases. If you moved to one of these regions and met 50 random people, in which region would you be most likely to meet at least one person diagnosed with HIV? In which of these regions would you be least likely to meet at least one person diagnosed with AIDS?

Step 1 ► Figure out the populations of the remaining regions in thousands, and add them to the table.

Step 2 ► For each region, divide the number of people living with AIDS by the population in thousands, and fill in column 6.

Step 3 ► Enter the ranks for the rates of AIDS patients per 1000 population, using 1 for the largest value and 6 for the smallest.

Step 4 ► Are the ranks for the rates the same as the ranks for the numbers of cases? If not, describe at least one difference.

Step 5 ► Finally, if you moved to one of these regions and met 50 random people, in which region would you be most likely to meet at least one person living with HIV? In which region would you be least likely to meet at least one person living with HIV?

State	AIDS	Rank Cases	Population	Population (thousands)	AIDS per 1000 population	Rank Rate
New York	192,753	1	19,421,005	19,421	9.92	2
California	160,293	2	37,341,989	37,342		
Florida	117,612	3	18,900,773			
Texas	77,070	4	25,258,418			
New Jersey	54,557	5	8,807,501			
District of Columbia	9,257	6	601,723			

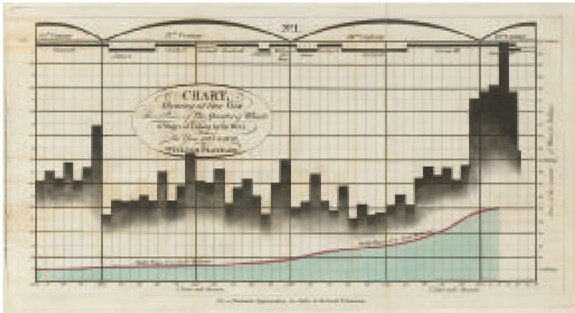
2 Picturing Variation with Graphs

THEME

Any collection of data exhibits variation. The most important tool for organizing this variation is called the distribution of the sample, and visualizing this distribution is the first step in every statistical investigation. We can learn much about a numerical variable by focusing on three components of the distribution: the shape, the center, and the variability, or horizontal spread. Examining a graph of a distribution can lead us to deeper understanding of the situation that produced the data.

One of the major concepts of statistics is that although individual events are hard to predict, large numbers of events usually exhibit predictable patterns. The search for patterns is a key theme in science and business. An important first step in this search is to identify and visualize the key features of your data.

Using graphics to see patterns and identify important trends or features is not new. One of the earliest statistical graphs dates back to 1786, when a Scottish engineer named William Playfair published a paper examining whether there was a relationship between the price of wheat and wages. To help answer this question, Playfair produced a graph (shown in Figure 2.1) that is believed to be the first of its kind. This graph became the prototype of two of the most commonly used tools in statistics: the bar chart and the histogram.


▲ **FIGURE 2.1** Playfair's chart explores a possible relationship between wages and the price of wheat. (Source: Playfair 1786)

Graphics such as these can be extraordinarily powerful ways of organizing data, detecting patterns and trends, and communicating findings. The graphs that we use have changed somewhat since Playfair's day, but graphics are of fundamental importance to analyzing data. The first step in any statistical analysis is to make a picture of some kind in order to check our intuition against the data. If our intuitions are wrong, it could very well be because the world works differently than we thought. Thus, by making and examining a display of data (as illustrated in this chapter's Case Study), we gain some insight into how the world works.

In Chapter 1 we discussed some of the methods used to collect data. In this chapter we'll cover some of the basic graphics used in analyzing the data we collect. Then, in Chapter 3, we'll comment more precisely on measuring and comparing key features of our data.

CASE STUDY

Student-to-Teacher Ratio at Colleges

Are private four-year colleges better than public four-year colleges? That depends on what you mean by "better." One measure of quality that many people find useful (and there are many other ways to measure quality) is the student-to-teacher ratio: the number of students enrolled divided by the number of teachers. For schools with small student-to-teacher ratios, we expect class sizes to be small; students can get extra attention in a small class.

The data in Table 2.1 on the next page were collected from some schools that award four-year degrees. The data are for the 2010–2011 academic year; 89 private colleges and 49 state-supported (public) colleges were sampled. Each ratio was rounded to the nearest whole number for simplicity. For example, the first private college listed has a

Private Colleges								
16	16	19	16	17	22	23	33	19
10	24	13	21	13	19	21	29	41
11	17	2	12	25	14	14	18	22
25	20	18	12	26	39	20	21	21
18	13	6	13	14	27	12	9	29
30	28	19	17	0	17	14	22	21
14	13	22	12	19	27	19	20	15
18	17	60	10	14	31	22	7	16
12	23	4	14	22	19	9	14	20
18	16	8	30	17	8	23	19	

Public Colleges				
16	28	37	23	22
20	28	17	27	59
9	27	19	23	20
6	21	17	23	26
20	17	26	3	15
17	23	11	32	24
22	30	20	19	26
20	18	26	14	15
21	21	25	26	18
23	36	24	21	

▲ **TABLE 2.1** Ratio of students to teachers at private and public colleges. (Source: http://nces.ed.gov/ipeds/)

student-to-teacher ratio of 16, which means that there are about 16 students for every teacher. What differences do you expect between the two groups? What similarities do you anticipate?

It is nearly impossible to compare the two groups without imposing some kind of organization on the data. In this chapter you will see several ways in which we can graphically organize groups of data like these so that we can compare the two types of colleges. At the end of this chapter, you'll see what the right graphical summaries can tell us about how these types of colleges compare.

SECTION 2.1

Visualizing Variation in Numerical Data

One of the most important conceptual tools in statistics and data analysis is the distribution. The **distribution of a sample** of data is simply a way of organizing the data. Think of the distribution of the sample as a list that records (1) the values that were observed and (2) the frequencies of these values. **Frequency** is another word for the *count* of how many times the value occurred in the collection of data.

KEY POINT The distribution of a sample is one of the central organizational concepts of data analysis. The distribution organizes data by recording all of the values observed in a sample, as well as how many times each value was observed.

Distributions are important because they capture much of the information we need in order to make comparisons between groups, examine data for errors, and learn about real-world processes. Distributions enable us to examine the variation of the data in our sample, and from this variation we can often learn more about the world.

The first step of almost every statistical investigation is to visualize the distribution of the sample. By creating an appropriate graphic, we can see patterns that might otherwise escape our notice.

For example, here are some raw data from the National Collegiate Athletic Association (NCAA), available online. This set of data shows the number of goals scored by NCAA female soccer players in Division III in the 2012 season. (Division III schools are colleges or universities that are not allowed to offer scholarships to athletes.) To make the data set smaller, we show only first-year students.

9, 11, 11, 11, 11, 12, 13, 13, 13, 13, 13, 14, 14, 14,
15, 15, 16, 16, 16, 16, 18, 18, 19, 19, 20, 20, 21, 35

This list includes only the values. A distribution lists the values and also the frequencies. The distribution of this sample is shown in Table 2.2.

It's hard to see patterns when the distribution is presented as a table. A picture makes it easier for us to answer questions such as "What's the typical number of goals scored by a player?" and "Is 19 goals an unusually high number?" Data are also available for male soccer players and for other divisions and classes. A picture would make it easier to compare the numbers of goals for different groups. For example, in a season, do men typically score more goals or fewer goals than women?

When examining distributions, we use a two-step process:

1. See it.
2. Summarize it.

In this section we explain how to visualize the distribution. In the next section, we discuss the characteristics you should look for to help you summarize it.

All of the methods we use for visualizing distributions are based on the same idea: Make some sort of mark that indicates how many times each value occurred in our data set. In this way, we get a picture of the sample distribution so that we can see at a glance which values occurred and how often.

Two very useful methods for visualizing distributions of numerical variables are dotplots and histograms. Dotplots are simpler; histograms are more commonly used and perhaps more useful.

Value	Frequency
9	1
11	4
12	1
13	5
14	3
15	2
16	4
18	2
19	2
20	2
21	1
35	1

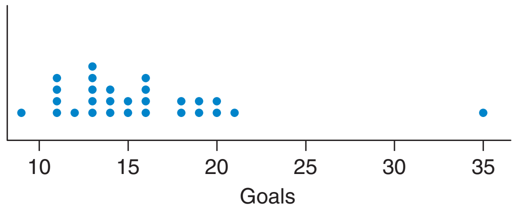
▲ **TABLE 2.2** Distribution of the number of goals scored by first-year women soccer players in NCAA Division III in 2012.

Details

Making Dotplots
Dotplots are easy to make with pen and paper, but don't worry too much about recording values to great accuracy. The purpose of a plot like this is to help us see the overall shape of the distribution, not to record details about individual observations.

Dotplots

In constructing a **dotplot**, we simply put a dot above a number line where each value occurs. We can get a sense of frequency by seeing how high the dots stack up. Figure 2.2 shows a dotplot for the number of goals for first-year female soccer players.



Tech

◀ **FIGURE 2.2** Dotplot of the number of goals scored by first-year women soccer players in NCAA Division III, 2012. Each dot represents a soccer player. Note that the horizontal axis begins at 9.

With this simple picture, we can see more than we could from Table 2.2. We can see from this dotplot that most women scored 20 or fewer goals in the 2012 season, but a couple scored more. Also, we can see that the woman who scored 35 goals was exceptional within this group. Not only is 35 the largest value in this data set, but it stands apart from the others by a large gap.

SNAPSHOT THE DOTPLOT

WHAT IS IT? ► A graphical summary.

WHAT DOES IT DO? ► Shows a picture of the distribution of a numerical variable.

HOW DOES IT DO IT? ► Each observation is represented by a dot on a number line.

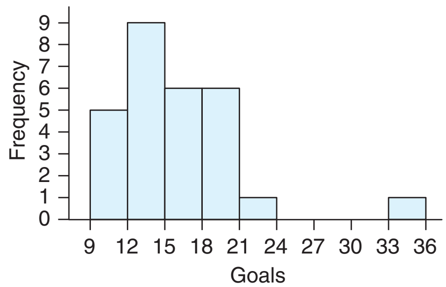
HOW IS IT USED? ► To see patterns in samples of data with variation.

Histograms

While dotplots have one dot for each observation in the data set, **histograms** produce a smoother graphic by grouping observations into intervals, called bins. These groups are formed by dividing the number line into bins of equal width and then counting how many observations fall into each bin. To represent the bins, histograms display vertical bars, where the height of each bar is proportional to the number of observations inside that bin.

For example, with the goals scored in the dotplot in Figure 2.2, we could create a series of bins that go from 9 to 12, 12 to 15, 15 to 18, and so on. Five women scored between 9 and 12 goals during the season, so the first bar has a height of 5. The second bin contains nine observations and consequently has a height of 9. The finished graph is shown in Figure 2.3. (Note that some statisticians use the word *interval* in place of *bin*. You might even see another word that means the same thing: *class*.) Figure 2.3 shows, among other things, that six women scored between 15 and 18 goals, six scored between 18 and 21 goals, and so on.

Tech

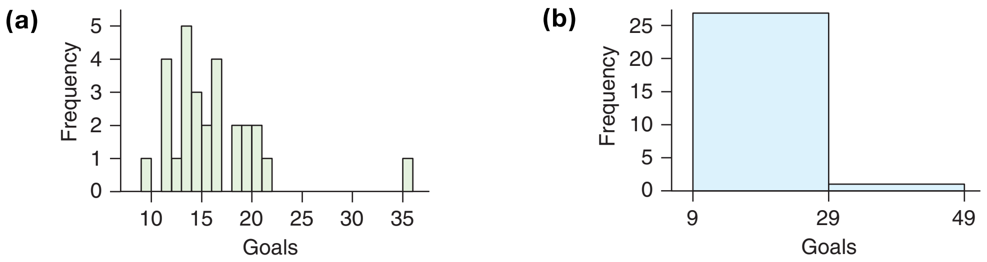


► **FIGURE 2.3** Histogram of goals for female first-year soccer players in NCAA Division III, 2012. The first bar, for example, tells us that five players scored between 9 and 12 goals during the season.

Making a histogram requires paying attention to quite a few details. For example, we need to decide on a rule for what to do if an observation lands exactly on the boundary of two bins. In which bin would we place an observation of 12 goals? A common rule is to decide always to put "boundary" observations in the bin on the right, but we could just as well decide always to put them in the bin on the left. The important point is to be consistent. The graphs here use the right-hand rule and put boundary values in the bin to the right.

Figure 2.4 shows two more histograms of the same data. Even though they display the same data, they look very different—both from each other and from Figure 2.3. Why?

Changing the width of the bins in a histogram changes its shape. Figure 2.3 has bins with width of 3 goals. In contrast, Figure 2.4a has much smaller bins, and Figure 2.4b has wider bins. Note that when we use small bins, we get a spiky histogram. When we use wider bins, the histogram gets less spiky. Using wide bins hides more detail. If you chose very wide bins, you would have no details at all. You would see just one big rectangle!

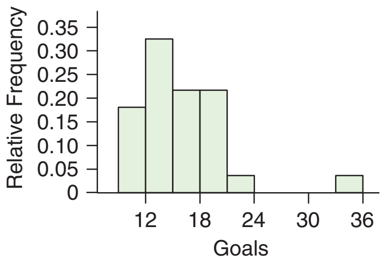


◀ **FIGURE 2.4** Two more histograms of goals scored in one season the same data as in Figure 2.3. **(a)** This histogram has narrow bins and is spiky. **(b)** This histogram has wide bins and offers less detail.

How large should the bins be? Too small and you see too much detail (as in Figure 2.4a). Too large and you don't see enough (as in Figure 2.4b). Most computer software will automatically make a good choice. Our software package (StatCrunch, in this case) automatically chose a binwidth of 3. Still, if you can, you should try different sizes to see how different choices change your impression of the distribution of the sample. Fortunately, most statistical software packages make it quite easy to change the bin width.

Relative Frequency Histograms A variation on the histogram (and statisticians, of course, love variation) is to change the units of the vertical axis from frequencies to relative frequencies. A **relative frequency** is simply a proportion. So instead of reporting that the first bin had 5 observations in it, we would report that the proportion of observations in the first bin was $5/28 = 0.18$. We divide by 28 because there were a total of 28 observations in the data set. Figure 2.5 is the same as the first histogram shown for the distribution of goals (Figure 2.3); however, Figure 2.5 reports relative frequencies, and Figure 2.3 reports frequencies.

Using relative frequencies does not change the shape of the graph; it just communicates different information to the viewer. Rather than answering the question "How many players scored between 9 and 12 goals?" (5 players), it now answers the question "What *proportion* of players scored between 9 and 12 goals?" (0.18).



◀ **FIGURE 2.5** Relative frequency histogram of goals scored by first-year women soccer players in NCAA Division III, 2012.

SNAPSHOT THE HISTOGRAM

WHAT IS IT? ▶ A graphical summary for numerical data.

WHAT DOES IT DO? ▶ Shows a picture of the distribution of a numerical variable.

HOW DOES IT DO IT? ▶ Observations are grouped into bins, and bars are drawn to show how many observations (or what proportion of observations) lie in each bin.

HOW IS IT USED? ▶ By smoothing over details, histograms help our eyes pick up more important, large-scale patterns. Be aware that making the bins wider hides detail, and making the bins smaller can show too much detail. The vertical axis can display frequency, relative frequency, or percents.

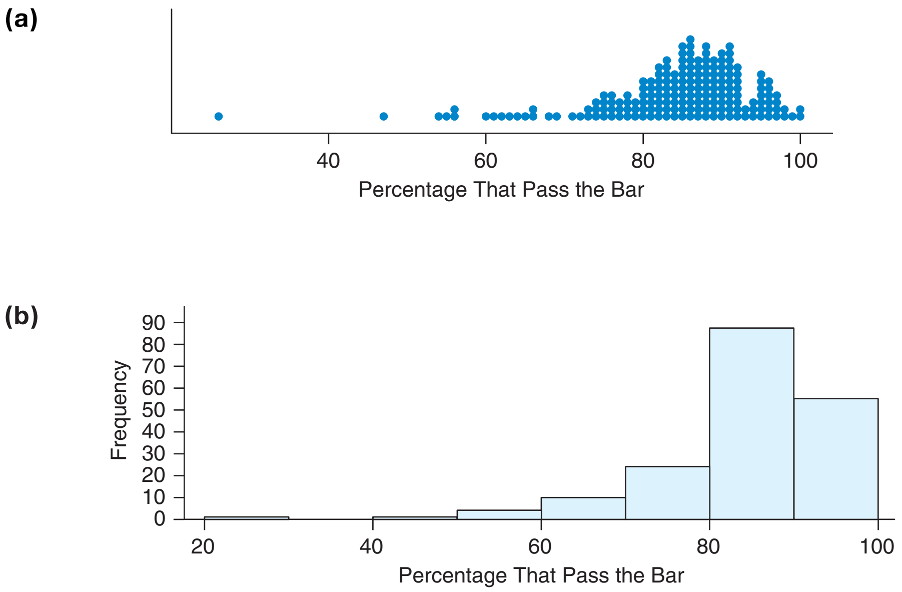
EXAMPLE 1 Visualizing Bar Exam Pass Rates at Law Schools

In order to become a lawyer, you must pass your state's bar exam. When you are choosing a law school, it therefore makes great sense to choose one that has a high percentage of graduates passing the bar exam. The Internet Legal Research Group website provides the pass rate for 184 law schools in the United States in 2009 (Internet Legal Research Group 2013).

QUESTION Is 80% a good pass rate for a law school?

SOLUTION This is a subjective question. An 80% pass rate might be good enough for a prospective student, but another way of looking at this is to determine whether there are many schools that do better than 80%, or whether 80% is a typical pass rate, or whether there are very few schools that do so well. Questions such as these can often be answered by considering the distribution of our sample of data, so our first step is to choose an appropriate graphical presentation of the distribution.

Either a dotplot or a histogram would show us the distribution. Figure 2.6a shows a dotplot generated by Minitab. It's somewhat difficult (but not impossible) to read a dotplot with 184 observations; there's too much detail to allow us to answer broad questions such as these. It is easier to use a histogram, as in Figure 2.6b. In this histogram, each bar has a width of 10 percentage points, and the y-axis tells us how many law schools had a pass rate within those limits.



▶ **FIGURE 2.6** **(a)** A dotplot for the bar-passing rate for 184 law schools. Each dot represents a law school, and the dot's location indicates the bar-passing rate for that school. **(b)** A histogram shows the same data as in part (a), except that the details have been smoothed.

We see that about 85 law schools had a pass rate between 80% and 90%. Another 55 or so had a pass rate over 90%. Adding these together, 140 schools ($85 + 55$) had a pass rate higher than 80%, and this is $140/184 = 0.761$, or about 76% of the schools. This tells us that a pass rate of 80% is not all that unusual; most law schools do this well or better.

TRY THIS! Exercise 2.5

Stemplots

Tech

Stemplots, which are also called stem-and-leaf plots, are useful for visualizing numerical variables when you don't have access to technology and the data set is not large. Stemplots are also useful if you want to be able to easily see the actual values of the data.

To make a **stemplot**, divide each observation into a "stem" and a "leaf." The **leaf** is the last digit in the observation. The **stem** contains all the digits that precede the leaf. For the number 60, the *6* is the stem and the *0* is the leaf. For the number 632, the *63* is the stem and the *2* is the leaf. For the number 65.4, the *65* is the stem and the *4* is the leaf.

A stem-and-leaf plot can help us understand data such as drinking behaviors. Alcohol is a big problem at many colleges and universities. For this reason, a collection of college students who said that they drink alcohol were asked how many alcoholic drinks they had consumed in the last seven days. Their answers were

1, 1, 1, 1, 1, 2, 2, 2, 3, 3, 3, 3, 3, 4, 5, 5, 5, 6, 6, 6, 8, 10, 10, 15, 17, 20, 25, 30, 30, 40

For one-digit numbers, imagine a 0 at the front. The observation of 1 drink becomes 01, the observation of 2 drinks becomes 02, and so on. Then each observation is just two digits; the first digit is the stem, and the last digit is the leaf. Figure 2.7 shows a stemplot of these data.

Stem	Leaves
0	111112223333345556668
1	0057
2	05
3	00
4	0

▲ **FIGURE 2.7** A stemplot for alcoholic drinks consumed by college students. Each digit on the right (the leaves) represents a student. Together, the stem and the leaf indicate the number of drinks for an individual student.

If you rotate a stemplot 90 degrees counterclockwise, it does not look too different from a histogram. Unlike histograms, stemplots display the actual values of the data. With a histogram, you know only that the values fall somewhere within an interval.

Stemplots are often organized with the leaves in order from lowest to highest, which makes it easier to locate particular values, as in Figure 2.7. This is not necessary, but it makes the plot easier to use.

From the stemplot, we see that most students drink a moderate amount in a week but that a few drink quite a bit. Forty drinks per week is almost six per day, which qualifies as problem drinking by some physicians' definitions.

Figure 2.8 shows a stemplot of some exam scores. Note the empty stems at 4 and 5, which show that there were no exam grades between 40 and 59. Most of the scores are between 60 and 100, but one student scored very low relative to the rest of the class.

Stem	Leaves
3	8
4	
5	
6	0257
7	00145559
8	0023
9	0025568
10	00

▲ **FIGURE 2.8** A stemplot for exam grades. Two students had scores of 100, and no students scored in the 40's or 50's.

SNAPSHOT **THE STEMPLOT**

WHAT IS IT? ► A graphical summary for numerical data.

WHAT DOES IT DO? ► Shows a picture of the distribution of a numerical variable.

HOW DOES IT DO IT? ► Numbers are divided into leaves (the last digit) and stems (the preceding digits). Stems are written in a vertical column, and associated leaves are "attached."

HOW IS IT USED? ► In very much the same way as a histogram. It is often useful when technology is not available and the data set is not large.

SECTION 2.2

Summarizing Important Features of a Numerical Distribution

When examining a distribution, pay attention to describing the shape of the distribution, the **typical value (center)** of the distribution, and the **variability (spread)** in the distribution. The typical value is subjective, but the location of the center of a distribution often gives us an idea of which values are typical for this variable. The variability is reflected in the amount of horizontal spread the distribution has.

KEY POINT When examining distributions of numerical data, pay attention to the shape, center, and horizontal spread.

Figure 2.9 compares distributions for two groups. You've already seen histogram (a)—it's the histogram for the goals scored in 2012 by first-year women soccer players in Division III. Histogram (b) shows goals scored for first-year male soccer players in Division III in the same year. How do these two distributions compare?

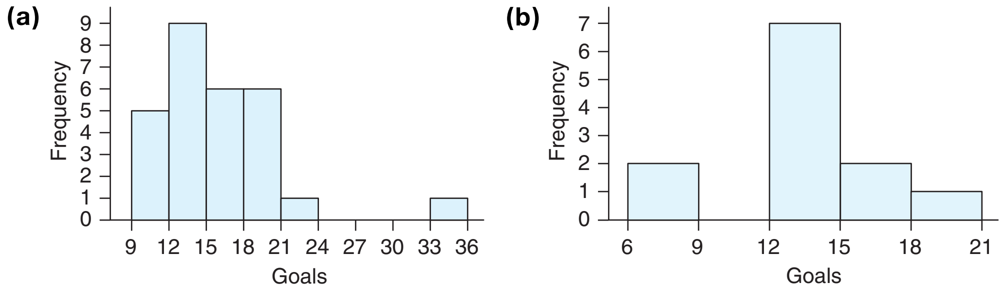


FIGURE 2.9 Distributions of the goals scored for **(a)** first-year women and **(b)** first-year men in Division III soccer in 2012.

1. *Shape*. Are there any interesting or unusual features about the distributions? Are the shapes very different? (If so, this might be evidence that men play the game differently than women.)
2. *Center*. What is the typical value of each distribution? Is the typical number of goals scored per game different for men than for women?
3. *Spread*. The horizontal spread presents the variation in goals per game for each group. How do the amounts of variation compare? If one group has low variation, it suggests that the soccer skills of the members are pretty much the same. Lots of variability might mean that there is a wider variety of skill levels.

Let's consider these three aspects of a distribution one at a time.

Details

Vague Words
For now, we have left the three ideas *shape*, *center*, and *spread* deliberately vague. Indeed, there are ways of measuring more precisely where the center is and how much spread exists. However, the first task in a data analysis is to examine a distribution to informally evaluate the shape, center, and spread.

Shape

You should look for three basic characteristics of a distribution's shape:

1. Is the distribution symmetric or skewed?
2. How many mounds appear? One? Two? None? Many?
3. Are unusually large or small values present?

Symmetric or Skewed? A symmetric distribution is one in which the left-hand side of the graph is roughly a mirror image of the right-hand side. The idealized distributions in Figure 2.10 show two possibilities. Figure 2.10a is a **symmetric distribution** with one mound. (Statisticians often describe a distribution with this particular shape as a **bell-shaped distribution**. Bell-shaped distributions play a major role in statistics, as you will see throughout this text.)

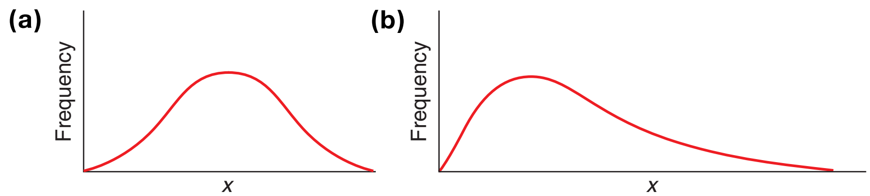


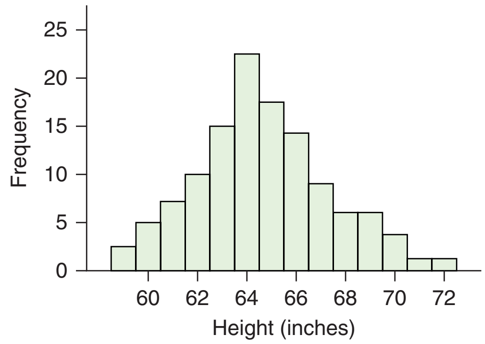
FIGURE 2.10 Sketches of **(a)** a symmetric distribution and **(b)** a right-skewed distribution.

Figure 2.10b represents a nonsymmetric distribution with a skewed shape that has one mound. Note that it has a "tail" that extends out to the right (toward larger values). Because the tail goes to the right, we call it a **right-skewed distribution**. This is a typical shape for the distribution of a variable in which most values are relatively small but there are also a few very large values. If the tail goes to the left, it is a **left-skewed distribution**, where most values are relatively large but there are also a few very small values.

Figure 2.11 shows a histogram of 123 college women's heights. How would you describe the shape of this distribution? This is a good real-life example of a symmetric distribution. Note that it is not perfectly symmetric, but you will never see "perfect" in real-life data.

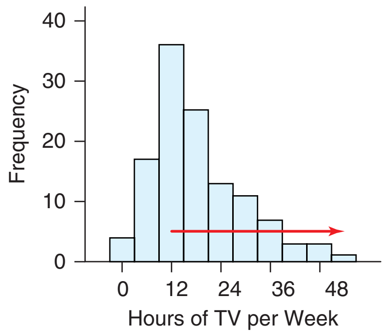
Details

Positive and Negative Skew
You may see some researchers use the terms *positively-skewed* instead of *right-skewed* and *negatively-skewed* instead of *left-skewed*.

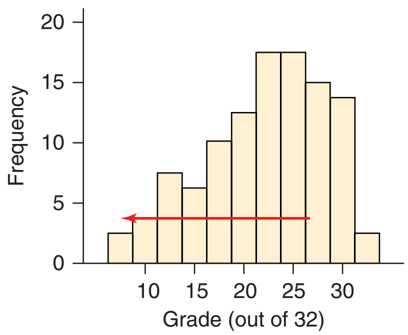


◀ **FIGURE 2.11** Histogram of heights of women. (Source: Brian Joiner in Tufte 1983)

Suppose we asked a sample of people how many hours of TV they watched in a typical week. Would you expect a histogram of these data to be bell-shaped? Probably not. The smallest possible value for this data set would be 0, and most people would probably cluster near a common value. However, a few people probably watch quite a bit more TV than most other people. Figure 2.12 shows the actual histogram. We've added an arrow to emphasize that the tail of this distribution points to the right; this is a right-skewed distribution.



▲ **FIGURE 2.12** This data set on TV hours viewed per week is skewed to the right. (Source: Minitab Program)



▲ **FIGURE 2.13** This data set on test scores is skewed to the left.

Figure 2.13 shows a left-skewed distribution of test scores. This is the sort of distribution that you should hope your next exam will have. Most people scored pretty high, so the few people who scored low create a tail on the left-hand side. A very difficult test, one in which most people scored very low and only a few did well, would be right-skewed.

Another circumstance in which we often see skewed distributions is when we collect data on people's income. When we graph the distribution of incomes of a large group of people, we quite often see a right-skewed distribution. You can't make less than 0 dollars per year. Most people make a moderate amount of money, but there is no upper limit on how much a person can make, and a few people in any large sample will make a very large amount of money.

Example 2 shows that you can often make an educated guess about the shape of a distribution even without collecting data.

EXAMPLE 2 Roller Coaster Endurance

A morning radio show is sponsoring a contest in which contestants compete to win a car. About 40 contestants are put on a roller coaster, and whoever stays on it the longest wins. Suppose we make a histogram of the amount of time the contestants stay on (measured in hours or even days).

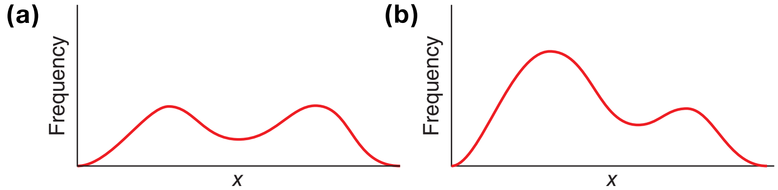
QUESTION What shape do we expect the histogram to have and why?

SOLUTION Probably most people will drop out relatively soon, but a few will last for a very long time. The last two contestants will probably stay for a very long time indeed. Therefore, we would expect the distribution to be right-skewed.

TRY THIS! Exercise 2.9

How Many Mounds? What do you think would be the shape of the distribution of heights if we included men in our sample as well as women? The distributions of women's heights by themselves and men's heights by themselves are usually symmetric and have one mound. But because we know that men tend to be taller than women, we might expect a histogram that combines men's and women's heights to have two mounds.

The statistical term for a one-mound distribution is **unimodal distribution**, and a two-mound distribution is called a **bimodal distribution**. Figure 2.14a shows a bimodal distribution. A **multimodal distribution** has more than two modes. The modes do not have to be the same height (in fact, they rarely are). Figure 2.14b is perhaps the more typical bimodal distribution.



▶ **FIGURE 2.14** Idealized bimodal distributions. **(a)** Modes of roughly equal height. **(b)** Modes that differ in height.

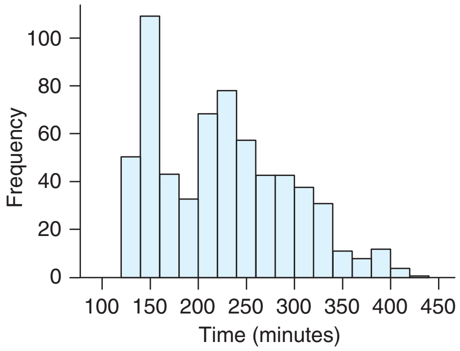
These sketches are idealizations. In real life you won't see distributions this neat. You will have to make a decision about whether a histogram is close enough to be called symmetric and whether it has one mound, two mounds, no mounds, or many mounds. The existence of multiple mounds is sometimes a sign that very different groups have been combined into a single collection (such as combining men's heights with women's heights). When you see multimodal distributions, you may want to look back at the original data and see whether you can examine the groups separately, if separate groups exist. At the very least, whenever you see more than one mound, you should ask yourself, "Could these data be from different groups?"

EXAMPLE 3 Two Marathons, Merged

Data were collected on the finishing times for two different marathons. One marathon consisted of a small number of elite runners: the 2012 Olympic Games. The other marathon included a large number of amateur runners: a marathon in Portland, Oregon.

QUESTION What shape would you expect the distribution of this sample to have?

SOLUTION We expect the shape to be bimodal. The elite runners would tend to have faster finishing times, so we expect one mound on the left for the Olympic runners and another mound on the right. Figure 2.15 is a histogram of the data.

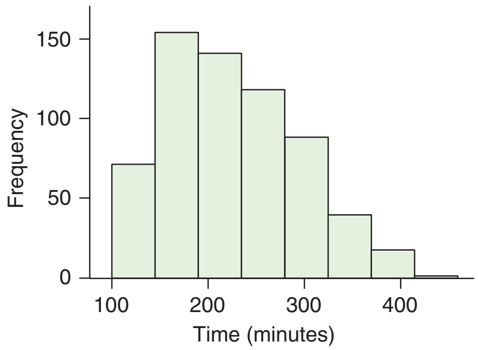


◀ **FIGURE 2.15** Histogram of finishing times for two marathons.

There appears to be one mound centered at about 150 minutes (2.5 hours) and another centered at about 250 minutes (about 4.1 hours).

TRY THIS! Exercise 2.11

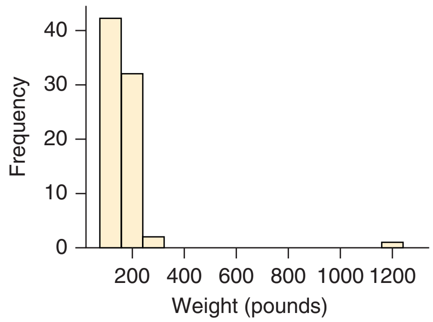
When we view a histogram, our understanding of the shape of a distribution is affected by the width of the bins. Figure 2.15 reveals the bimodality of the distribution partly because the width of the bins is such that we see the right level of detail. If we had made the bins too big, we would have got less detail and might not have seen the bimodal structure. Figure 2.16 shows what would happen. Experienced data analysts usually start with the bin width the computer chooses and then also examine histograms made with slightly wider and slightly narrower bins. This lets them see whether any interesting structure emerges.



◀ **FIGURE 2.16** Another histogram of the same running times as in Figure 2.15. Here, the bins are wider and "wash out" detail, so the distribution no longer looks bimodal, even though it should.

Do Extreme Values Occur? Sometimes when you make a histogram or dotplot, you see extremely large or extremely small observations. When this happens, you should report these values and, if necessary, take action (although *reporting* is often action enough). Extreme values can appear when an error is made in entering data. For example, we asked students in one of our classes to record their weights in pounds. Figure 2.17 shows the distribution. The student who wrote 1200 clearly made a mistake. He or she probably meant to write 120.

Extreme values such as these are called **outliers**. The term *outlier* has no precise definition. Sometimes you may think an observation is an outlier, but another person might disagree, and this is fine. However, if there is no gap between the observation in question and the bulk of the histogram, then the observation probably should not be considered an outlier. Outliers are points that don't fit the pattern of the rest of the data, so if a large percentage of the observations are extreme, it might not be accurate to label them as outliers. After all, if lots of points don't fit a pattern, maybe you aren't seeing the right pattern!



▶ **FIGURE 2.17** Histogram of weights with an extreme value.

KEY POINT Outliers are values so large or small that they do not fit into the pattern of the distribution. There is no precise definition of the term *outlier*. Outliers can be caused by mistakes in data entry, but genuine outliers are sometimes unusually interesting observations.

Sometimes outliers result from mistakes, and sometimes they do not. The important thing is to make note of them when they appear and to investigate further, if you can. You'll see in Chapter 3 that the presence of outliers can affect how we measure center and spread.

Center

An important question to ask about any data set is "What is the typical value?" The typical value is the one in the center, but we use the word *center* here in a deliberately vague way. We might not all agree on precisely where the center of a graph is. The idea here is to get a rough impression so that we can make comparisons later. For example, judging on the basis of the histogram shown in Figure 2.9, the center for the women soccer players is about 16 goals. Thus we could say that the typical first-year woman soccer player scored about 16 goals in 2012. In contrast, the center of the distribution for the male soccer players is about 13 goals. It would seem that the typical male soccer player scores fewer goals in a season than the typical female player, perhaps indicating that men's soccer is stronger on defense than women's soccer, at least among Division III first-year players.

If the distribution is bimodal or multimodal, it may not make sense to seek a "typical" value for a data set. If the data set combines two very different groups, then it might be more useful to find separate typical values for each group. What is the typical finishing time of the runners in Figure 2.15? There is no single typical time, because there are two distinct groups of runners. The elite group have their typical time, and the amateurs have a different typical time. However, it *does* make sense to ask about the typical test score for the student scores in Figure 2.13, because there is only one mound and only one group of students.

! Caution

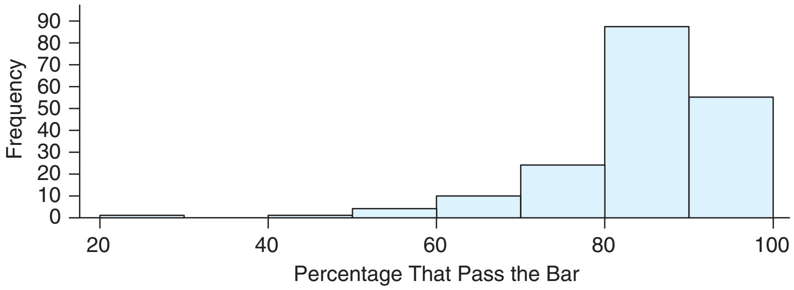
Multimodal Centers
Be careful about giving a single typical value for a multimodal or bimodal distribution. These distributions sometimes indicate the existence of different and diverse groups combined into the same data set. It may be better to report a different typical value for each group, if possible.

EXAMPLE 4 Typical Bar-Passing Rate for Law Schools

Examine the distribution of bar-passing rates for law schools in the United States (see Figure 2.6b, which is repeated here for convenience).

QUESTION What is a typical bar-passing rate for a law school?

SOLUTION We show Figure 2.6b again. The center is somewhere in the range of 80% to 90% passing, so we would say that the typical bar-passing rate is between 80% and 90%.



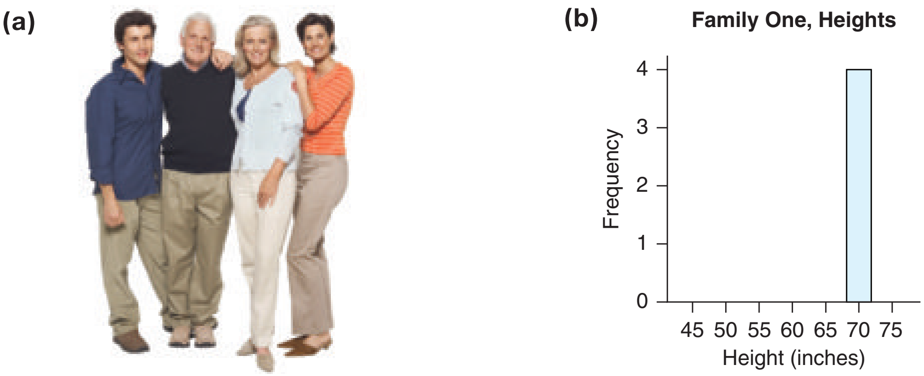
◀ **FIGURE 2.6b (repeated)** Percentage passing the bar exam from different law schools.

TRY THIS! Exercise 2.13

Variability

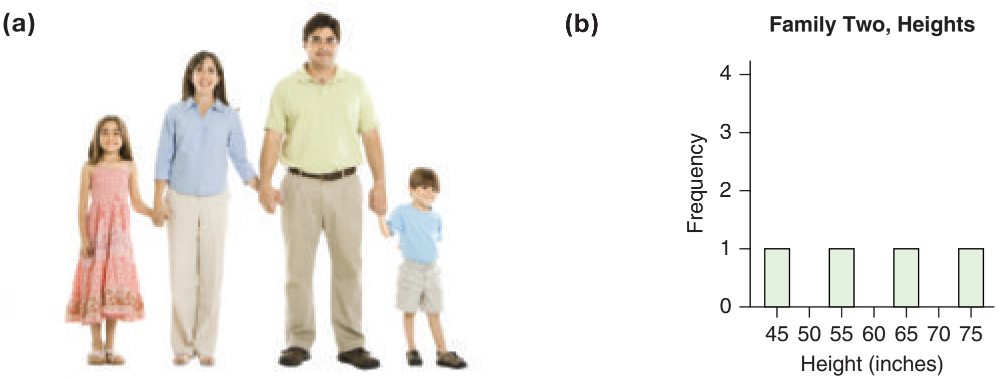
The third important feature to consider when examining a distribution is the amount of variation in the data. If all of the values of a numerical variable are the same, then the histogram (or dotplot) will be skinny. On the other hand, if there is great variety, the histogram will be spread out, thus displaying greater variability.

Here's a very simple example. Figure 2.18a shows a family of four people who are all very similar in height. Note that the histogram of these heights (Figure 2.18b) is quite skinny; in fact, it is just a single bar!



▲ **FIGURE 2.18** Family One with a Small Variation in Height

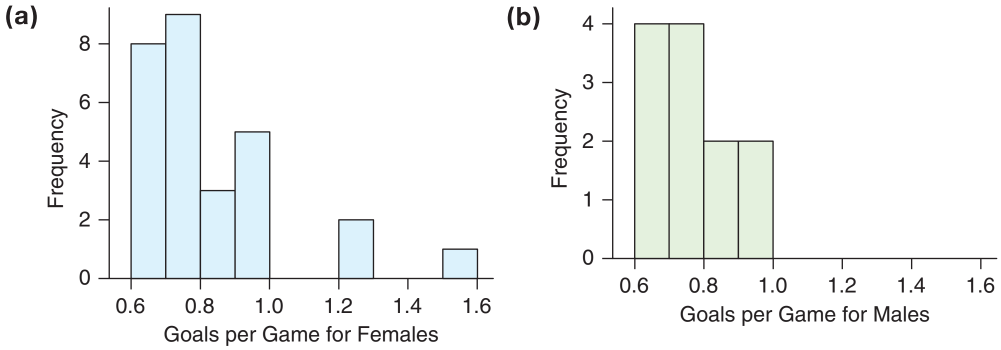
Figure 2.19a, on the other hand, shows a family that exhibits large variation in height. The histogram for this family (Figure 2.19b) is more spread out.



▲ **FIGURE 2.19** Family Two with a Large Variation in Height

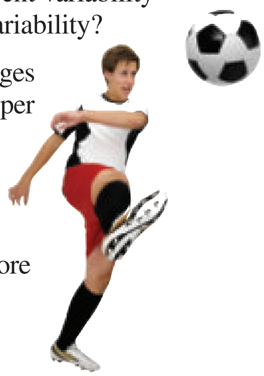
EXAMPLE 5 NCAA Soccer Players

Soccer players who play more games have more opportunities to score goals, so it isn't quite fair to simply look at the number of goals scored in a season. Instead, consider Figure 2.20, which shows the number of goals per game (the number of goals a player scores divided by the number of games that she or he played) for women and men first-year players.



▶ **FIGURE 2.20** Distribution of goals per game scored during one season for **(a)** women and **(b)** men.

QUESTION Do these men and women soccer players have different variability in the goals scored per game? If so, which group has the greater variability?



SOLUTION The number of goals scored per game for women ranges from about 0.6 to 1.6 goals per game; that's a difference of 1.0 goal per game. Goals scored per game by men range from about 0.6 to 1.0, a range of 0.4 goal per game. We see that the women have more variability in the number of goals scored per game. One interpretation of this difference in spread is that the women who play soccer exhibit a wider range of skill levels. In Chapter 3 you'll see some more precise ways of measuring the spread of a numerical variable.

TRY THIS! Exercise 2.15

Describing Distributions

When you are asked to describe a distribution or to compare two distributions, your description should include the center of the distribution (What is the typical value?), the spread (How much variability is there?), and the shape. If the shape is bimodal, it might not be appropriate to mention the center, and you should instead identify the approximate location of the mounds. You should also mention any unusual features, such as extreme values.

EXAMPLE 6 Body Piercings

How common are body piercings among college students? How many piercings does a student typically have? One statistics professor asked a large class of students to report (anonymously) the number of piercings they possessed.

QUESTION Describe the distribution of body piercings for students in a statistics class.

SOLUTION The first step is to "see it" by creating an appropriate graphic of the distribution of the sample. Figure 2.21 shows a histogram for these data. To summarize this distribution, we examine the shape, center, and spread and comment on any unusual features.

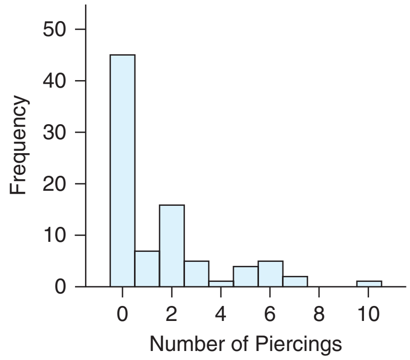


FIGURE 2.21 Numbers of piercings of students in a statistics class.

The distribution of piercings is right-skewed, as we might expect, given that many people will have 0, 1, or 2 piercings but a few people are likely to have more. The typical number of piercings (the center of the distribution) seems to be about 1, although a majority of students have none. The number of piercings ranges from 0 to about 10. An interesting feature of this distribution is that it appears to be multimodal. There are three peaks: at 0, 2, and 6, which are even numbers. This makes sense, because piercings often come in pairs. But why is there no peak at 4? (The authors do not know.) What do you think the shape of the distribution would look like if it included only the men? Only the women?

TRY THIS! Exercise 2.17

SECTION 2.3

Visualizing Variation in Categorical Variables

When visualizing data, we treat categorical variables in much the same way as numerical variables. We visualize the distribution of the categorical variable by displaying the values (categories) of the variable and the number of times each value occurs.

To illustrate, consider the Statistics Department at UCLA. UCLA offers an introductory statistics course every summer, and it needs to understand what sorts of students are interested in this class. In particular, understanding whether the summer students are mostly first-year students (eager to complete their general education requirements) or seniors (who put off the class as long as they could) can help the department better plan its course offerings.

Table 2.3 shows data from a sample of students in an introductory course offered during the 2013 summer term at UCLA. The "unknown" students are probably not enrolled in any university (adult students taking the course for business reasons or high school students taking the class to get a head start).

Class is a categorical variable. Table 2.4 on the next page summarizes the distribution of this variable by showing us all of the values in our sample and the frequency with which each value appears. Note that we added a row for first-year students.

Two types of graphs that are commonly used to display the distribution of a sample of categorical data are bar charts and pie charts. Bar charts look, at first glance, very similar to histograms, but they have several important differences, as you will see.

Student ID	Class
1	Senior
2	Junior
3	Unknown
4	Unknown
5	Senior
6	Graduate
7	Senior
8	Senior
9	Unknown
10	Unknown
11	Sophomore
12	Junior
13	Junior
14	Sophomore
15	Unknown
16	Senior
17	Unknown
18	Unknown
19	Sophomore
20	Junior

TABLE 2.3 Identification of classes for students in statistics.

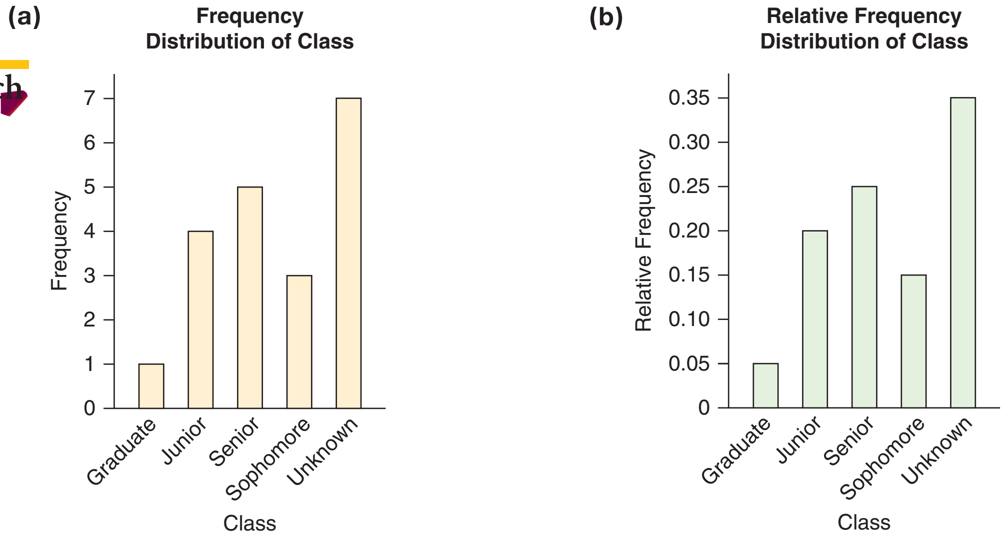
Bar Charts

A **bar graph** (also called a **bar chart**) shows a bar for each observed category. The height of the bar is proportional to the frequency of that category. Figure 2.22a on the next page shows a bar graph for the UCLA statistics class data. The vertical axis

Class	Frequency
Unknown	7
First-year student	0
Sophomore	3
Junior	4
Senior	5
Graduate	1
Total	**20**

▲ **TABLE 2.4** Summary of classes for students in statistics.

Tech



▲ **FIGURE 2.22** **(a)** Bar chart showing numbers of students in each class enrolled in an introductory statistics section. The largest "class" is the group made up of seven unknowns. First-year students are not shown because there were none in the data set. **(b)** The same information as shown in part (a), but now with relative frequencies. The unknowns are about 0.35 (35%) of the sample.

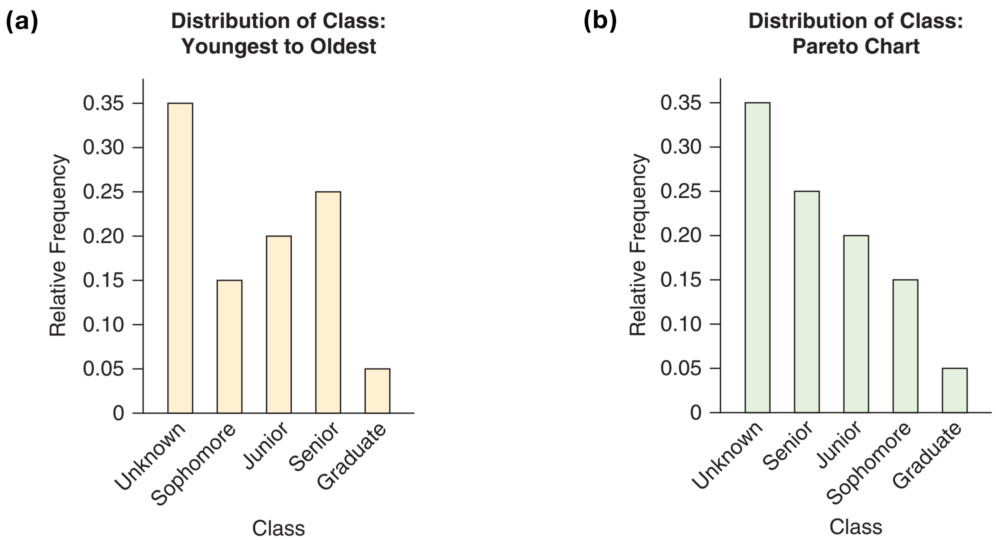
measures frequency. We see that the sample has one graduate student and four juniors. We could also display relative frequency if we wished (Figure 2.22b). The shape does not change; only the numbers on the vertical axis change.

Note that there are no first-year students in the sample. We might expect this of a summer course, because entering students are unlikely to take courses in the summer before they begin college, and students who have completed a year of college are generally no longer first-year students (assuming they have completed enough units).

Bar Charts vs. Histograms Bar charts and histograms look a lot alike, but they have some very important differences.

- In a bar chart, it sometimes doesn't matter in which order you place the bars. Quite often, the categories of a categorical variable have no natural order. If they do have a natural order, you might want to sort them in that order. For example, in Figure 2.23a we've sorted the categories into a fairly natural order,

► **FIGURE 2.23** **(a)** Bar chart of classes using natural order. **(b)** Pareto chart of the same data. Categories are ordered with the largest frequency on the left and arranged so the frequencies decrease to the right.



from "Unknown" to "Graduate." In Figure 2.23b we've sorted them from most frequent to least frequent. Either choice is acceptable.

Figure 2.23b is interesting because it shows more clearly that the most populated category consists of "Unknown." This result suggests that there might be a large demand for this summer course outside of the university. Bar charts that are sorted from most frequent to least frequent are called **Pareto charts**. (These charts were invented by the Italian economist and sociologist Vilfredo Pareto, 1848–1943.) They are often an extremely informative way of organizing a display of categorical data.

- Another difference between histograms and bar charts is that in a bar chart, it doesn't matter how wide or narrow the bars are. The widths of the bars have no meaning.
- A final important difference is that a bar chart has gaps between the bars. This indicates that it is impossible to have observations between the categories. In a histogram, a gap indicates that no values were observed in the interval represented by the gap.

SNAPSHOT THE BAR CHART

WHAT IS IT? ► A graphical summary for categorical data.

WHAT DOES IT DO? ► Shows a picture of the distribution of a categorical variable.

HOW DOES IT DO IT? ► Each category is represented by a bar. The height of the bar is proportional to the number of times that category occurs in the data set.

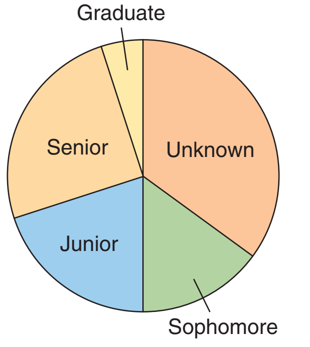
HOW IS IT USED? ► To see patterns of variation in categorical data. The categories can be presented in order of most frequent to least frequent, or they can be arranged in another meaningful order.

Pie Charts

Pie charts are another popular format for displaying relative frequencies of data. A **pie chart** looks, as you would expect, like a pie. The pie is sliced into several pieces, and each piece represents a category of the variable. The area of the piece is proportional to the relative frequency of that category. The largest piece in the pie in Figure 2.24 belongs to the category "Unknown" and takes up about 35% of the total pie.

Some software will label each slice of the pie with the percentage occupied. This isn't always necessary, however, because a primary purpose of the pie chart is to help us judge how frequently categories occur relative to one another. For example, the pie chart in Figure 2.24 shows us that "Unknown" occupies a fairly substantial portion of the whole data set. Also, labeling each slice gets cumbersome and produces cluttered graphs if there are many categories.

Although pie charts are very common (we bet that you've seen them before), they are not commonly used by statisticians or in scientific settings. One reason for this is that the human eye has a difficult time judging how much area is taken up by the wedge-shaped slices of the pie chart. Thus, in Figure 2.24, the "Sophomore" slice looks only slightly smaller than the "Junior" slice. But you can see from the bar chart (Figure 2.23) that they're actually noticeably different. Pie charts are extremely difficult to use to compare the distribution of a variable across two different groups (such as comparing males and females). Also, if there are many different categories, it can be difficult to provide easy-to-read labels for the pie charts.



▲ **FIGURE 2.24** Pie chart showing the distribution of the categorical variable *Class* in a statistics course.

SNAPSHOT THE PIE CHART

WHAT IS IT? ► A graphical summary for categorical data.

WHAT DOES IT DO? ► Shows the proportion of observations that belong to each category.

HOW DOES IT DO IT? ► Each category is represented by a wedge in the pie. The area of the wedge is proportional to the relative frequency of that category.

HOW IS IT USED? ► To understand which categories are most frequent and which are least frequent. Sometimes it is useful to label each wedge with the proportion of occurrence.

SECTION 2.4

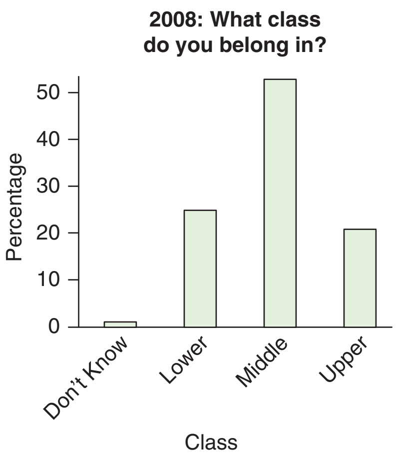
Summarizing Categorical Distributions

The concepts of *shape*, *center*, and *spread* that we used to summarize numerical distributions sometimes don't make sense for categorical distributions, because we can often order the categories any way we please. The center and shape would be different for every ordering of categories. However, we can still talk about typical outcomes and the variability in the sample.

The Mode

When describing a distribution of a categorical variable, pay attention to which category occurs most often. This value, the one with the tallest bar in the bar chart, can sometimes be considered the "typical" outcome. There might be a tie for first place, and that's okay. It just means there's not as much variability in the sample. (Read on to see what we mean by that.)

The category that occurs most often is called the **mode**. This meaning of the word *mode* is similar to its meaning when we use it with numerical variables. However, one big difference between categorical and numerical variables is that we call a categorical variable bimodal only if two categories are nearly tied for most frequent outcomes. (The two bars don't need to be exactly the same height, but they should be very close.) Similarly, a categorical variable's distribution is multimodal if more than two categories all have roughly the tallest bars. For a numerical variable, the heights of the mounds do not need to be the same height for the distribution to be multimodal.



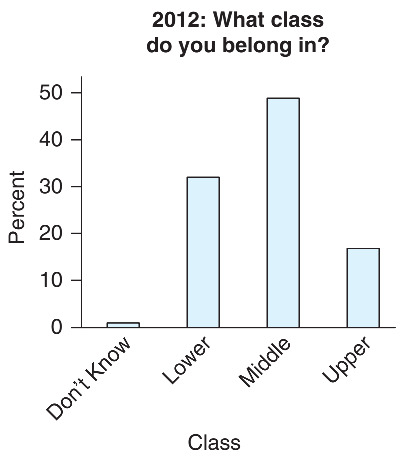
▲ **FIGURE 2.25** Percents of respondents who, in 2008, reported the economic class they felt they belonged to.

For an example of a mode, let's examine this study from the Pew Research Center. The researchers interviewed 2413 Americans in 2008 and asked them what economic class they felt they belonged in: the lower class, the middle class, or the upper class. Figure 2.25 shows the results.

We see that the mode is the middle class; over half of the people (in fact, 53%) identified themselves as members of the middle class. The remaining people were almost equally divided between the lower class (25%) and the upper class (21%).

EXAMPLE 7 A Matter of Class

In 2012, the Pew survey asked a new group of 2508 Americans which economic class they identified with. The bar chart in Figure 2.26 shows the distribution of the responses in 2012.



◀ **FIGURE 2.26** Percents of respondents who, in 2012, reported the economic class they felt they belonged to.

QUESTION What is the typical response? How would you compare the responses in 2012 with the responses in 2008?

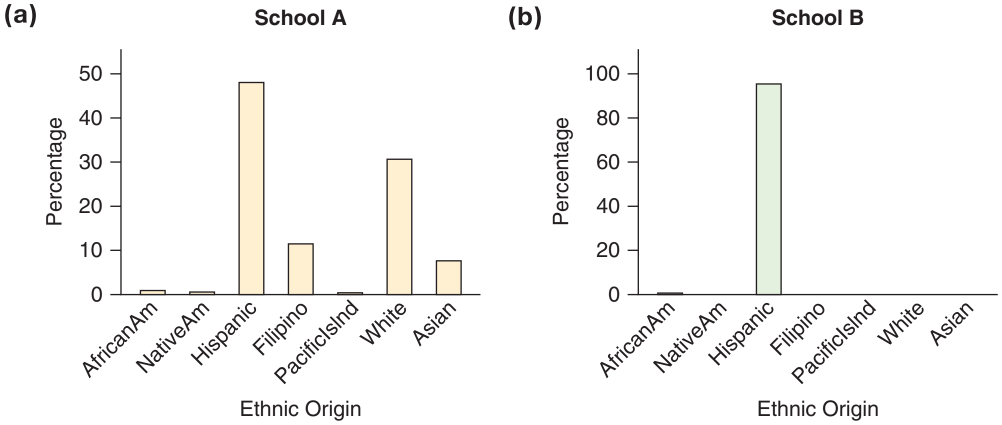
SOLUTION The mode, the typical response, is still the middle class. However, in 2012, fewer than 50% identified themselves as members of the middle class. This is a lower percentage than in 2008. It also appears that the percentage who identified themselves as members of the lower class increased from 2008, while the percentage of those identifying themselves as members of the upper class decreased.

TRY THIS! Exercise 2.37

Variability

When thinking about the variability of a categorical distribution, it is sometimes useful to think of the word *diversity*. If the distribution has a lot of diversity (many observations spread across many different categories), then its variability is high. Bar charts of distributions with high variability sometimes have several different modes, or several categories that are close contenders for being the mode. On the other hand, if all of the observations are in one single category, then diversity is low. If you are examining a bar chart and there is a single category that is clearly the only mode—a category with far more observations than the other categories—then variability is low.

For example, Figure 2.27 shows bar charts of the ethnic composition of two schools in the Los Angeles City School System. Which school has the greater variability in ethnicity?



◀ **FIGURE 2.27** Percents of students at two Los Angeles schools who are identified with several ethnic groups. Which school has more ethnic variability?

School A (Figure 2.27a) has much more diversity, because it has observations in more than four categories, whereas School B has observations in only two. At School A, the mode is clearly Hispanic, but the second-place group, Whites, is not too far behind.

School B (Figure 2.27b) consists almost entirely of a single ethnic group, with very small numbers in the other groups. The fact that there is one very clear mode means that School B has lower variability than School A.

Comparing variability graphically for categorical variables is not easy to do. But sometimes, as in the case of Figure 2.27, there are clear-cut instances where you can generally make some sort of useful comparison.

EXAMPLE 8 The Shrinking Middle Class?

Compare the distribution of responses to the Pew survey in 2008 (Figure 2.25) with that in 2012 (Figure 2.26).

QUESTION In which year is more variability apparent? Explain.

SOLUTION In 2008 the mode was very clearly the middle class. In 2012 this is still clearly the mode; however, the bar labeled Middle is lower than it was in 2008 because a lower percentage of people identified themselves as members of the middle class. At the same time, the percentage of people who identified themselves as members of the lower class grew, so the two bars labeled Lower and Middle are closer in height. This indicates that there is more variability in the 2012 distribution than in the 2008 distribution.

TRY THIS! Exercise 2.39

Describing Distributions of Categorical Variables

When describing a distribution of categorical data, you should mention the mode (or modes) and say something about the variability. Example 9 illustrates what we mean.

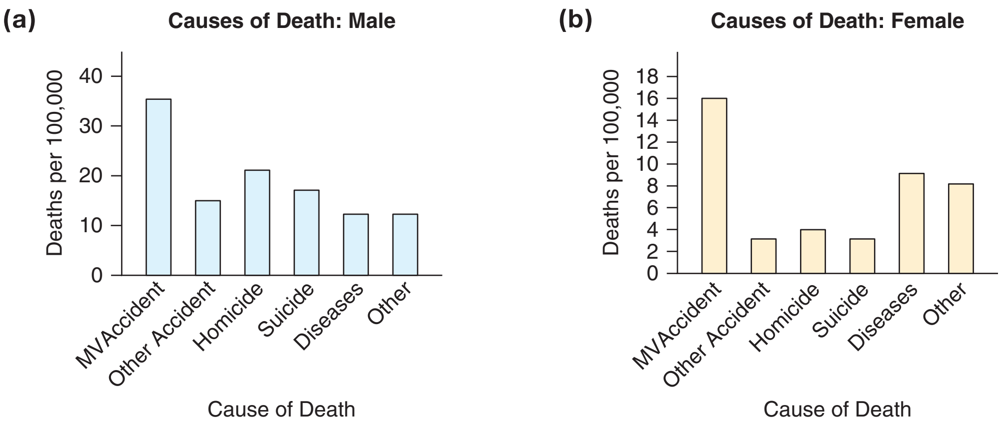
KEY POINT When summarizing graphs of categorical data, report the mode or modes and describe the variability (diversity).

EXAMPLE 9 Causes of Death

According to some experts, about 51.5% of babies currently born in the United States are male. But among people between 100 and 104 years old, there are four times as many women as men (U.S. Census Bureau, 2000). How does this happen? One possibility is that the percent of boys born has changed over time. Another possibility is presented in the two bar charts in Figure 2.28. These bar charts show the numbers of deaths per 100,000 people in one year, for people aged 15–24 years, for both males and females.

QUESTION Compare the distributions depicted in Figure 2.28. Note that the categories are put into the same order on both graphs.

SOLUTION First, note that the histograms are not on the same scale, as you can see by comparing the values on the y-axes. This presentation is typical of most software



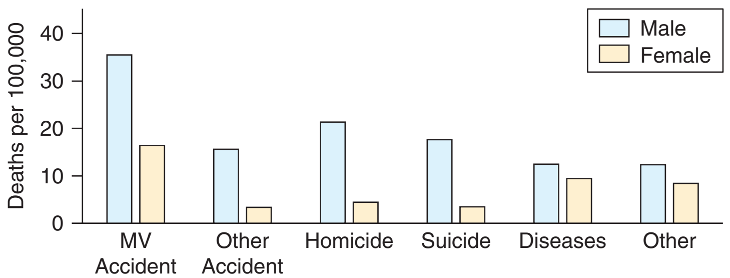
◀ **FIGURE 2.28** The number of deaths per 100,000 males **(a)** and females **(b)** for people 15–24 years old in a one-year period.

and is sometimes desirable, because otherwise, one of the bar charts might have such small bars that we couldn't easily discern differences.

Although motor vehicle accidents are the mode for both groups, males show a consistently high death rate for all other causes of death, whereas females have relatively low death rates in the categories for other accident, homicide, and suicide. In other words, the cause of death for females is less variable than that for males. It is also worth noting that the death rates are higher for males in every category. For example, roughly 16 out of every 100,000 females died in a motor vehicle accident in one year, while roughly 35 out of every 100,000 males died in car accidents in the same year.

TRY THIS! Exercise 2.43

We can also make graphics that help us compare two distributions of categorical variables. When comparing two groups across a categorical variable, it is often useful to put the bars side-by-side, as in Figure 2.29. This graph makes it easier to compare rates of death for each cause. The much higher death rate for males is made clear.



◀ **FIGURE 2.29** Death rates of males and females, graphed side by side.

SECTION 2.5

Interpreting Graphs

The first step in every investigation of data is to make an appropriate graph. Many analyses of data begin with visualizing the distribution of a variable, and this requires that you know whether the variable is numerical or categorical. When interpreting these graphics, you should pay attention to the center, spread, and shape.

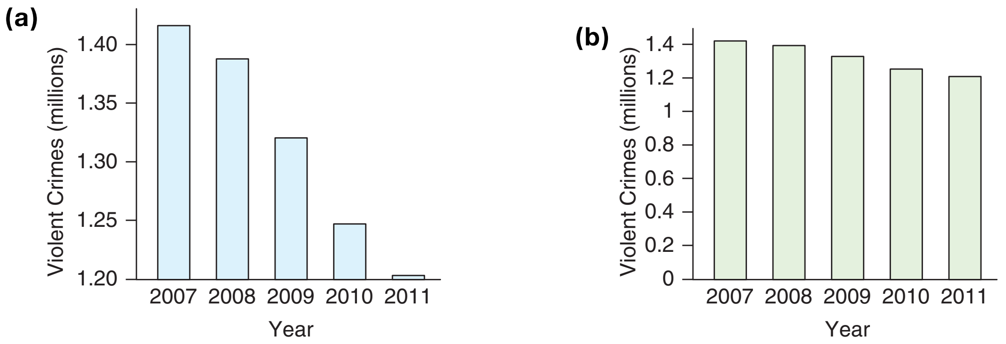
Often, you will come across graphics produced by other people, and these can take extra care to interpret. In this section, we'll warn you about some potential pitfalls when interpreting graphs and show you some unusual visualizations of data.

Misleading Graphs

A well-designed statistical graphic can help us discover patterns and trends and can communicate these patterns clearly to others. However, our eyes can play tricks on us, and manipulative people can take advantage of this to use graphs to give false impressions.

The most common trick—one that is particularly effective with bar charts—is to change the scale of the vertical axis so that it does not start at the origin (0). Figure 2.30a shows the number of violent crimes per year in the United States as reported by the FBI (http://www.fbi.gov). The graphic seems to indicate a dramatic drop in crime since 2007.

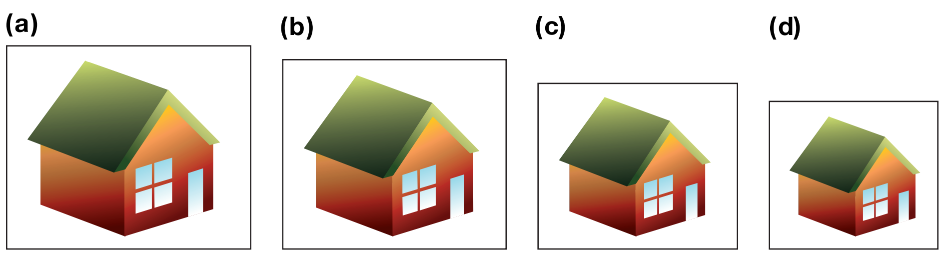
However, note that the vertical axis starts at about 1,200,000 (that is, 1.20 million) crimes. Because the origin begins at 1.20 million and not at 0, the bars are all shorter than they would be if the origin were at 0. The drop in 2010 seems particularly dramatic, in part because the height of the bar for 2010 is less than half the height of the bar for 2009. What does this chart look like if we make the bars the correct height? Figure 2.30b shows the same data, but to the correct scale. It's still clear that there has been a decline, but the decline doesn't look nearly so dramatic now, does it? Why not?



▲ **FIGURE 2.30** **(a)** This bar chart apparently shows a dramatic decline in the number of violent crimes since 2007. The origin for the vertical axis begins at about 1.20 million, not at 0. **(b)** This bar chart reports the same data as part (a), but here the vertical axis begins at the origin (0).

The reason is that when the origin is correctly set to 0, as in Figure 2.30b, it is clear that the percentage decline has not been so great. For instance, the number of crimes is clearly lower in 2010 than in 2009, but it is not half as low, as Figure 2.30a might suggest.

Most of the misleading graphics you will run across exploit a similar theme: Our eye tends to compare the relative sizes of objects. Many newspapers and magazines like to use pictures to represent data. For example, the plot in Figure 2.31 attempts to illustrate the number of homes sold for some past years, with the size of each house



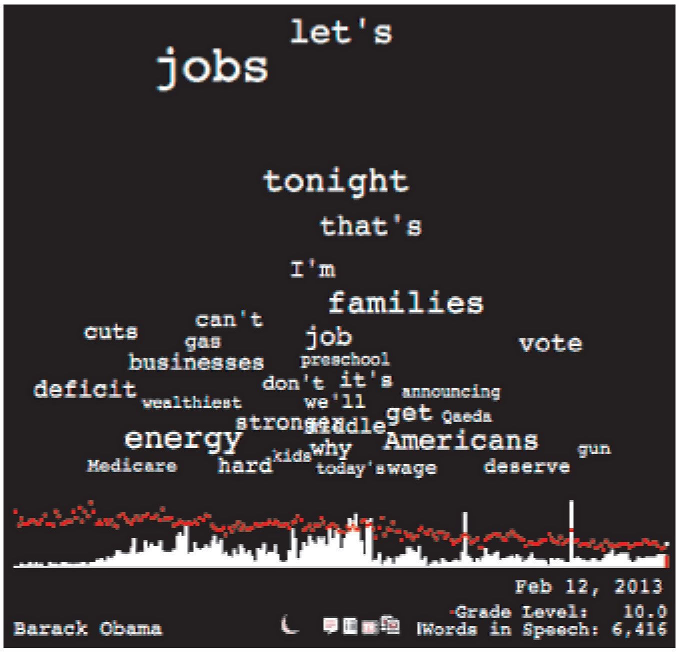
▶ **FIGURE 2.31** Deceptive graphs: Image **(a)** represents 7.1 million homes sold in 2005, image **(b)** represents 6.5 million homes sold in 2006, image **(c)** represents 5.8 million homes sold in 2007, and image **(d)** represents 4.9 million homes sold in 2008. (Source: *L.A. Times*, April 30, 2008) representing the number of homes sold that year. Such graphics can be very misleading, because the pictures are often not to scale with the actual numbers.

In Figure 2.31 the *heights* of the homes are indeed proportional to the sales numbers, but our eye reacts to the *areas* instead. The smallest house is 69% as tall as the largest house (because 4.9 is 69% of 7.1), but the area of the smallest house is only about 48% of that of the largest house, so our tendency to react to area rather than to height exaggerates the difference.

The Future of Statistical Graphics

The Internet allows for a great variety of graphical displays of data that take us beyond simple visualizations of distributions. Many statisticians, computer scientists, and designers are experimenting with new ways to visualize data. Most exciting is the rise of interactive displays. The *State of the Union Visualization*, for example (http://stateoftheunion.onetwothree.net), makes it possible to compare the content of State of the Union speeches. Every U.S. president delivers a State of the Union address to Congress near the beginning of each year. This interactive graphic enables users to compare words from different speeches and "drill down" to learn details about particular words or speeches.

For example, Figure 2.32 is based on the State of the Union address that President Barack Obama delivered on February 12, 2013. The largest words are the words that appear most frequently in the speech. The words that appear to the left are words that typically appear earlier in the speech, and the words located to the right typically appear later. Thus we see that near the beginning, President Obama talked about the economy, using words such as *cuts, jobs, deficit*, and *businesses*. Toward the end, there was talk about more general things, and we see *vote*, *deserve*, and *gun*. The word *families* is more or less in the middle, which suggests that it was used frequently throughout the speech. The vertical position of a word indicates how unusual it is compared to the content of other State of the Union addresses. Words that appear near the top, such as *let's* and *jobs*, are words that are particular to this speech, compared to all other State of the Union addresses. Words near the bottom, such as *today's, wage,* and *deserve*, are words that occur relatively frequently throughout all State of the Union addresses.



◀ **FIGURE 2.32** President Obama's 2013 State of the Union speech, visualized as a "word cloud." This array shows us the most commonly used words, approximately where these words occurred in the speech, how often they occurred, and how unusual they are compared to the content of other State of the Union speeches. (Courtesy of Brad Borevitz, http://onetwothree.net. Used with permission.)

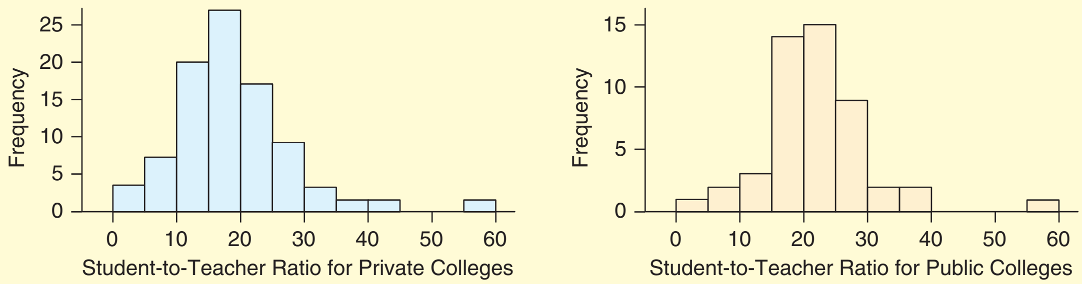
Figure 2.32 is packed with information. On the bottom, in red, you see dots that represent the grade level required to read and understand the speech. Each dot represents a different speech, and the speeches are sorted from first (George Washington) to most recent (Barack Obama). Near the bottom right we see that the grade level of the speech is 10.0, or about tenth grade, and that this is considerably lower than it was 100 years ago but consistent with more recent speeches. The white "bars" represent the numbers of words in the State of the Union addresses. This speech contains 6416 words, which is fairly typical of the last few years. The tallest spike represents Jimmy Carter's speech in 1981, which had 33,613 words and a grade level of 15.3!

CASE STUDY REVISITED

Student-to-Teacher Ratio at Colleges

How do public colleges compare to private colleges in student-to-teacher ratios? The data were presented at the beginning of the chapter. This list of raw data makes it hard to see patterns, so it is very difficult to compare groups. But because the student-to-teacher ratio is a numerical variable, we can display these distributions as two histograms to enable us to make comparisons between public and private schools.

In Figure 2.33, the student-to-teacher ratio for the private schools is shown in the left histogram, and that for public schools is shown in the right histogram. The differences in the distributions tell us about differences in these types of colleges. The typical student-to-teacher ratio for private schools is around 18 students per teacher, while for public schools it's a little over 20. Although both schools have quite a bit of spread in these ratios, there may be a little less spread—a little less variation—in the public schools. The private schools' distribution is right-skewed. The public schools' distribution is more symmetric. Both have outliers.



▲ **FIGURE 2.33** Histograms of the student-to-teacher ratio at samples of private colleges and public colleges.

By the way, many of the schools with the smallest student-to-teacher ratios deal with specialties such as health or design. In case you are wondering which public college has the smallest student-to-teacher ratio in Figure 2.33, it is the University of Medicine and Dentistry of New Jersey. At medical facilities like this one, the physicians are often listed as faculty. The private school with the 60-to-1 ratio is Mountain State University in Beckley, West Virginia. Mountain State University closed permanently in 2013 after losing its accreditation.

EXPLORING STATISTICS
CLASS ACTIVITY

Personal Distance

How much personal distance do people require when they're using an automatic teller machine?

GOALS	MATERIALS
In this activity, you will learn to make graphs of sample distributions in order to answer questions about data in comparing two groups of students.	Meter stick (or tape measure).

ACTIVITY

Work in groups of three students. Each group must have a meter stick. The first person stands (preferably in front of a wall) and imagines that she or he is at an ATM getting cash. The second student stands behind the first. The first student tells the second student how far back he or she must stand for the first student to be just barely comfortable, saying, for example, "Move back a little, now move forward just a tiny bit," and so on. When that distance is set, the third student measures the distance between the heel of the first person's right shoe and the toe of the second person's right shoe. That will be called the "personal distance."

For each student in your group, record the gender and personal distance. Your instructor will help you pool your data with the rest of the class.

Note: Be respectful of other people's personal space. Do not make physical contact with other students during this activity.

BEFORE THE ACTIVITY

1. Do you think men and women will have different personal distances? Will the larger distances be specified by the men or the women?
2. Which group will have distances that are more spread out?
3. What will be the shape of the distributions?

AFTER THE ACTIVITY

Do men and women have different personal distances? Create appropriate graphics to compare personal distances of men and women to answer this question. Then describe these differences.

CHAPTER REVIEW

KEY TERMS

LEARNING OBJECTIVES

After reading this chapter and doing the assigned homework problems, you should

- Understand that a distribution of a sample of data displays a variable's values and the frequencies (or relative frequencies) of those values.
- Know how to make graphs of distributions of numerical and categorical variables and how to interpret the graphs in context.
- Be able to compare centers and spreads of distributions of samples informally.

SUMMARY

The first step in any statistical investigation is to make plots of the distributions of the data in your data set. You should identify whether the variables are numerical or categorical so that you can choose an appropriate graphical representation.

If the variable is numerical, you can make a dotplot, histogram, or stemplot. Pay attention to the shape (Is it skewed or symmetric? Is it unimodal or multimodal?), to the center (What is a typical outcome?), and to the spread (How much variability is present?). You should also look for unusual features, such as outliers.

Be aware that many of these terms are deliberately vague. You might think a particular observation is an outlier, but another person might not agree. That's okay, because the purpose isn't to determine whether such points are "outliers" but to indicate whether further investigation is needed. An outlier might, for example, be caused by a typing error someone made when entering the data.

If you see a bimodal or multimodal distribution, ask yourself whether the data might contain two or more groups combined into the single graph.

If the variable is categorical, you can make a bar chart, a Pareto chart (a bar chart with categories ordered from most frequent to least frequent), or a pie chart. Pay attention to the mode (or modes) and to the variability.

SOURCES

Guerin, C., et al. 2013. Prone positioning in severe acute respiratory distress syndrome. *New England Journal of Medicine*, vol. 368 (June 6): 2159–2168).

Internet Research Legal Group. 2013. http://www.ilrg.com/rankings/

Leiserowitz, A., E. Maibach, C. Roser-Renouf, and D.J. Hmielowski. 2012. *Climate change in the American mind: Americans' global warming beliefs and attitudes in March 2012.* Yale University and George Mason University. New Haven, CT: Yale Project on Climate Change Communication. (http://environment.yale.edu/climate/files/Climate-Beliefs-March-2012.pdf)

Minitab 15 Statistical Software (2007). [Computer software]. State College, PA: Minitab, Inc. (www.minitab.com)

National Center for Educational Statistics: http://www.nces.ed.gov/ipeds/

National Safety Council. 2002. *Injury Facts*, 2002 edition. Itasca, Illinois.

Playfair, W. 1786. Commercial and political atlas: Representing, by copper-plate charts, the progress of the commerce, revenues, expenditure, and debts of England, during the whole of the eighteenth century. London: Corry. Reprinted in H. Wainer and I. Spence (eds.), *The commercial and political atlas and statistical breviary*, New York: Cambridge University Press, 2005. ISBN 0-521-85554-3. http://www.math.yorku.ca/SCS/Gallery/milestone/refs.html#Playfair:1786)

Tufte, E. 1983. *The visual display of quantitative information*, Graphics Press: Cheshire, Connecticut. 1st ed., p. 141. Photo by Brian Joiner.

U.S. Census Bureau. 2000. http://ceic.mt.gov/C2000/SF12000/Pyramid/pptab00.htm

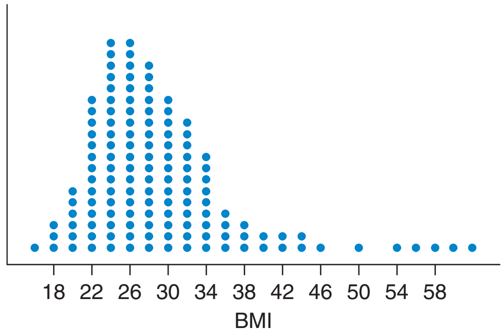
SECTION EXERCISES

SECTIONS 2.1 AND 2.2

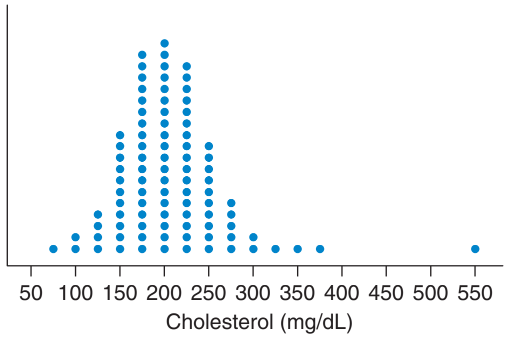
2.1 Body Mass Index The dotplot shows body mass index (BMI) for 134 people according to the National Health and Nutrition Examination Survey (NHANES) in 2010, as reported in *USA Today*.

a. A BMI of more than 40 is considered morbidly obese. Report the number of morbidly obese shown in the dotplot.

b. Report the percentage of people who are morbidly obese. Compare this with an estimate from 2005 that 3% of people in the United States at that time were morbidly obese.



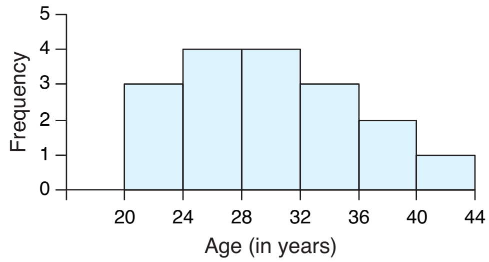
2.2 Cholesterol Levels The dotplot shows the cholesterol level of 93 adults from the 2010 NHANES data.



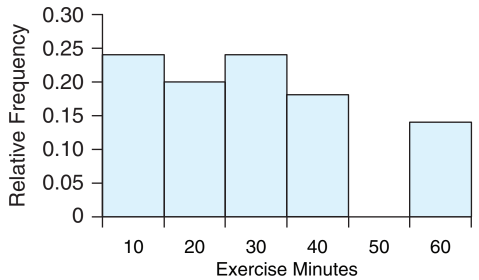
a. A total cholesterol level of 240 mg/dL (milligrams per deciliter) or more is considered unhealthy. Report the number of people in this group with unhealthy cholesterol levels.

b. Knowing there are a total of 93 people in this sample, report the percentage of people with unhealthy total cholesterol levels. How does this compare with an estimate from 2010 that 18% of people in the United States had unhealthy cholesterol levels?

2.3 When Billionaires Made Their First Million The histogram shows frequencies for the ages of 17 billionaires when they made their first million listed at inc.com. Convert this histogram to one showing relative frequencies by relabeling the vertical axis, with the appropriate relative frequencies. You may just report the new labels for the vertical axis because that is the only thing that changes.



* **2.4 Exercise Minutes** The relative frequency histogram shows the average number of minutes spent on exercising (Exercise Minutes) per day reported as experienced "last month" for 50 working women.



*a. About how many women had 30 or fewer minutes of exercise per day?

b. The graph is bimodal. What are the two modes?

TRY **2.5 Paper Shredders (Example 1)** The histogram shows the distribution of the number of paper shredders in the offices of 120 lawyers.

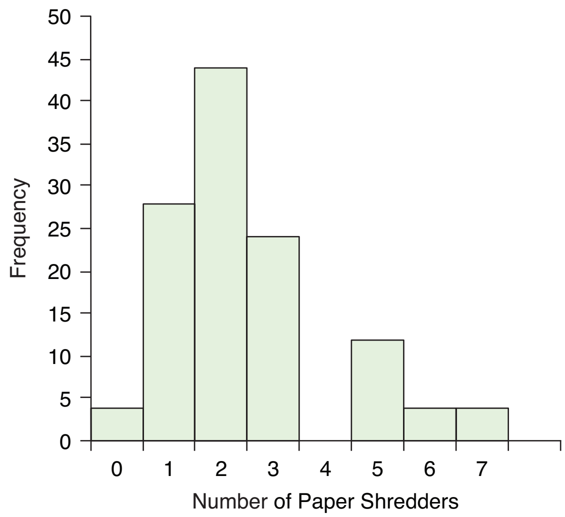
a. According to the histogram, about how many offices do not have a paper shredder?

b. How many paper shredders are in the offices that have the most paper shredders?

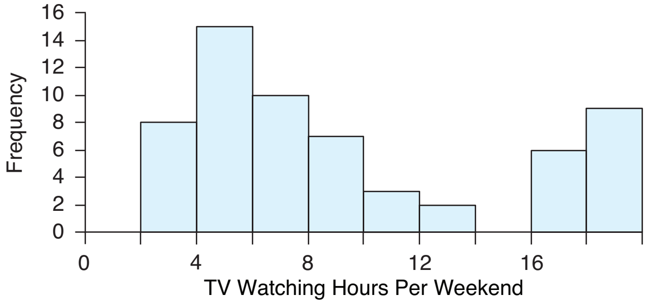
c. How many offices have two paper shredders?

d. How many offices have six or more paper shredders?

e. What proportion of offices have six or more paper shredders?



2.6 TV Time The histogram shows the distribution of self-reported numbers of hours of watching TV per weekend for 60 eighth-grade students. This graph uses a right-hand rule: Someone who watched TV for exactly 4 hours would be in the second bin, the bin to the right of 4.



a. According to the histogram, there are two possible values for the minimum number of hours of watching TV. What are they?

b. How many students watched TV for 18 or 19 hours (more than 17 hours)?

c. How many students watched TV for 7 or fewer hours?

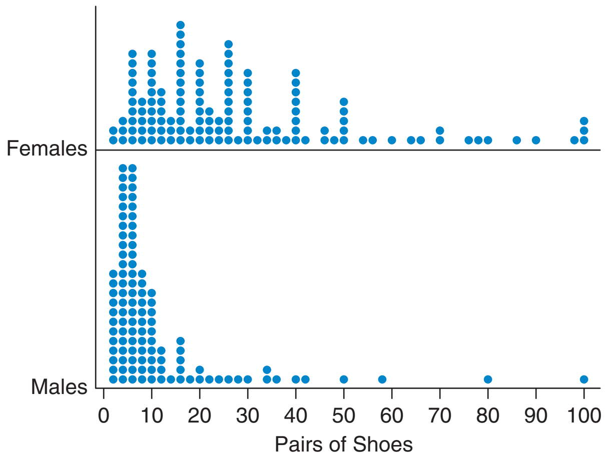
d. What proportion of students watched TV for 6 or fewer hours?

2.7 Shoes The graph is a dotplot of the number of pairs of shoes owned by men and women who took a survey on StatCrunch. (Source: StatCrunch Responses to Shoe Survey. Owner: scsurvey)

a. Shape: What is the shape of each dotplot?

b. Center: Is it the males or the females who typically have more pairs of shoes?

* c. Spread: For which group is the data set more spread out?

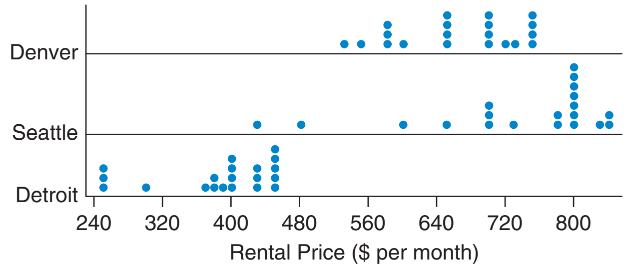


2.8 Condo Rental The dotplot shows rental prices per month for condos listed in three cities. The prices were obtained from Zillow.com.

a. Center: Which city typically has the lowest rents?

b. Spread: Which city has the greatest spread?

c. Shape: What is the direction of skew for the distribution of rental prices in Seattle?



TRY **2.9 Practice Hours (Example 2)** A coach asks 50 football players how many hours they practiced a day before the match. Predict the shape of the distribution and explain.

2.10 Penalty Predict the shape of the distribution of the number of times a group of 300 employees has been fined for coming in late over one week.

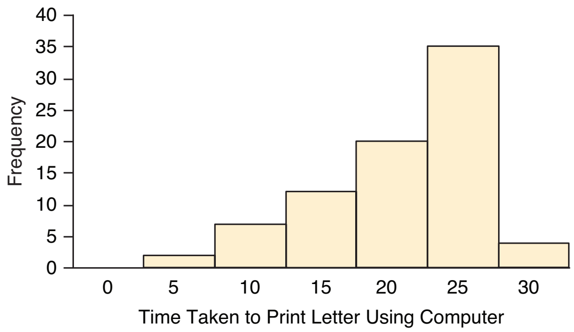
TRY **2.11 Armspans (Example 3)** According to the ancient Roman architect Vitruvius, a person's armspan (the distance from fingertip to fingertip with the arms stretched wide) is approximately equal to his or her height. For example, people 5 feet tall tend to have an armspan of 5 feet. Explain, then, why the distribution of armspans for a class containing roughly equal numbers of men and women might be bimodal.

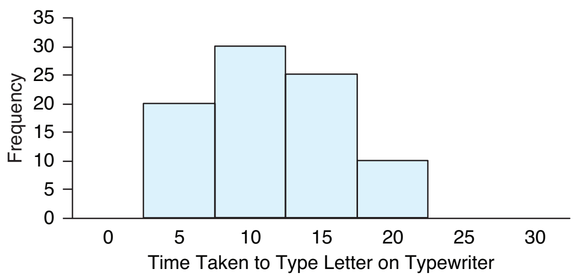
* **2.12 Tuition** The distribution of in-state annual tuition for all colleges and universities in the United States is bimodal. What is one possible reason for this bimodality?

TRY **2.13 When Billionaires Made Their First Million (Example 4)** From the histogram in Exercise 2.3, approximately what is a typical age of a billionaire in this sample?

2.14 Exercise Minutes From the histogram shown in Exercise 2.4, what is the typical number of exercise minutes for these women?

TRY **2.15 Printing Times (Example 5)** Use the histograms to compare the time spent in getting a final copy of a letter typed on a typewriter with the time spent on getting a copy of the same letter typed on a computer and then printed. Which method typically takes longer? Which group has the more variable printing time?



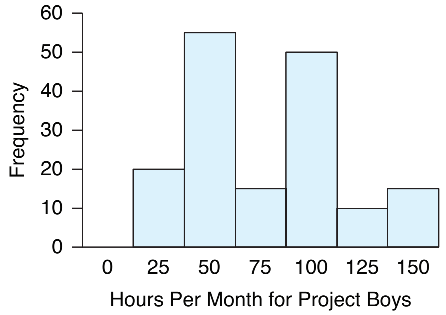


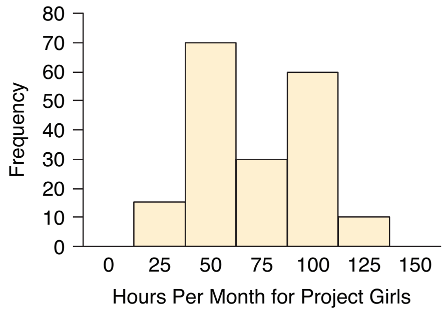
2.16 Spending on Social Service The histograms show the distribution of the estimated numbers of hours per month spent on social services for project boys (left) and project girls (right).

a. Compare and describe the shape of the distributions.

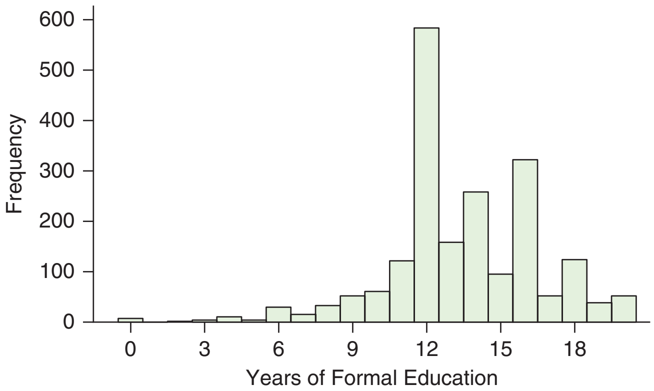
b. Which group tends to spend more time?

c. Which group has more variation in the hours spent?





TRY **2.17 Education (Example 6)** In 2012, the General Social Survey (GSS), a national survey conducted nearly every year, reported the number of years of formal education for 2018 people. The histogram shows the distribution of data.

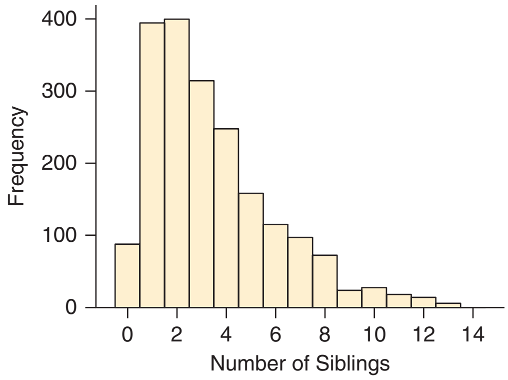


a. Describe and interpret the distribution of years of formal education. Mention any unusual features.

b. Assuming that those with 16 years of education completed a bachelor's degree, estimate how many of the people in this sample got a bachelor's degree or higher.

c. The sample includes 2018 people. What percentage of people in this sample have a bachelor's degree or higher? How does this compare with Wikipedia's estimate that 27% have a bachelor's degree?

2.18 Siblings The histogram shows the distribution of the numbers of siblings (brothers and sisters) for 2000 adults surveyed in the 2012 General Social Survey.



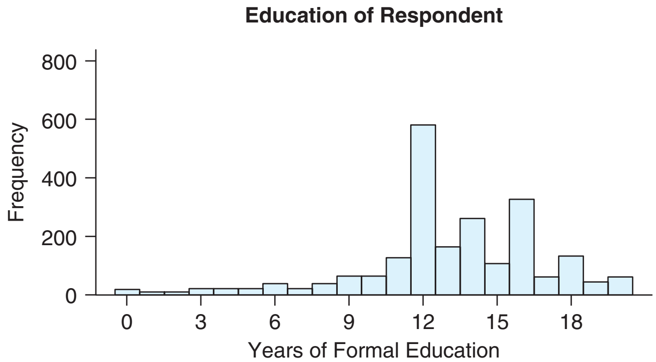
a. Describe the shape of the distribution.

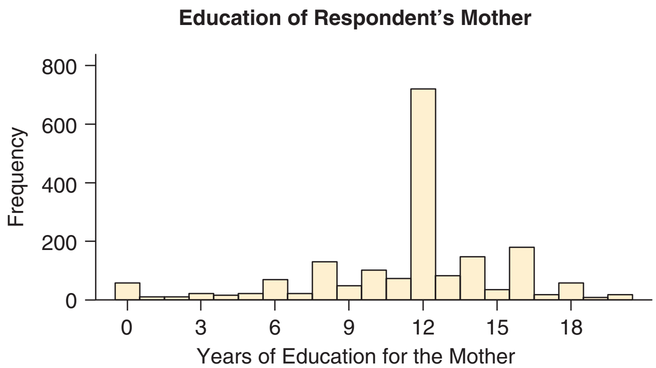
b. What is the typical number of siblings, approximately?

c. About how many people in this survey have no siblings?

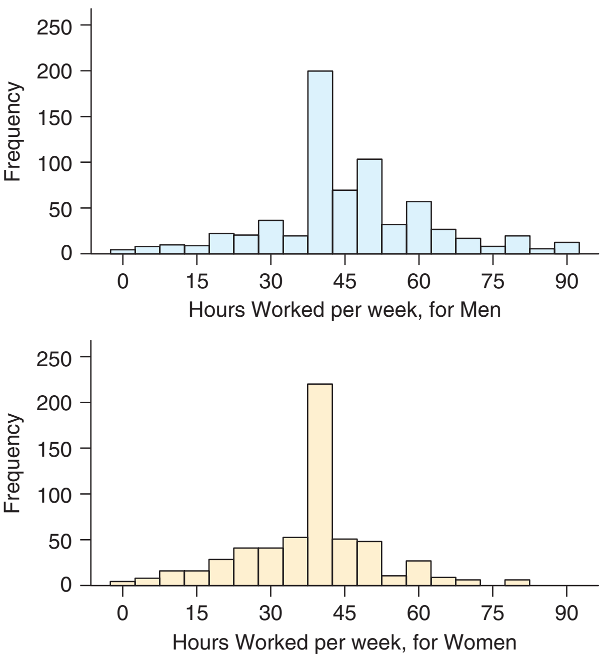
d. What percentage of the 2000 adults surveyed have no siblings?

* **2.19 Years of Education** The GSS asked people how many years of education they had and how many years their mothers had. If people who responded to the survey (respondents) completed high school but had no further education, they reported 12 years of education. If they stopped after a bachelor's degree, they reported 16 years. There were 2018 people who answered the question about their own education and 1780 who answered the question about the education of their mothers. Compare the distribution of years of education of the respondents with the distribution of years of education of their mothers.



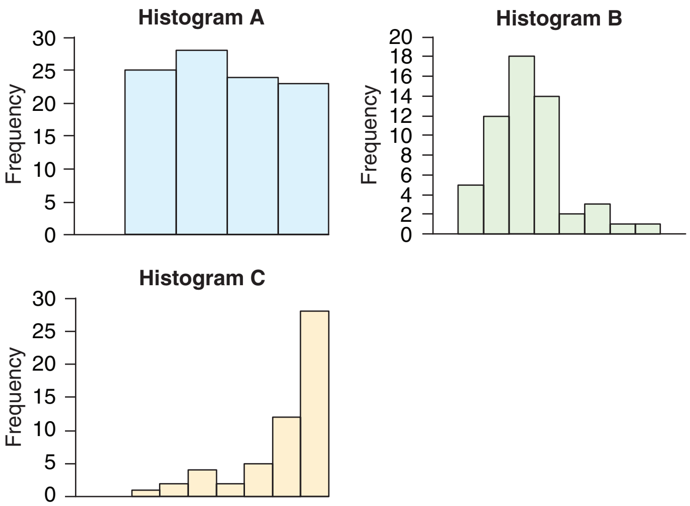


* **2.20 Hours Worked** In the 2012 General Social Survey, 636 male paid employees and 537 female paid employees were asked how many hours they worked in the last week. (Those who said, "I don't know" or "I don't work" were not included in the data set.) Compare the distributions of hours of work for the men and the women.



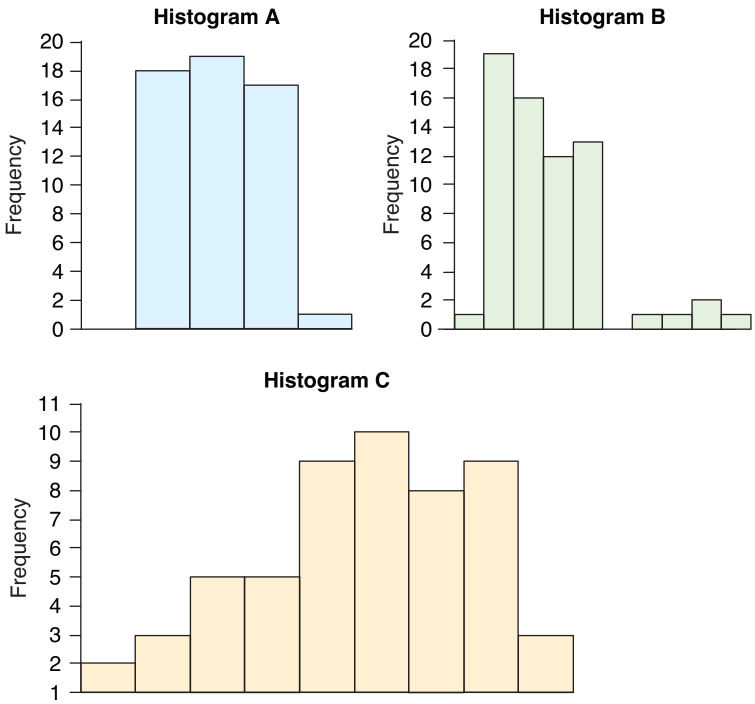
2.21 Matching Histograms Match each of the following histograms to the correct situation.

1. Ages of graduates pursuing higher education in business management.
2. Semester (1 = First Semester, 2 = Second Semester, 3 = Third Semester, and 4 = Fourth Semester) of graduates studying business management at a two-year college.
3. Number of times these same students reported studying in the last week.



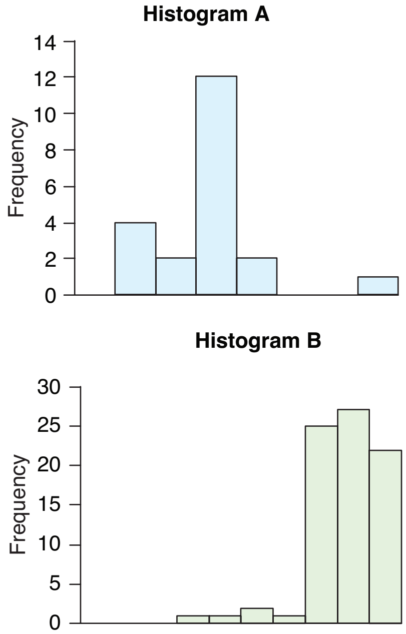
2.22 Matching Histograms Match each of the following histograms to the current situation.

1. An easy question attempted first by students in an exam.
2. The number of hours of practice by players of a national team.
3. The age of a large, typical group of tenth-grade students.

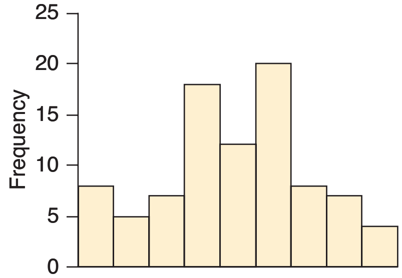


2.23 Matching Match each description with the correct histogram (A through C).

1. Shoe size of employees in an office that contains about equal numbers of men and women.
2. Number of working hours the previous week in the same office.
3. Time taken to reach office by a car.



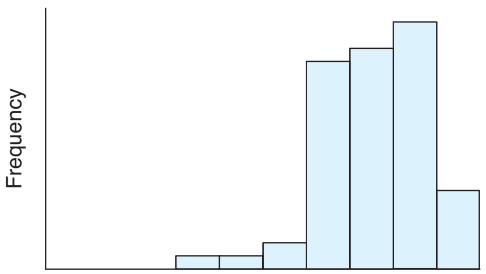
Histogram C



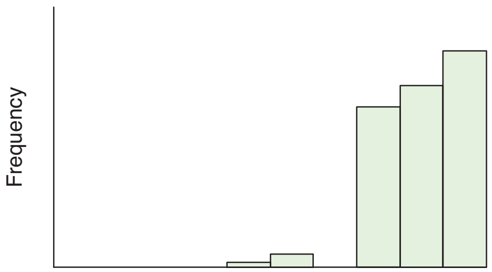
2.24 Matching Match each description with the correct histogram (A through C).

1. Running speed of 100 athletes who are participating in Olympics.
2. Hemoglobin count of around 1000 students, about half of whom are boys and half of whom are girls.
3. Ages of all 50 students in a college class that is made up of part-time workers and conducted in the evening.

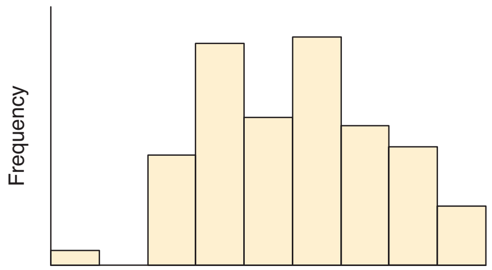
Histogram A



Histogram B



Histogram C



2.25 Eating Out and Jobs College student Jacqueline Loya asked students who had full-time jobs and students who had part-time jobs how many times they went out to eat in the last month.

Briefly compare the distributions of the responses in the two groups. Include appropriate graphics. *See page 100 for guidance.*

Full-time: 5, 3, 4, 4, 4, 2, 1, 5, 6, 5, 6, 3, 3, 2, 4, 5, 2, 3, 7, 5, 5, 1, 4, 6, 7
Part-time: 1, 1, 5, 1, 4, 2, 2, 3, 3, 2, 3, 2, 4, 2, 1, 2, 3, 2, 1, 3, 3, 2, 4, 2, 1

2.26 Comparing Weights of Baseball and Soccer Players College students Edward Lara and Anthony Dugas recorded the self-reported weights (in pounds) of some community college male baseball players and male soccer players. Write a brief comparison of the distributions of weights for the two groups. Include appropriate graphics. For one approach, look at the guidance given for Exercise 2.25.

Baseball Players' Weights

190	205	230	198	195
200	185	195	180	182
187	177	169	182	193
181	207	186	193	190
192	225	210	200	186

Soccer Players' Weights

165	189	167	173	184
190	170	190	158	174
185	182	185	150	190
187	172	156	172	156
183	180	168	180	163

2.27 Textbook Prices The table shows prices of 50 college textbooks in a community college bookstore, rounded to the nearest dollar. Make an appropriate graph of the distribution of the data, and describe the distribution.

76	19	83	45	88	70	62	84	85	87
86	37	88	45	75	83	126	56	30	33
26	88	30	30	25	89	32	48	66	47
115	36	30	60	36	140	47	82	138	50
126	66	45	107	112	12	97	96	78	60

2.28 SAT scores The table shows a random sample of 50 quantitative SAT scores of first-year students admitted at a university. Make an appropriate graph of the distribution of the data, and describe the distribution.

649	557	734	653	652	538	674	705	729	737
672	583	729	677	618	662	692	692	672	624
669	529	609	526	665	724	557	647	719	593
624	611	490	556	630	602	573	575	665	620
629	593	665	635	700	665	677	653	796	601

2.29 Animal Longevity The table below Exercise 2.30 shows the average lifespan for some mammals in years, according to infoplease.com. Graph these average lifespans and describe the distribution. What is a typical lifespan? Identify the three outliers and report their lifespans. If you were to include humans in this graph, where would the data point be? Humans live an average of about 75 years.

2.30 Animal Gestation Periods The accompanying table also shows the gestation period (in days) for some animals. The gestation period is the length of pregnancy. Graph the gestation period and describe the distribution. If there are any outliers, identify the animal(s) and give their gestation periods. If you were to include humans in this graph, where would the data point be? The human gestation period is about 266 days.

Animal	Gestation (days)	Lifespan (years)
Baboon	187	12
Bear, grizzly	225	25
Beaver	105	5
Bison	285	15
Camel	406	12
Cat, domestic	63	12
Chimp	230	20
Cow	284	15
Deer	201	8
Dog, domestic	61	12
Elephant, African	660	35
Elephant, Asian	645	40
Elk	250	16
Fox, red	52	7
Giraffe	457	10
Goat	151	8
Gorilla	258	20

Animal	Gestation (days)	Lifespan (years)
Hippo	238	41
Horse	330	20
Leopard	98	12
Lion	100	15
Monkey, rhesus	166	15
Moose	240	12
Pig, domestic	112	10
Puma	90	12
Rhino, black	450	15
Sea Lion	350	12
Sheep	154	12
Squirrel, gray	44	10
Tiger	105	16
Wolf, maned	63	5
Zebra, Grant's	365	15

2.31 Tax Rate A StatCrunch survey asked people what they thought the maximum income tax rate should be in the United States. Make separate dotplots of the responses from Republicans and Democrats. If possible, put one above the other, using the same horizontal axis. Then compare the groups by commenting on the shape, center, and spread of each distribution. The data are at this text's website. (Source: StatCrunch Survey: Responses to Taxes in the U.S. Owner: scsurvey)

2.32 Pets A StatCrunch survey asked people whether they preferred cats or dogs and how many pets they had. (Those who preferred both cats and dogs and those who preferred another type of pet are not included in these data.) Make two dotplots showing the distribution of the numbers of pets: one for those preferring dogs and one for those preferring cats. If possible, put one dotplot above the other, using the same horizontal axis. Then compare the distributions. (Source: StatCrunch: Pet Ownership Survey Responses. Owner: chitt71)

Cat
15
5
1
1
5
3
1
2
3
6
1
1
1
4
0
1
3
0
2

Dog	
6	2
1	1
1	1
2	1
2	1
1	1
2	7
1	1
1	1
2	1
1	1
3	1
4	1
1	2
4	1
3	2
4	0
2	2
1	4
2	2
1	3
1	1
3	3
1	2
1	2
2	1
7	
4	

2.33 Law School Tuition Data are shown for the cost of one year of law school at 30 of the top law schools in the United States, in 2013. The numbers are in thousands of dollars. Make a histogram of the costs, and describe the distribution. If there are any outliers, identify the school(s). (Source: http://grad-schools.usnews.rankingsandreviews.com. Accessed via StatCrunch.)

Yale University	53.6
Harvard University	50.9
Stanford University	50.8
University of Chicago	50.7
Columbia University	55.5
University of Pennsylvania	53.1
New York University	51.1
Duke University	50.8
Cornell University	55.2
Georgetown University	48.8
Vanderbilt University	46.8
Washington University	47.5
University of Boston	44.2
University of Southern California	50.6
George Washington University	47.5
University of Notre Dame	46.0
Boston College	43.2
Washington and Lee University	43.5
Emory University	46.4
Fordham University	49.5
Brigham Young University	21.9
Tulane University	45.2
American University	45.1
Wake Forest University	39.9
College of William and Mary	37.8
Loyola Marymount University	44.2
Baylor University	46.4
University of Miami	42.9
Syracuse University	45.7
Northeastern University	43.0

2.34 Text Messages Recently, 115 users of StatCrunch were asked how many text messages they sent in one day. Make a histogram to display the distribution of the numbers of text messages sent, and describe the distribution. Some sample data are shown. The full data set is at this text's website. (Source: StatCrunch Survey: Responses to How Often Do You Text? Owner: Webster West)

Sent Texts	Sent Texts
1	50
1	6
0	5
5	300
5	30

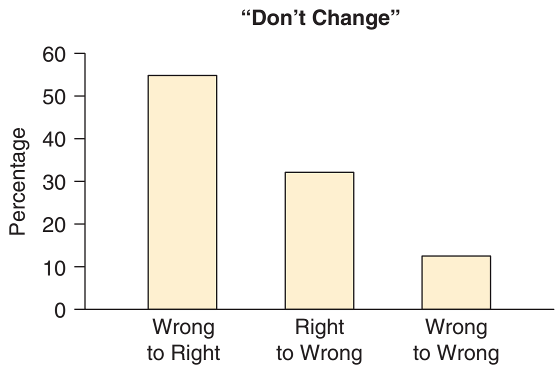
2.35 Fruits, Sugar Data are available at a website on the weight of total sugar (in grams) per 100 grams of fruit for 84 different fruits. Make a histogram to show the distribution of the quantity of total sugar, and describe the distribution. The first few entries are shown in the accompanying table. (Source: thepaleodiet.com)

Fruit	Total Sugar	Glucose
Almond Joy	44.9	0
Apples	13.3	2.3
Apricots	9.3	1.6
Avocado	0.9	0.5
Baby Ruth	42	0

2.36 Fruits, Glucose Data are available at a website on the weight of glucose in 84 different fruits. Make a histogram of the data, and describe that distribution. (Source: thepaleodiet.com)

SECTIONS 2.3 AND 2.4

TRY **2.37 Changing Multiple-Choice Answers When Told *Not* to Do So (Example 7)** One of the authors wanted to determine the effect of changing answers on multiple-choice tests. She studied the tests given by another professor, who had told his students before their exams that if they had doubts about an answer they had written, they would be better off *not changing* their initial answer. The author went through the exams to look for erasures, which indicate that the first choice was changed. In these tests, there is only one correct answer for each question. Do the data support the view that students should not change their initial choice of an answer?



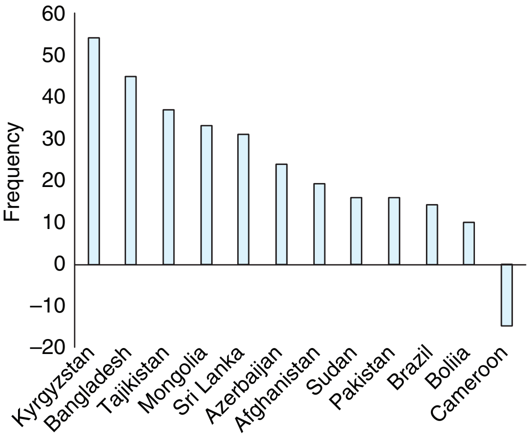
2.38 Change in Prices According to worldbank.org, there were 12 countries that recorded the largest price change in wheat from June to December 2010. These are shown in the graph.

a. Estimate how much price change happened in Bangladesh.

b. Estimate how much price change happened in Cameroon.

c. Does this graph support the theory that the country with the greatest price change in wheat during the period was Sri Lanka, as contended by some research results?

d. This is a bar chart with a special name because of the decreasing order of the bars. What is that name?



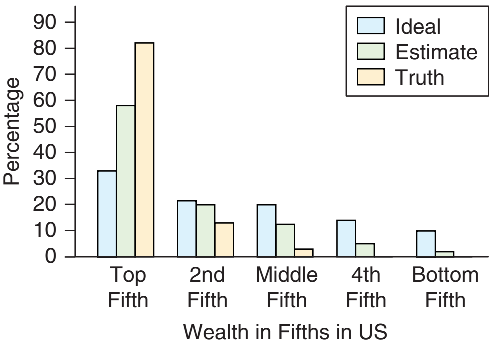
TRY **2.39 U.S. Distribution of Wealth (Example 8)** Michael Norton (of Harvard) and Dan Ariely (of Duke) did a survey reported in *Harvard Magazine* (Nov–Dec 2011). They asked respondents to estimate how much of the wealth is held by each quintile (fifth) of the people in the United States. The top quintile is the wealthiest 20% of the people in the United States, and the bottom quintile is the poorest 20%. They also asked respondents what the distribution should be ideally. The bar chart shows these responses, as well as the true values. For instance, the true value for the bottom fifth is 0.1% and the true value for the fourth fifth is 0.2%, so they are hard to see on the chart. This means that the poorest fifth of the people have only 0.1% of the country's wealth. (Source: http://harvard-magazine.com/2011/11/what-we-know-about-wealth)

a. In truth, the top fifth (in terms of wealth) actually held about what percentage of the wealth?

b. Which figures (Ideal, Estimate, or Truth) show the least variation?

c. Which figures (Ideal, Estimate, or Truth) show the most variation?

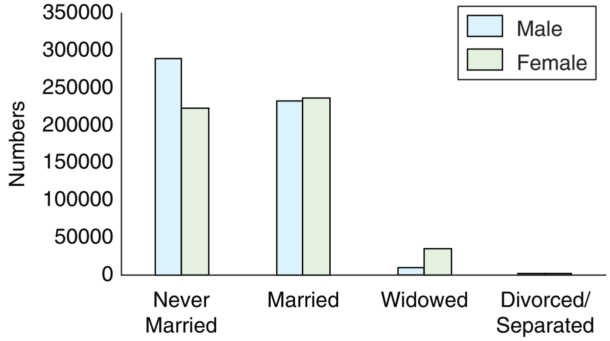
d. Do people making estimates tend to underestimate or to overestimate the proportion of wealth held by the top 20%?



2.40 Marital Status The graph shows the marital status of men and women in India according to Tables C2 and C14 of Census of India 2001.

a. Which gender tends to have more widowed members? Explain.

b. Which gender tends to have a low percentage of married people?

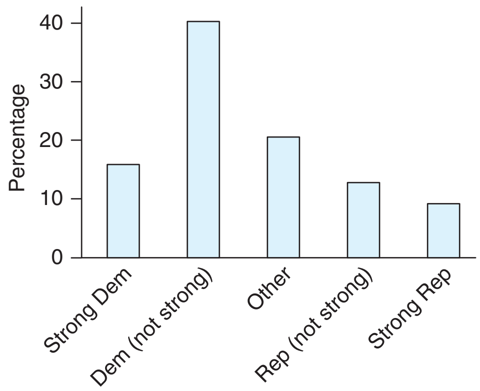


2.41 Political Party Affiliation: Men The 2012 General Social Survey (GSS) asked its respondents to report their political party affiliation. The graphs show the results for 879 men.

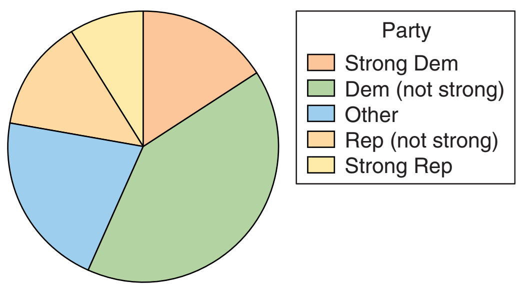
a. Which political affiliation has the most men?

b. What political affiliation has the second highest number of men? Is this easier to determine with the bar chart or with the pie chart? Explain.

Bar Chart of Political Party for Men



Pie Chart of Political Party for Men

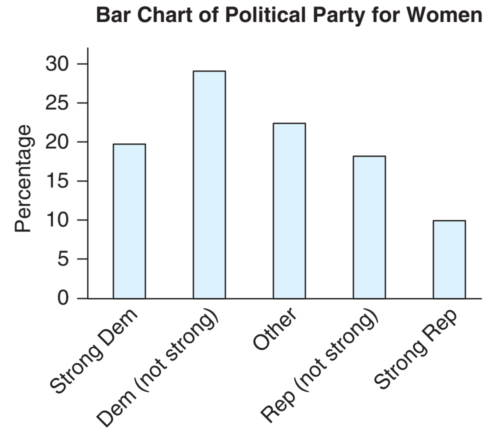


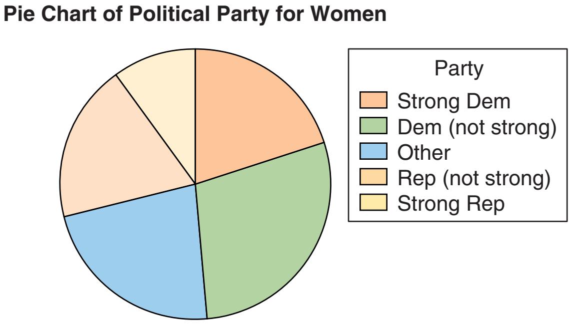
2.42 Political Party Affiliation: Women The 2012 General Social Survey (GSS) asked its respondents to report their political party affiliation. The graphs show the results for 1081 women.

a. Which political affiliation has the most women?

b. What political affiliation has the second largest number of women? Is this easier to determine with the bar chart or with the pie chart? Explain.

c. Some people believe that women tend to lean more than men toward liberal political positions (such as those advocated by the Democrats). Compare the graphs for Exercises 2.41 and 2.42. Do you see evidence of this? Explain.

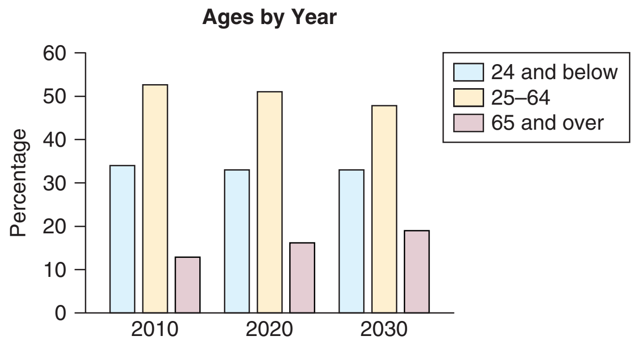




TRY **2.43 Age by Year (Example 9)** The bar chart shows the projected percentage of U.S. residents in different age categories by year, according to the *2009 World Almanac and Book of Facts*.

a. Comment on the predicted changes from 2010 through to 2030. Which age groups are predicted to become larger, which are predicted to become smaller, and which are predicted to stay roughly the same?

b. Comment on the effect this might have on Social Security, a government program that collects money from those currently working and gives it to retired people.

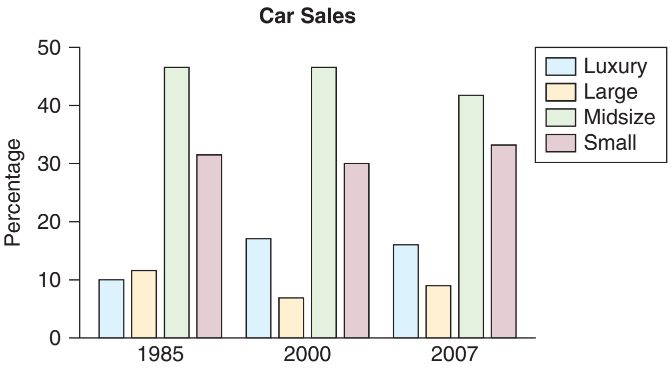


2.44 Retail Car Sales With gas prices rising, as they did between 1985 and 2007, you might expect people to move toward buying smaller cars. Compare the types of cars sold in 1985, 2000, and 2007 as shown in the figure. (Source: *2009 World Almanac and Book of Facts*)

a. Which type of car sold the most in all three years?

b. What is the trend for small cars? Has a higher or a lower percentage of small cars sold in more recent years?

c. What is the trend for large cars?



2.45 Nobel Laureates The table gives information on Nobel laureates affiliated with the University of Cambridge, United Kingdom. Sketch an appropriate graph of the distribution, and comment on its important features.

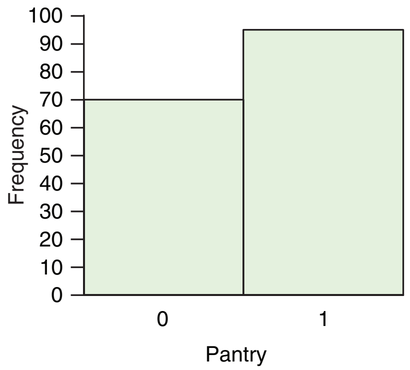
Category	Percentage
Literature	3%
Science	84%
Economics	11%
Peace	2%

2.46 Adoptions The table gives information on the top five countries from which U.S. residents adopted children in 2007. Sketch an appropriate graph of the distribution, and comment on its important features. (Source: *2009 World Almanac and Book of Facts*)

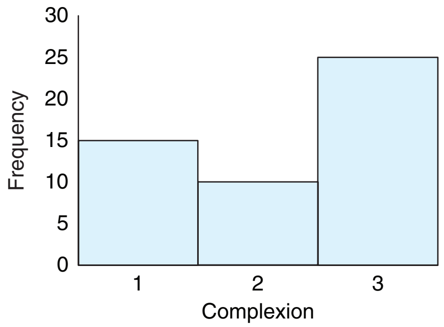
Country	Number
China	5453
Guatemala	4728
Russia	2310
Ethiopia	1255
South Korea	939

SECTION 2.5

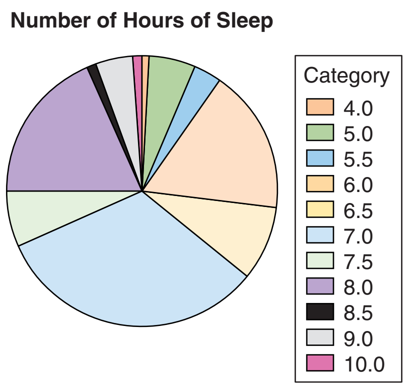
2.47 Pantry The accompanying graph shows the distribution of data on whether offices in a district have pantries. (A 1 indicates the office has a pantry and 0 indicates it does not have a pantry). Is this a bar graph or a histogram? How could the graph be improved?



2.48 Complexion A girl has gathered data on self-perceived complexion types, where 1 represents dark, 2 represents medium, and 3 represents light. A graph of these data is shown. What type of graph would be a better choice to display these data, and why?



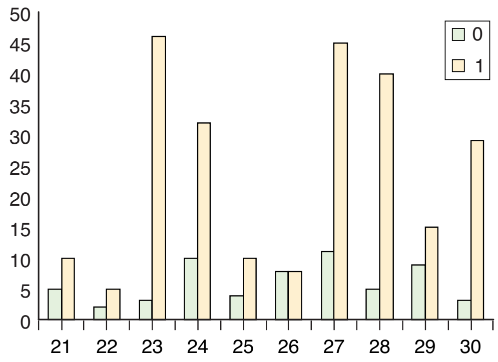
2.49 Pie Chart of Sleep Hours The pie chart reports the number of hours of sleep "last night" for 118 college students. What would be a better type of graph for displaying these data? Explain why this pie chart is hard to interpret.



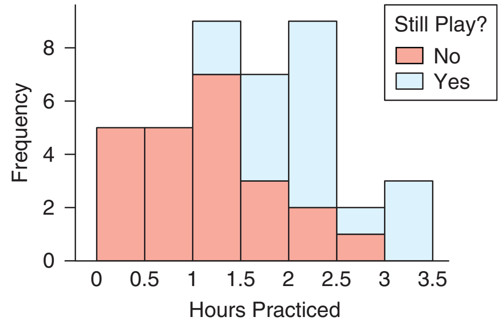
2.50 Literacy and Age The graph shows the ages of literates (labeled 1) and illiterates (labeled 0) who are living in a particular village.

a. Is this a histogram or a bar graph? How do you know?

b. What type(s) of graph(s) would be more appropriate?



2.51 Musicians Survey: StatCrunch Graph The accompanying graph is a special histogram with additional information; it was made using StatCrunch. People who studied music as children were asked how many hours a day they practiced when they were teenagers, and also whether they still play now that they are adults. To understand the graph, look at the third bar (spanning 1.0 to 1.5); it shows that there were seven people (the light red part of the bar) who practiced between 1.0 and 1.5 hours and did not still play as adults, and there were two people (the light blue part of the bar) who practiced 1.0 to 1.5 hours and still play as adults. Comment on what the graph shows. What other types of graphs could be used for this data set?



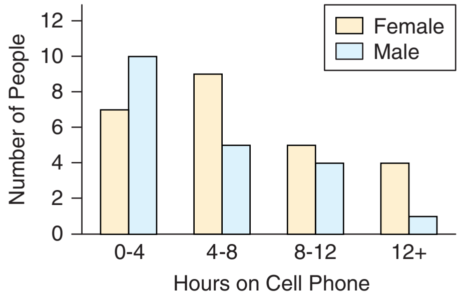
2.52 Cell Phone Use Refer to the accompanying bar chart, which shows the time spent on a typical day talking on the cell phone for some men and women. Each person was asked to choose the one of four intervals that best fitted the amount of time they spent on the phone (for example, "0 to 4 hours" or "12 or more hours").

* a. Identify the two variables. Then state whether they are categorical or numerical and explain.

b. Is the graph a bar chart or a histogram? Which would be the better choice for these data?

c. If you had the actual number of hours for each person, rather than just an interval, what type of graph should you use to display the distribution of the actual numbers of hours?

d. Compare the modes of the two distributions, and interpret what you discover: What does this say about the difference between men's and women's cell phone use?



CHAPTER REVIEW EXERCISES

2.53 PlayStations The table shows the first few entries for the number of hours spent per week playing games on PlayStations by some eighth-grade students, stacked and coded, where 1 represents a boy and 0 represents a girl.

What would be the appropriate graphs to compare the distributions of hours of games played on PlayStation per week for boys and girls if you had all the data? Explain.

Hours	Boy
8	0
12	1
10	1
3	1
9	0
8	1
6	1

2.54 Skills A factory has vacancies for various posts. What type(s) of graph(s) would be appropriate to compare the distribution of jobs for skilled and unskilled if you had all of the data? Explain.

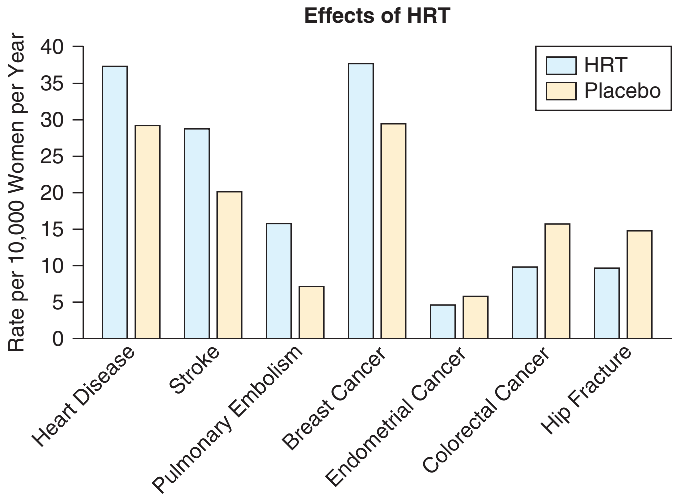
Skilled	Posts
0	Assistant
1	Welder
0	Storekeeper
1	Mechanic
1	Storekeeper

2.55 Hormone Replacement Therapy The use of the drug Prempro, a combination of two female hormones that many women take after menopause, is called hormone replacement therapy (HRT). In July 2002, a medical article reported the results of a study that was done to determine the effects of Prempro on many diseases. (Source: Women's Writing Group, Risks and benefits of estrogen plus progestin in healthy postmenopausal women, *JAMA*, 2002)

The study was placebo-controlled, randomized, and double-blind. From studies like these, it is possible to make statements about cause and effect. The figure shows comparisons of disease rates in the study.

a. For which diseases was the disease rate higher for those who took HRT? And for which diseases was the rate lower for those who took HRT?

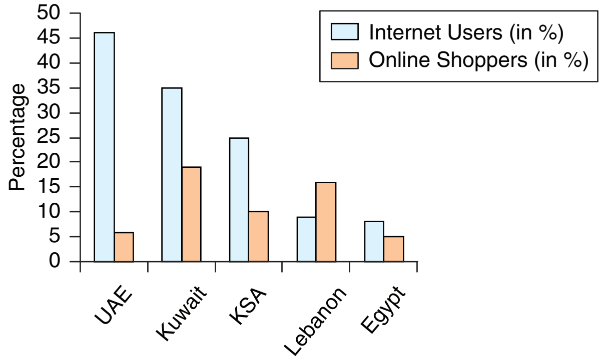
b. Why do you suppose we compare the rate per 10,000 women (per year), rather than just reporting the numbers of women observed who got the disease?



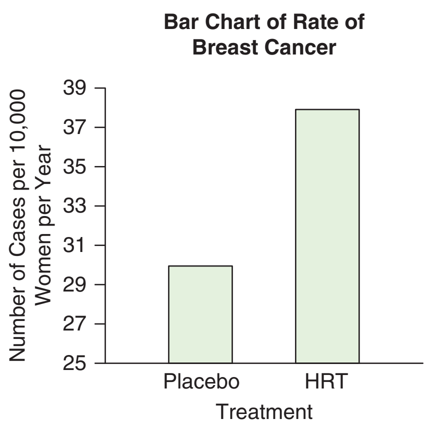
2.56 E-Commerce The bar graph shows information reported in tfour.me. For countries in the Middle East, the percentage of online purchase of goods and services and the percentage of people with access to the Internet are displayed.

a. Which two countries in the Middle East have the largest percentage of people with access to the Internet?

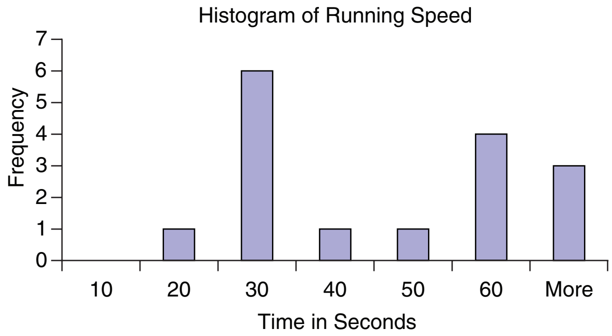
b. Which two countries in the Middle East have the largest percentage of online shoppers?



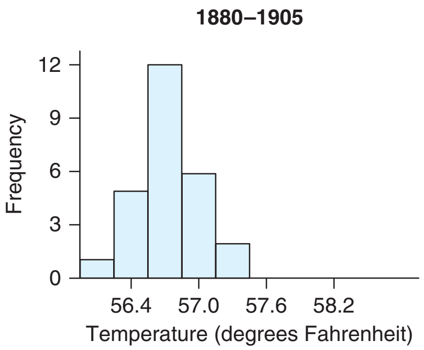
2.57 Hormone Replacement Therapy Again The bar chart shows a comparison of breast cancer rates for those who took HRT and those who took a placebo. Explain why the graph is deceptive, and indicate what could be done to make it less so.

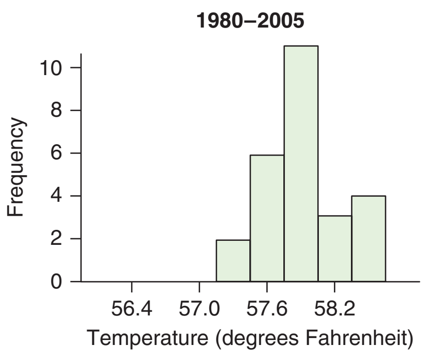


2.58 Running Speed A student was practicing for a track event and recorded the time taken (to complete running) in seconds. The times went from the best of 20 seconds to the worst of 75 seconds, as you can see in the stemplot. Suggest improvements to the histogram below generated by Excel, assuming that what is wanted is a histogram of the data (not a bar chart).

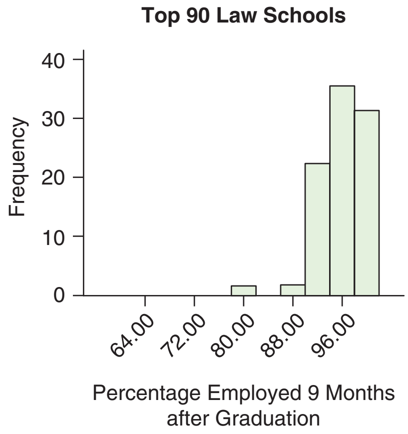


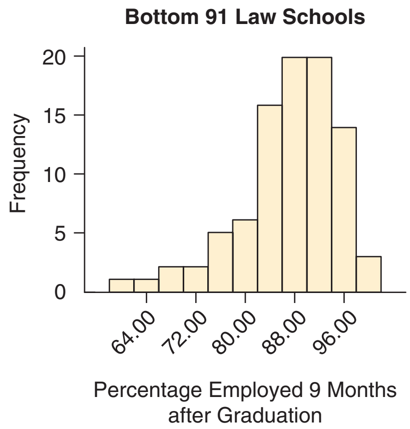
* **2.59 Global Temperatures** The histograms show the average global temperature per year for two 26-year ranges in degrees Fahrenheit. The range for 1880 to 1905 is on the top, and the range for 1980 to 2005 is on the bottom. Compare the two histograms for the two time periods, and explain what they show. Also estimate the difference between the centers. That is, about how much does the typical global temperature for the 1980–2005 time period differ from that for the 1880–1905 period? (Source: Goddard Institute for Space Studies, NASA, 2009)





* **2.60 Employment after Law School** Accredited law schools were ranked from 1 for the best (Harvard) down to number 181 by the Internet Legal Research Group (ILRG). When you decide on a law school to attend, one of the things you might be interested in is whether, after graduation, you will be able to get a job for which your law degree is required. We split the group of 181 law schools in half, with the top-ranked schools in one group and the lower-ranked schools in the other. The histograms show the distribution of the percentages of graduates who, 9 months after graduating, have obtained jobs that require a law degree. Compare the two histograms.

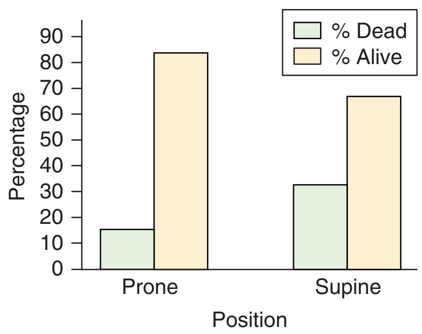




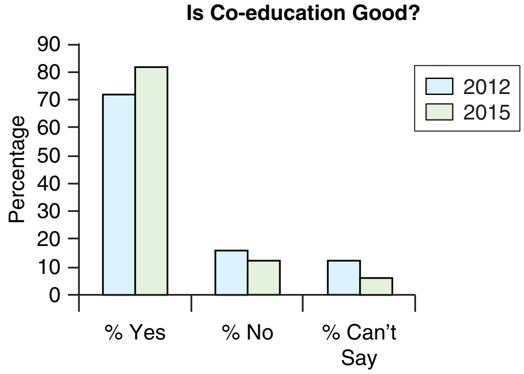
2.61 Position for Breathing The graph shows the results of a study done by Guerin et al. (2013) and reported in the *New England Journal of Medicine*. Patients arriving at emergency rooms who were having trouble breathing were randomly assigned to lie on their backs (supine) or on their stomachs (prone). For each of these groups, the graph displays the number of patients who died from acute respiratory distress syndrome (ARDS) and the number who did not die from this condition.

a. Based on this graph, which treatment would you recommend for these patients and why?

b. Explain why a histogram was not used.



2.62 Opinions on Co-education People were asked whether they thought co-education is good or not. The graph shows results of two surveys, one done in 2012 and the other done in 2015. *Yes* means they believe co-education is good and *No* means they do not believe that. What does the graph tell us about changes in opinions about co-education? Explain.



2.63 Create a dotplot that has at least 10 observations and is right-skewed.

2.64 Create a dotplot that has at least 10 observations and does not have skew.

2.65 Traffic Cameras College students Jeannette Mujica, Ricardo Ceja Zarate, and Jessica Cerda conducted a survey in Oxnard, California, of the number of cars going through a yellow light at intersections with and without traffic cameras that are used to automatically fine drivers who run red lights. The cameras were very noticeable to drivers. The amount of traffic was constant throughout the study period (the afternoon commute.) The data record the number of cars that crossed the intersection during a yellow light for each light cycle. A small excerpt of the data is shown in the table; see this text's website for all the data. What differences, if any, do you see between intersections with cameras and those without? Use an appropriate graphical summary, and write a comparison of the distributions.

# Cars	Cam
1	Cam
2	Cam
1	No Cam
3	No Cam

2.66 Ideal Weight Thirty-nine students (26 women and 13 men) reported their ideal weight (in most cases, not their current weight). The tables show the data.

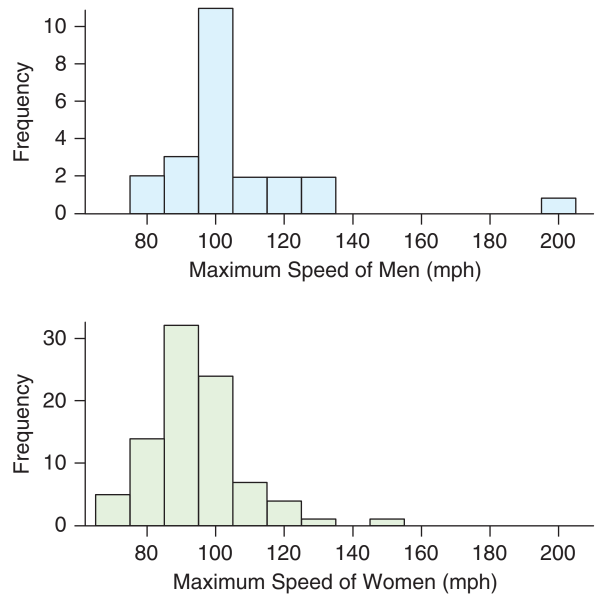
Women					
110	115	123	130	105	110
130	125	120	115	120	120
120	110	120	150	110	130
120	118	120	135	130	135
90	110				
Men					
160	130	220	175	190	190
135	170	165	170	185	155
160					

a. Explain why the distribution of ideal weights is likely to be bimodal if men and women are both included in the sample.

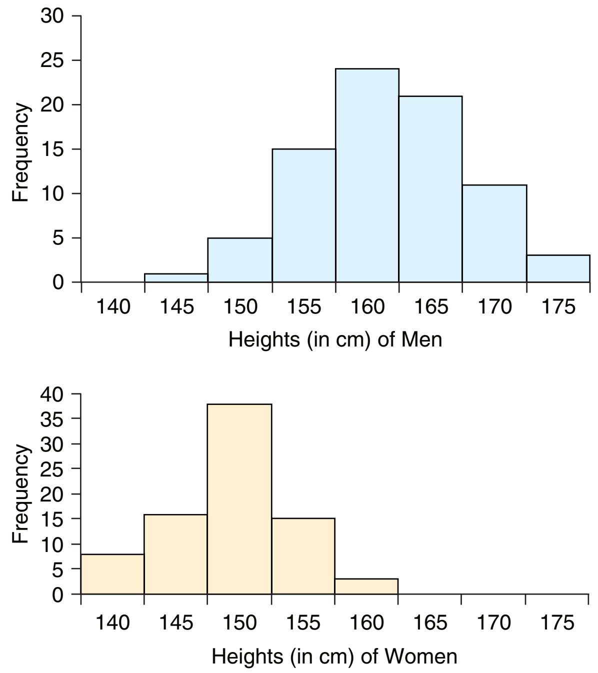
b. Make a histogram combining the ideal weights of men and women. Use the default histogram provided by your software. Report the bin width and describe the distribution.

c. Vary the number of bins, and print out a second histogram. Report the bin width and describe this histogram. Compare the two histograms.

2.67 MPH The graphs show the distribution of self-reported maximum speed ever driven by men and women college students who drive. Compare shapes, centers, and spreads, and mention any outliers.



2.68 Heights The graphs show heights of men and women working in an office. Compare shapes, centers, and spreads, and indicate whether there are outliers.



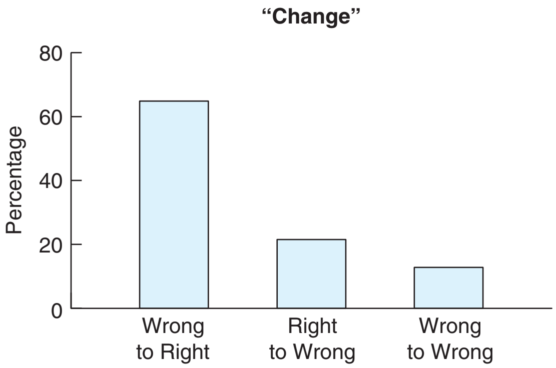
2.69 Ages of Billionaires Predict the shape of the distribution of the ages of 20 billionaires. A typical value is about 54 years, but there is an outlier at about 32 years.

2.70 Water A coach asks all of his trainees to report the approximate quantity of water they drink in a day. Predict the shape of the distribution of the quantity of water drunk per day.

2.71 Changing Multiple-Choice Answers When Told to Do So One of the authors wanted to determine the effect of changing answers on multiple-choice tests. She had advised her students that if they had changed their minds about a previous answer, they should replace their first choice with their new choice. By looking for erasures on the exam, she was able to count the number of changed answers that went from wrong to right, from right to wrong, and from wrong to wrong. The results are shown in the bar chart.

a. Do the data support her view that it is better to replace your initial choice with the revised choice?

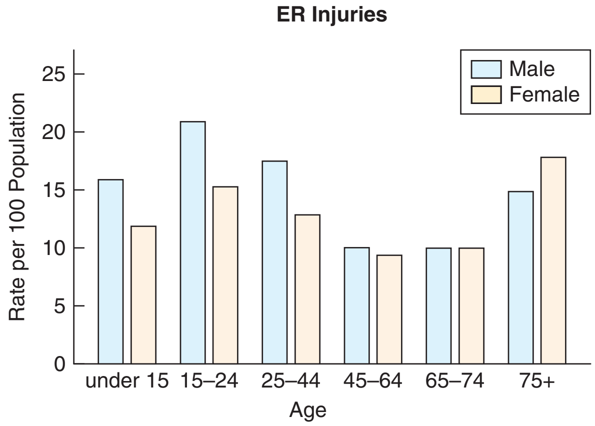
b. Compare this bar chart with the one in Exercise 2.37. Does changing answers generally tend to lead to higher or to lower grades?



2.72 ER Visits for Injuries The graph shows the rates of visits to the ER for injuries by gender and by age. Note that we are concerned with the rate per 100 people of that age and gender in the population. (Source: National Safety Council 2004)

a. Why does the National Safety Council give us rates instead of numbers of visits?

b. For which ages are the males more likely than the females to have an ER visit for an injury? For which ages are the men and women similar? For which ages do the women have more visits?



GUIDED EXERCISES

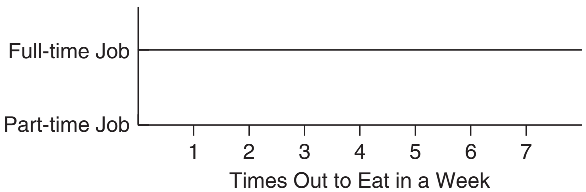
2.25 Eating Out and Jobs College student Jacqueline Loya asked students who had full-time jobs and students who had part-time jobs how many times they went out to eat in the last month.

Full-time: 5, 3, 4, 4, 4, 2, 1, 5, 6, 5, 6, 3, 3, 2, 4, 5, 2, 3, 7, 5, 5, 1, 4, 6, 7

Part-time: 1, 1, 5, 1, 4, 2, 2, 3, 3, 2, 3, 2, 4, 2, 1, 2, 3, 2, 1, 3, 3, 2, 4, 2, 1

QUESTION Compare the two groups by following the steps below. Include appropriate graphics.

Step 1 ▶ Create graphs.
Make dotplots (or histograms) using the same axis with one set of data above the other, as shown in the figure.



Step 2 ▶ Examine shape.
What is the shape of the data for those with part-time jobs? What is the shape of the data for those with full-time jobs?

Step 3 ▶ Examine center.
Which group tends to go out to eat more often, those with full-time or those with part-time jobs?

Step 4 ▶ Examine variation.
Which group has a wider spread of data?

Step 5 ▶ Check for outliers.
Were there any numbers separated from the other numbers as shown in the dotplots? (In other words, were there any gaps in the dotplots?)

Step 6 ▶ Summarize.
Finally, in one or more sentences, compare the shape, center, and variation (and mention outliers if there were any).

TechTips

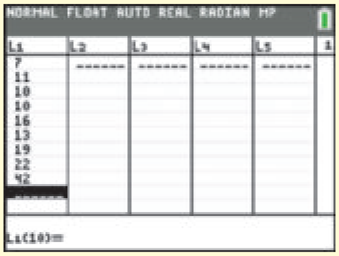
General Instructions for All Technology

EXAMPLE: ► Use the following ages to make a histogram:

7, 11, 10, 10, 16, 13, 19, 22, 42

Columns

Data sets are generally put into columns, not rows. The columns may be called variables, lists, or something similar. Figure 2A shows a column of data.



▲ **FIGURE 2A** TI-84 Data Screen

TI-84

Resetting the Calculator (Clearing the Memory)

If you turn off the TI-84, it does not reset the calculator. All the previous data and choices remain. If you are having trouble and want to start over, you can reset the calculator.

1. **2nd Mem** (with the + sign)
2. **7** for **Reset**
3. **1** for **All RAM**
4. **2** for **Reset**

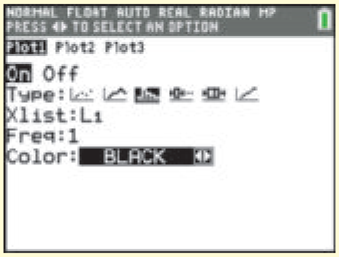
It will say **RAM cleared** if it has been done successfully.

Entering Data into the Lists

1. Press **STAT**, and select **EDIT** (by pressing **ENTER** when **EDIT** is highlighted).
2. If you find that there are data already in **L1** (or any list you want to use), you have three options:
   a. Clear the entire list by using the arrow keys to highlight the **L1** label and then pressing **CLEAR** and then **ENTER**. Do not press **DELETE**, because then you will no longer have an **L1**. If you delete **L1**, to get it back you can **Reset** the calculator.
   b. Delete the individual entries by highlighting the top data entry, then pressing **DELETE** several times until all the data are erased. (The numbers will scroll up.)
   c. Overwrite the existing data. CAUTION: Be sure to **DELETE** data in any cells not overwritten.
3. Type the numbers from the example into **L1** (List1). After typing each number, you may press **ENTER** or use ▼ (the arrow down on the keypad). Double-check your entries before proceeding.

Histogram

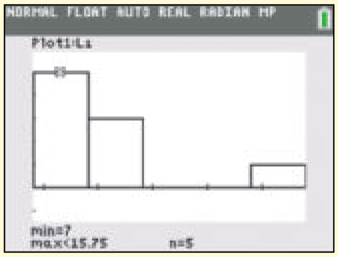
1. Press **2nd**, **STATPLOT** (which is in the upper left corner of the keypad).
2. If more than the first plot is **On**, press **4** (for **PlotsOff**) and **ENTER** to turn them all off.
3. Press **2nd**, **STATPLOT**, and **1** (for **Plot 1**).
4. Turn on **Plot1** by pressing **ENTER** when **On** is flashing; see Figure 2B.



▲ **FIGURE 2B** TI-84

5. Use the arrows on the keypad to get to the histogram icon (highlighted in Figure 2B) and press **ENTER** to choose it. Use the down arrow to get to Xlist and press **2nd** and **1** for **L1**. The settings shown in Figure 2B will lead to a histogram of the data in List 1 (L1).
6. Press **GRAPH**, press **ZOOM** and **9** (**ZoomStat**) to create the graph.
7. To see the numbers shown in Figure 2C, press **TRACE** (in the top row on the keypad) and move around using the group of four arrows on the keypad to see other numbers.

   Figure 2C shows a histogram of the numbers in the example.



▲ **FIGURE 2C** TI-84 Histogram

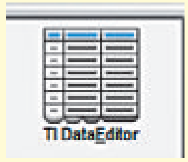
The TI-84 cannot make stemplots, dotplots, or bar charts.

Downloading Numerical Data from a Computer into a TI-84

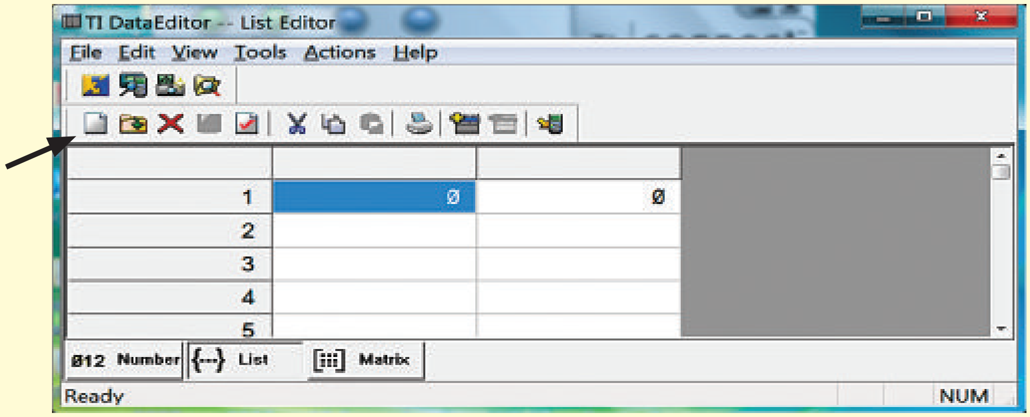
Before you can use your computer with your TI-84, you must install (on your computer) the software program TI Connect™, and a driver for the calculator. You need to do this only once. If you have done steps 1–3 and are ready to download the data, start with step 4.

1. Downloading and saving the TI Connect™ setup program.
   Insert the CD that came with the TI-84 into the computer disk drive and follow the on screen instructions. This will copy the setup file into the Downloads folder in your hard drive. (If you no longer have the CD, the programs can be downloaded free from the TI website, www.TI.com.)
2. Installing TI Connect™.
   Click on the globe in the lower left corner of your desktop and in the **Search** box, type **Downloads**. When you get to the **Downloads**, double click on **TIConnect** to begin the installation. Follow the on screen instructions.
   This should also create a TI Connect icon on your desktop screen.

3. Installing the calculator driver on the computer.
   To do this, connect the TI-84 to the computer using the USB cable that came with the calculator; then follow the on screen instructions.
   This completes the installation and the CD may now be removed from the computer
4. Double click the TI Connect icon on your desktop screen. If it doesn't exist, Click **Start**, **Programs**, **TI Tools**, and **TI-Connect**.

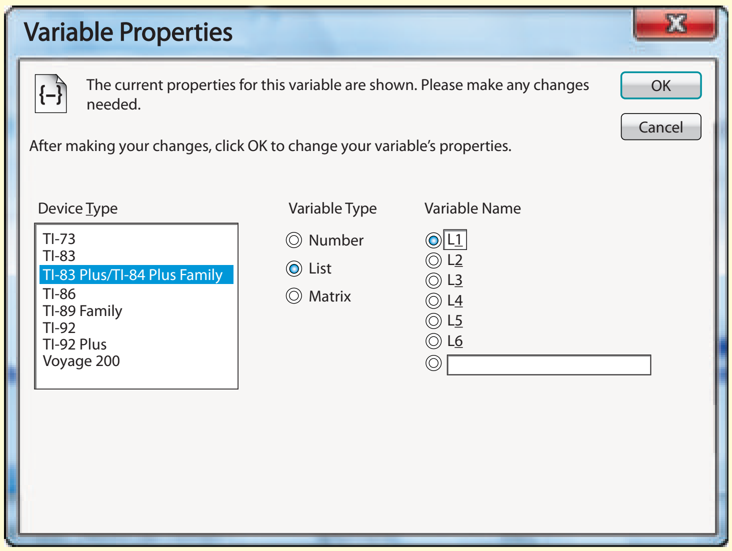


5. Click on **TI-Data Editor**.
6. Refer to Figure 2D.
   Click on the icon for the white sheet of paper. The arrow points to it. This will give you a white column to use for the data. If you have more than one column of data click on the piece of paper until you have the correct number of columns for your use. Figure 2D shows two columns.



▲ **FIGURE 2D** TI Data Editor

7. Copy a column or more than one column of *numerical* data from your computer. You may use Excel, Minitab, or any spreadsheet for the source of the column(s) of numbers, but do not include any labels or words. (If there are any letters or special characters in the column, they will not transfer and may show up as zeros.) Then click on the 0 in the first cell of the column (see Figure 2D). The cell *must* become colored (blue, as shown in the figure) before pasting. If it is not colored, click out of the column (for example, on the gray area to the right of the column) and then click in the first cell of the first column again. When that cell is blue, paste the column(s) of numbers into the TI-84 data editor. Alternatively, you can just type numbers in the column on the data editor.
8. While your cursor is in the column you want to name, choose **File** and **Properties** from the **Data Editor**. Then refer to Figure 2E: Check a list number like **L1** as shown (or, if you want a name, it cannot be more than 8 characters) and click **OK**. Then go back to any other column and do the same, but pick a different name such as **L2**.



▲ **FIGURE 2E** Variable Properties

9. Connect the TI-84 to the computer with the cable and turn on the TI-84 calculator.
10. Refer to Figure 2D. To paste the column(s) of data click **Actions, Send All Lists**.
11. When you get the **Warning**, click **Replace or Replace All** to overwrite the old data in the lists.

Look in your calculator lists (**STAT**, **EDIT**) to see the data there. If the data are not there, check the cable connection, check that the TI-84 is turned on, and start again with step 6.
Caution: While the data are transferring, you will not be able to use the calculator; it is thinking.

MINITAB

Entering the Data

When you open Minitab, you will see a blank spreadsheet for entering data. Type the data from the example into **C1**, column 1. Be sure your first number is put into Row 1, not above Row 1 in the label region. Be sure to enter only numbers. (If you want a label for the column, type it in the label region *above* the numbers.) You may also paste in data from the computer clipboard. Double-check your entries before proceeding.

All Minitab Graphs

After making the graph, double click on what you want to change, such as labels.

Histogram

1. Click **Graph > Histogram**
2. Leave the default option **Simple** and click **OK**.
3. Double click **C1** (or the name for the column) and click **OK**. (Another way to get **C1** in the big box is to click **C1** and click **Select**.)
4. After obtaining the histogram, if you want different bins (intervals), double click on the *x*-axis and look for **Binning**.

Figure 2F shows a Minitab histogram of the ages.

Stemplot

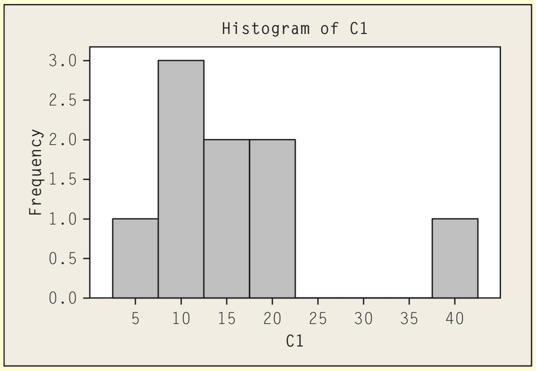
Click **Graph > Stem-and-Leaf**

Dotplot

Click **Graph > Dotplot**

Bar Chart

Click **Graph > Bar Chart**



▲ **FIGURE 2F** Minitab Histogram

EXCEL

When you open Excel and choose (click on) **Blank workbook**, you will see a blank spreadsheet for entering data. Before entering data for the first time, click the **Data** tab (top of screen, middle); you should see **Data Analysis** just below and on the far right. If you do not see it, you will need to load the Data Analysis Toolpak (instructions below). Now click the **Add-Ins** tab; you should see the XLSTAT icon ⊠ just below and on the far left. If you do not see it, you should install XLSTAT (instructions below). You will need both of these add-ins in order to perform all of the statistical operations described in this text.

Data Analysis Toolpak

1. Click **File** > **Options** > **Add-Ins**.
2. In **Manage** Box, select **Excel Add-Ins**, and click **Go**.
3. In **Add-ins available** box, check **Analysis Toolpak**, and click **OK**.

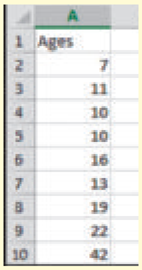
XLSTAT

1. Close Excel.
2. Download XLSTAT from www.myPearsonstore.com.
3. Install XLSTAT.
4. Open Excel, click **Add-Ins tab**, and click the XLSTAT icon ⊠.

You only need to install the Data Analysis Toolpak once. The Data Analysis tab should be available now every time Excel is opened. For XLSTAT, however, step 4 above may need to run each time Excel is opened (if you expect to run XLSTAT routines).

Entering Data

See Figure 2G. Enter the data from the example into column A with **Ages** in cell A1. Double-check your entries before proceeding.

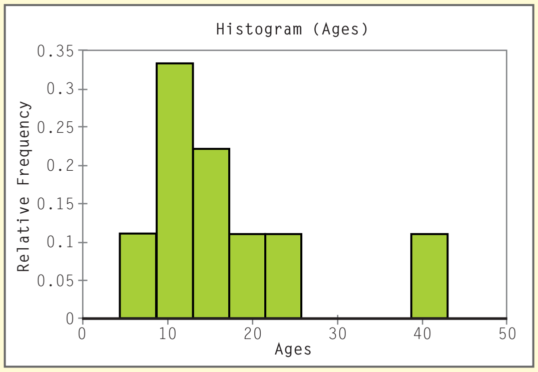


▲ **FIGURE 2G** Excel Data Screen

Histogram

1. Click **Add-Ins**, **XLSTAT**, **Describing data**, and **Histograms**.
2. When the box under **Data** is activated, drag your cursor over the column containing the data including the label **Ages**.
3. Select **Continuous**, **Sheet**, and **Sample labels**. Click **OK**, **Continue**, and **OK**.

Figure 2H shows the histogram.



▲ **FIGURE 2H** XLSTAT Histogram

Dotplot and Stemplot

1. Click **Add-Ins**, **XLSTAT**, **Visualizing data**, and **Univariate plots**.
2. When the box under **Quantitative Data** is activated, click on Sheet 1 (near the bottom of your screen, on the left side), and then drag your cursor over the column containing the data including the label **Ages**. Select **Sheet** and **Sample labels**.

Dotplot

3. Click **Options, Charts, Charts(1)** and select **Scattergrams** and **Horizontal**.
4. Click **OK** and **Continue**.

Stemplot

3. Click **Options, Charts, Charts(1)** and select **Stem-and-leaf plots**.
4. Click **OK** and **Continue**.

Bar Chart

After typing a summary of your data in table form (including labels), drag your cursor over the table to select it, click **Insert**, **Column** (in the **Charts** group), and select the first option.

STATCRUNCH

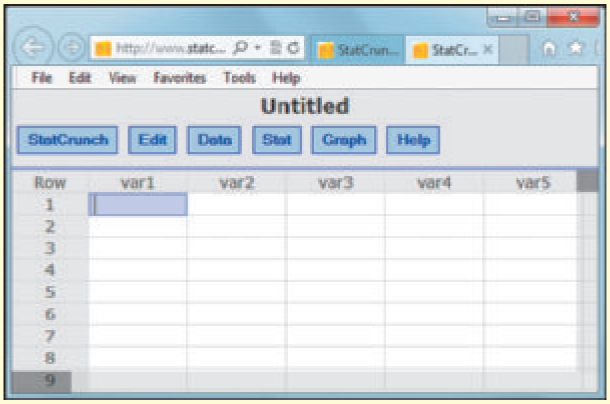
For Help: After logging in, click on **Help** or **Resources** and **Watch StatCrunch Video Tutorials on YouTube**.

Entering Data

1. Click **Open StatCrunch** and you will see a spreadsheet as shown in Figure 2I.
2. Enter the data from the example into the column labeled var1.
3. If you want labels on the columns, click on the variable label, such as **var1**, and backspace to remove the old label and type the new label. Double-check your entries before proceeding.

Pasting Data

1. If you want to paste data from your computer clipboard, click in an empty cell in Row 1 and press **Ctrl+V** on your keyboard.
2. If your clipboard data include labels, click on the **var1** cell instead, and then press **Ctrl+V** on your keyboard.



▲ **FIGURE 2I** StatCrunch Data Table

Histogram

1. Click **Graph > Histogram**
2. Under **Select columns**, click the variable you want a histogram for.
3. Click **Compute!**
4. To copy the graph for pasting into a document for submission, click **Options** and **Copy** and then paste it into a document.

Figure 2J shows the StatCrunch histogram of the ages.

Stemplot

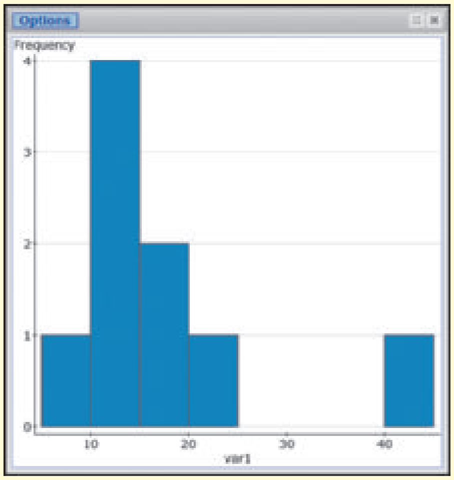
Click **Graph > Stem and Leaf**

Dotplot

Click **Graph > Dotplot**

Bar Chart

Click **Graph > Bar Plot > with data**



▲ **FIGURE 2J** StatCrunch Histogram

3 Numerical Summaries of Center and Variation

THEME

The complexity of numerical distributions can often be neatly summarized. In many cases, two numbers—one to measure the typical value and one to measure the variability—are all we need to summarize a set of data and make comparisons between groups. We can use many different ways of measuring what is typical and also of measuring variability; which method is best to use depends on the shape of the distribution.

Google the phrase *average American* and you'll find lots of entertaining facts. For example, the average American lives within 3 miles of a McDonald's, showers 10.4 minutes a day, and prefers smooth peanut butter to chunky. The average American woman is 5'4" tall and weighs 140 pounds, while the average American female fashion model is considerably taller and weighs much less: 5'11" and 117 pounds.

Whether or not these descriptions are correct, they are attempting to describe a typical American. The reason for doing this is to try to understand a little better what Americans are like, or perhaps to compare one group (American women) to another (female fashion models).

These summaries can seem odd, because we all know that people are too complex to be summarized with a single number. Characteristics such as weight, distance from a McDonald's, and length of a shower vary quite a bit from person to person. If we're describing the "typical" American, shouldn't we also describe how much variation exists among Americans? If we're making comparisons between groups, as we do in this chapter's Case Study, how do we do it in a way that takes into account the fact that individuals may vary considerably?

In Chapter 2, we talked about looking at graphs of distributions of data to get an intuitive, informal sense of the typical value (the center) and the amount of variation (the spread). In this chapter, we explore ways of making these intuitive concepts more precise by assigning numbers to them. We will see how this step makes it much easier to compare and interpret sets of data, for both symmetric and skewed distributions. These measures are important tools that we will use throughout the text.

CASE STUDY

Living in a Risky World

Our perception of how risky an activity is plays a role in our decision making. If you think that flying is very risky, for example, you will be more willing to put up with a long drive to get where you want to go. A team of UCLA psychologists were interested in understanding how people perceive risk and whether a simple reporting technique was enough to detect differences in perceptions between groups. The researchers asked over 500 subjects to consider various activities and rate them in terms of how risky they thought the activities were. The ratings were on a scale of 0 (no risk) to 100 (greatest possible risk). For example, subjects were asked to assign a value to the following activities: "Use a household appliance" and "Receive a diagnostic X-ray every 6 months." One question of interest to the researchers was whether men and women would assign different risk levels to these activities (Carlstrom et al. 2000).

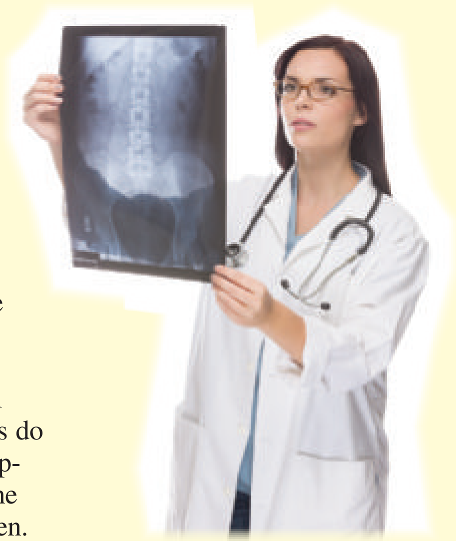
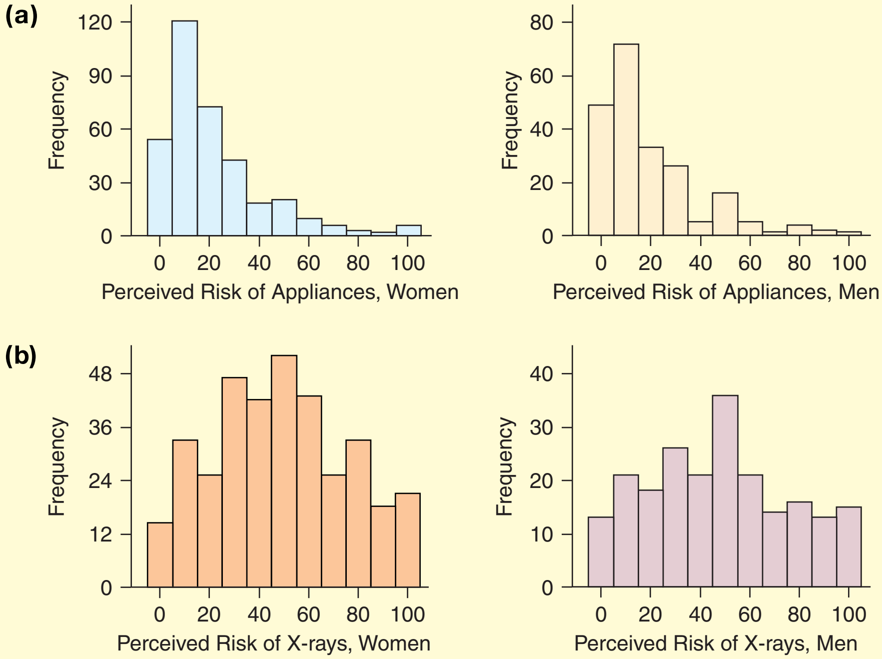


Figure 3.1a on the next page shows a histogram for the perceived risk of using household appliances, and Figure 3.1b shows a histogram for the risk of twice-annual X-rays. (Women are represented in the left panels, men in the right.) What differences do you see between the genders? How would you quantify these differences? In this chapter, you'll learn several techniques for answering these questions. And at the end of the chapter, we'll use these techniques to compare perceived risk between men and women.

▶ **FIGURE 3.1** Histograms showing the distributions of perceived risk by gender, for two activities. **(a)** Perceived risk of using everyday household appliances. **(b)** Perceived risk of receiving a diagnostic X-ray twice a year. The left panel in each part is for women, and the right panel for men.



SECTION 3.1

Summaries for Symmetric Distributions

In Chapter 2 you learned that we can characterize the typical value of a distribution by the center of the distribution, and the variability in that distribution by the horizontal spread. We left these concepts somewhat vague, but now our goal is to quantify them—to measure these concepts with numbers. But, assigning a number to the center and spread of a distribution is not all that straightforward. Statisticians have different ways of thinking about both center and spread, and these different ways of thinking play different roles, depending on the context of the data.

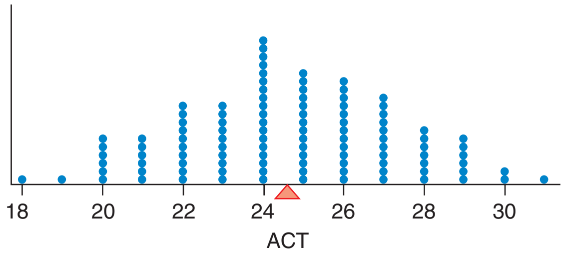
In this chapter, the two different ways in which we will think about the concept of center are (1) center as the balancing point (or center of mass), and (2) center as the halfway point. In this section we introduce the idea of the center as a balancing point, useful for symmetric distributions. Then in Section 3.3, we introduce the idea of the center as the halfway point, useful for skewed distributions. Each of these approaches results in a different measure, and our choice also affects the method we use to measure variability.

The Center as Balancing Point: The Mean

The most commonly used measure of center is the mean. The **mean** of a collection of data is the arithmetic average. The mean can be thought of as the balancing point of a distribution of data, and when the distribution is symmetric, the mean closely matches our concept of the "typical value."

> **Looking Back**
> **Symmetric Distributions**
> Recall that symmetric distributions are those in which the left-hand side of the graph of the distribution is roughly a mirror image of the right-hand side.

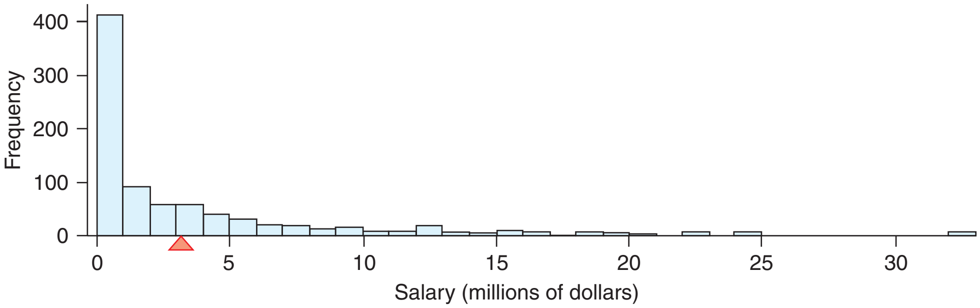
Visualizing the Mean Different groups of statistics students often have different backgrounds and levels of experience. Some instructors collect student data to help them understand whether one class of students might be very different from another. For example, if one class is offered earlier in the morning, or later in the evening, it might attract a different composition of students. An instructor at Peoria Junior College in Illinois collected data from two classes, including the students' ACT scores. (ACT is a standardized national college entrance exam.) Figure 3.2 shows the distribution of self-reported ACT scores for one statistics class.



◀ **FIGURE 3.2** The distribution of ACT scores for one class of statistics students. The mean is indicated here with a fulcrum (triangle). The mean is at the point on the dotplot that would balance if the points were placed on a seesaw. (Source: StatCrunch, Statistical_Data_499)

The balancing point for this distribution is roughly in the middle, because the distribution is fairly symmetric. The mean ACT for these students is calculated to be 24.6. You'll see how to do this calculation shortly. At the moment, rather than worrying about how to get the precise number, note that 24.6 is about the point where the distribution would balance if it were on a seesaw.

When the distribution of the data is more or less symmetric, the balancing point is roughly in the center, as in Figure 3.2. However, when the distribution is not symmetric, as in Figure 3.3, the balancing point is off-center, and the average may not match what our intuition tells us is the center.

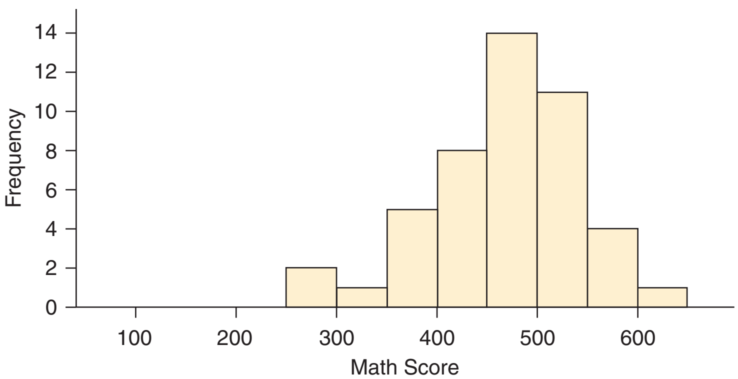


◀ **FIGURE 3.3** Salaries of professional baseball players in 2010.

Figure 3.3 shows a very skewed distribution: the salaries of professional baseball players in 2010. Salaries are in millions of dollars. (The highest paid player was Alex Rodriguez at $33 million. The lowest salary was $400,000, which 58 players received. The first bin includes over 400 players who earned $1 million dollars or less.) Where would you place the center of this distribution? Because the distribution is skewed to the right, the balancing point is fairly high; the mean is at 3.2 million dollars. However, when you consider that almost 70% of the players made less than this amount, you might not think that 3.2 million is what the "typical" player made. In other words, in this case the mean might not represent our idea of a typical salary, even for this group of famously well-paid professionals.

EXAMPLE 1 Math Scores

Figure 3.4 shows the distribution of math achievement scores for 46 countries as determined by the National Assessment of Education Progress. The scores are meant to measure the accomplishment of each country's eighth grade students with respect to mathematical achievement.



◀ **FIGURE 3.4** Distribution of international math scores for 46 countries. The scores measure the mean level of mathematical achievement for a country's eighth grade students.

QUESTION Based on the histogram, approximately what value do you think is the mean math achievement score?

SOLUTION The mean score is at the point where the histogram would balance if it were placed on a seesaw. The distribution looks roughly symmetric, and the balancing point looks to be somewhere between 400 and 500 points.

CONCLUSION The mean international math achievement score looks to be about 450 points. By the way, the U.S. score was 504, and the highest three scores were from Singapore (highest at 605), South Korea, and Hong Kong. The value of the mean tells us that the typical score, among the 46 countries, on the international math achievement test was between 400 and 500 points.

TRY THIS! Exercise 3.3

The Mean in Context As mentioned, the mean tells us the typical value in a data set with variability. We know that different countries had different math achievement scores, but typically, what is the math score of these countries? When you report the mean of a sample of data, you should not simply report the number but, rather, should report it in the context of the data so that your reader understands what you are measuring: The typical score on the international math test among these countries is between 400 and 500 points.

Knowing the typical value of one group enables us to compare this group to another group. For example, that same instructor at Peoria Junior College recorded ACT scores for a second classroom and found that the mean there was also 24.6. This tells us that these two classrooms were comparable, at least in terms of their mean ACT score. However, one classroom had a slightly younger mean age: 26.9 years compared to 27.2 years.

KEY POINT The mean of a collection of data is located at the "balancing point" of a distribution of data. The mean is one representation of the "typical" value of a variable.

Calculating the Mean for Small Data Sets The mean is used so often in statistics that it has its own symbol: $\bar{x}$, which is pronounced "x-bar." To calculate the mean, find the (arithmetic) **average** of the numbers; that is, simply add up all the numbers and divide that sum by the number of observations. Formula 3.1 shows you how to calculate the mean, or average.

$$\textbf{Formula 3.1:} \quad \text{Mean} = \bar{x} = \frac{\sum x}{n}$$

The symbol $\sum$ is the Greek capital sigma, or capital S, which stands for *summation*. The x that comes after $\sum$ represents the value of a single observation. Therefore, $\sum x$ means that you should add all the values. The letter n represents the number of observations. Therefore, this equation tells us to add the values of all the observations and divide that sum by the number of observations.

The mean shown in Formula 3.1 is sometimes called the sample mean in order to make it clear that it is the mean of a collection (or sample) of data.

EXAMPLE 2 Gas Buddy

According to GasBuddy.com (a website that invites people to submit prices at local gas stations), the prices of 1 gallon of regular gas at 12 service stations in a neighborhood in Austin, Texas, were as follows on one fall day in 2013:

\$3.19, \$3.09, \$3.09, \$2.93, \$2.95, \$3.09, \$2.99, \$2.99, \$2.95, \$2.99, \$2.99, \$2.97

A dotplot (not shown) indicates that the distribution is fairly symmetric.

QUESTION Find the mean price of a gallon of regular gas at these service stations. Explain what the value of the mean signifies in this context (in other words, interpret the mean).

SOLUTION Add the 12 numbers together to get \$36.22. We have 12 observations, so we divide \$36.22 by 12 to get \$3.018.

$$\bar{x} = \frac{3.19 + 3.09 + 3.09 + 2.93 + 2.95 + 3.09 + 2.99 + 2.99 + 2.95 + 2.99 + 2.99 + 2.97}{12}$$

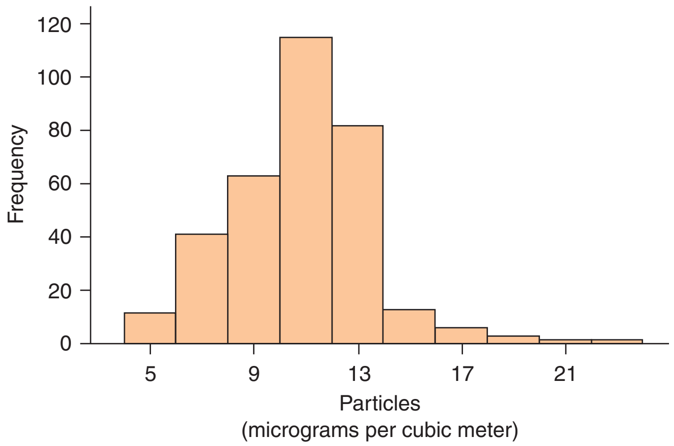
$$= \frac{36.22}{12} = 3.018$$

CONCLUSION The typical price of 1 gallon of gas at these gas stations in Austin, Texas, was \$3.02 on this particular day.

TRY THIS! Exercise 3.7a

Calculating the Mean for Larger Data Sets Formula 3.1 tells you how to compute the mean if you have a small set of numbers that you can easily type into a calculator. But for large data sets, you are better off using a computer or a statistical calculator. That is true for most of the calculations in this text, in fact. For this reason, we will often just display what you would see on your computer or calculator and describe in the TechTips section the exact steps used to get the solution.

For example, the histogram in Figure 3.5 shows the distribution of the amount of particulate matter, or smog, in the air in 333 cities in the United States in 2008, as reported by the Environmental Protection Agency. (The units are micrograms of particles per cubic meter.) When inhaled, these particles can affect the heart and lungs, so you would prefer your city to have a fairly low amount of particulate matter. (The EPA says that levels over 15 micrograms per cubic meter are unsafe.) Looking at the histogram, you can estimate the mean value using the fact that, since the distribution is fairly symmetric, the average will be about in the middle (around 11 micrograms per cubic meter). If you were given the list of 333 values, you could find the mean using Formula 3.1. But you'll find it easier to use the pre-programmed routines of your calculator or software. Example 3 demonstrates how to do this.



◀ **FIGURE 3.5** Levels of particulate matter for 333 U.S. cities in 2008. Because the distribution is fairly symmetric, the balancing point is roughly in the middle at about 11 micrograms per cubic meter.

EXAMPLE 3 Mean Smog Levels

Tech

We used four different statistical software programs to find the mean particulate level for the 333 cities. The data were uploaded into StatCrunch, Minitab, Excel, and the TI-84 calculator.

QUESTION For each of the computer outputs shown in Figure 3.6, find the mean particulate matter. Interpret the mean.

▶ **FIGURE 3.6(a)** StatCrunch output.

Column Statistics

Options

Summary statistics:

Column	n	Mean	Variance	Std. Dev.	Std. Err.	Median	Range	Min	Max	Q1	Q3
pm25_wtd	333	10.738439	6.6927953	2.5870438	0.14176913	10.8	19.1	4.4	23.5	9.2	12.3

▶ **FIGURE 3.6(b)** Minitab output.

Minitab Descriptive Statistics: pm25_wtd

```
Variable    N   N*    Mean  SE Mean  StDev  Minimum     Q1  Median      Q3
pm25_wtd  333  213  10.738    0.142  2.587    4.400  9.200  10.800  12.300

Variable  Maximum
pm25_wtd   23.500
```

▶ **FIGURE 3.6(c)** TI-84 calculator output.

```
NORMAL FLOAT AUTO REAL RADIAN MP
1-Var Stats
 x̄=10.73843844
 Σx=3575.9
 Σx²=40621.59
 Sx=2.587043708
 σx=2.583156337
 n=333
 minX=4.4
↓Q1=9.2
```

```
NORMAL FLOAT AUTO REAL RADIAN MP
1-Var Stats
↑Sx=2.587043708
 σx=2.583156337
 n=333
 minX=4.4
 Q1=9.2
 Med=10.8
 Q3=12.3
 maxX=23.5
```

▶ **FIGURE 3.6(d)** Excel output.

Column1	
Mean	10.73843844
Standard Error	0.141769122
Median	10.8
Mode	11.4
Standard Deviation	2.587043708
Sample Variance	6.692795145
Kurtosis	2.494762112
Skewness	0.542367196
Range	19.1
Minimum	4.4
Maximum	23.5
Sum	3575.9
Count	333

SOLUTION We have already seen, from Figure 3.5, that the distribution is close to symmetric, so the mean is a useful measure of center. With a large data set like this, it makes sense to use a computer to find the mean. Minitab and many other statistical software packages produce a whole slew of statistics with a single command, and your job is to choose the correct value.

CONCLUSION The software outputs give us a mean of 10.7 micrograms per cubic meter. We interpret this to mean that the typical level of particulate matter for these cities is 10.7 micrograms per cubic meter. For StatCrunch, Minitab, and Excel the mean is not hard to find; it is labeled clearly. For the TI output, the mean is labeled as $\bar{x}$, "x-bar."

TRY THIS! Exercise 3.11

KEY POINT In this text, pay more attention to how to apply and interpret statistics than to individual formulas. Your calculator or computer will nearly always find the correct values without your having to know the formula. However, you need to tell the calculator *what* to compute, you need to make sure the computation is meaningful, and you need to be able to explain what the result tells you about the data.

SNAPSHOT THE MEAN OF A SAMPLE

WHAT IS IT? ▶ A numerical summary.

WHAT DOES IT DO? ▶ Measures the center of the distribution of a sample of data.

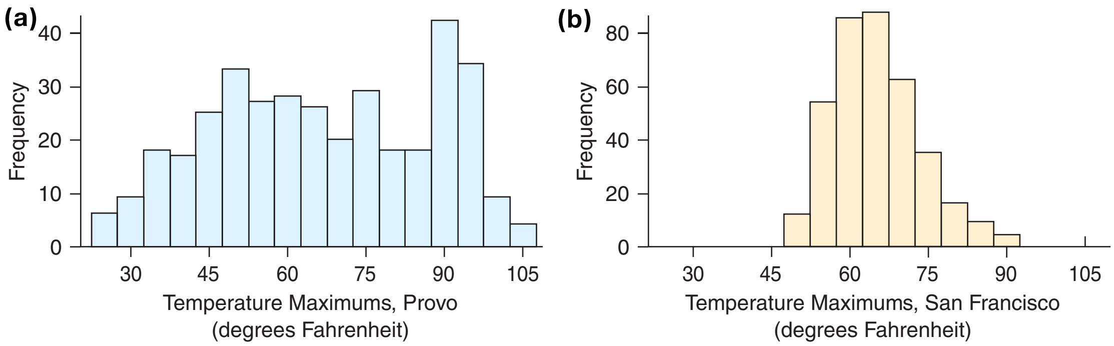
HOW DOES IT DO IT? ▶ The mean identifies the "balancing point" of the distribution, which is the arithmetic average of the values.

HOW IS IT USED? ▶ The mean represents the typical value in a set of data when the distribution is roughly symmetric.

Measuring Variation with the Standard Deviation

The mean amount of particulate matter in the air, 10.7 micrograms per cubic meter, does not tell the whole story. Even though the mean of the 333 cities is at a safe level (below 15 micrograms per cubic meter), this does not imply that any particular city—yours for example—is at a healthful level. Are most cities close to the mean level of 10.7? Or do cities tend to have levels of particulate matter far from 10.7? Values in a data set vary, and this variation is measured informally by the horizontal spread of the distribution of data. A measure of variability, coupled with a measure of center, helps us understand whether most observations are close to the typical value or far from it.

Visualizing the Standard Deviation The histograms in Figure 3.7 show daily high temperatures in degrees Fahrenheit recorded over one recent year at two locations in the United States. The histogram shown in Figure 3.7a records data collected in Provo, Utah, a city far from the ocean and at an elevation of 4500 feet.



▲ **FIGURE 3.7** Distributions of daily high temperatures in two cities: **(a)** Provo, Utah; **(b)** San Francisco. Both cities have about the same mean temperature, although the variation in temperatures is much greater in Provo than in San Francisco.

The histogram shown in Figure 3.7b records data collected in San Francisco, California, which sits on the Pacific coast and is famously chilly in the summer. (Mark Twain is alleged to have said, "The coldest winter I ever spent was a summer in San Francisco.")

The distributions of temperatures in these cities are similar in several ways. Both temperature distributions are fairly symmetric. You can see from the histograms that both cities have about the same mean temperature. For San Francisco the mean daily high temperature was about 65 degrees; for Provo it was about 67 degrees. But note the difference in the spread of temperatures!

One effect of the ocean on coastal climate is to moderate the temperature: The highs are not that high, and the lows not too low. This suggests that a coastal city should have less spread in the distribution of temperatures. How can we measure this spread, which we can see informally from the histograms?

Note that in San Francisco, most days were fairly close to the mean temperature of 65 degrees; rarely was it more than 10 degrees warmer or cooler than 65. (In other words, rarely was it colder than 55 or warmer than 75 degrees.) Provo, on the other hand, had quite a few days that were more than 10 degrees warmer or cooler than average.

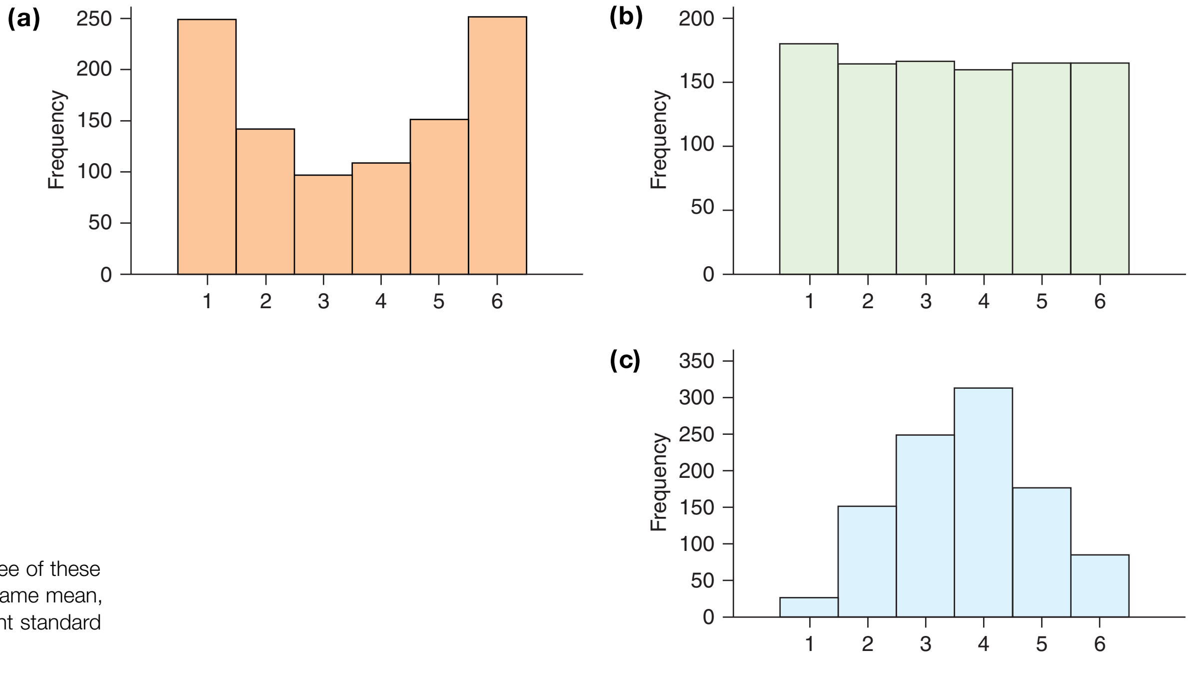
The **standard deviation** is a number that measures how far away the typical observation is from the mean. Distributions such as San Francisco's temperatures have smaller standard deviations because more observations are fairly close to the mean. Distributions such as Provo's have larger standard deviations because more observations are farther from the mean.

As you'll soon see, for most distributions, a majority of observations are within one standard deviation of the mean value.

KEY POINT The standard deviation should be thought of as the typical distance of the observations from their mean.

EXAMPLE 4 Comparing Standard Deviations from Histograms

Each of the three graphs in Figure 3.8 shows a histogram for a distribution of the same number of observations, and all the distributions have a mean value of about 3.5.



▶ **FIGURE 3.8** All three of these histograms have the same mean, but each has a different standard deviation.

QUESTION Which distribution has the largest standard deviation, and why?

SOLUTION All three groups have the same minimum and maximum values. However, the distribution shown in Figure 3.8a has the largest standard deviation. Why? The standard deviation measures how widely spread the points are from the mean. Note that the histogram in Figure 3.8a has the greatest number of observations farthest from the mean (at the values of 1 and 6). Figure 3.8c has the smallest standard deviation because so many of the data are near the center, close to the mean, which we can see because the taller bars in the center show us that there are more observations there.

CONCLUSION Figure 3.8a has the largest standard deviation, and Figure 3.8c has the smallest standard deviation.

TRY THIS! Exercise 3.15

The Standard Deviation in Context The standard deviation is somewhat more abstract, and harder to understand, than the mean. Fortunately, in a symmetric, unimodal distribution, a handy rule of thumb helps make this measure of spread more comprehensible. In these distributions, the majority of the observations (in fact, about two-thirds of them) are less than one standard deviation from the mean.

For temperatures in San Francisco (see Figure 3.7), the standard deviation is about 8 degrees, and the mean is 65 degrees. This tells us that in San Francisco, on a majority of days, the high temperature is within 8 degrees of the mean temperature of 65 degrees—that is, usually it is no colder than $65 - 8 = 57$ degrees and no warmer than $65 + 8 = 73$ degrees. In Provo, the standard deviation is substantially greater: 21 degrees. On a typical day in Provo, the high temperature is within 21 degrees of the mean. Provo has quite a bit more variability in temperature.

EXAMPLE 5 Standard Deviation of Smog Levels

The mean particulate matter in the 333 cities graphed in Figure 3.5 is 10.7 micrograms per cubic meter, and the standard deviation is 2.6 micrograms per cubic meter.

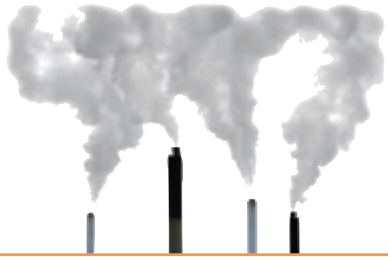
QUESTION Find the level of particulate matter one standard deviation above the mean and one standard deviation below the mean. Keeping in mind that the EPA says that levels over 15 micrograms per cubic meter are unsafe, what can we conclude about the air quality of most of the cities in this sample?

SOLUTION The typical city has a level of 10.7 micrograms per cubic meter, and because the distribution is unimodal and (roughly) symmetric, most cities have levels within 2.6 micrograms per cubic meter of this value. In other words, most cities have levels of particulate matter between

$$10.7 - 2.6 = 8.1 \text{ micrograms per cubic meter and}$$
$$10.7 + 2.6 = 13.3 \text{ micrograms per cubic meter}$$

CONCLUSION Because the value of 13.3 (one standard deviation above the mean) is lower than 15, most cities are below the safety limit. (The three cities reporting the highest levels of particulate matter were Phoenix, Arizona; Visalia, California; and Hilo, Hawaii. The three cities reporting the lowest levels of particulate matter were Cheyenne, Wyoming; Santa Fe, New Mexico; and Dickinson, North Dakota.)

As you'll soon see, we can say even more about this example. In a few pages you'll learn about the Empirical Rule, which tells us that about 95% of all cities should be within two standard deviations of the mean particulate level.



TRY THIS! Exercise 3.17

Calculating the Standard Deviation The formula for the standard deviation is somewhat more complicated than that for the mean, and a bit more work is necessary to calculate it. A calculator or computer is pretty much required for all but the smallest data sets. Just as the mean of a sample has its own symbol, the standard deviation of a sample of data is represented by the letter *s*.

$$\textbf{Formula 3.2:} \quad \text{Standard deviation} = s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

Think of this formula as a set of instructions. Essentially, the instructions say that we need first to calculate how far away each observation is from the mean. This distance, including the positive or negative sign, $(x - \bar{x})$, is called a **deviation**. We square these deviations so that they are all positive numbers, and then we essentially find the average. (If we had divided by n, and not $n - 1$, it would have been the average. We'll demonstrate in the Chapter 9 exercises why we divide by $n - 1$ and not n.) Finally, we take the square root, which means that we're working with the same units as the original data, not with squared units.

EXAMPLE 6 A Gallon of Gas

From the website GasBuddy.com, we collected the prices of a gallon of regular gas at 12 gas stations in a neighborhood in Austin, Texas, for one day in October 2013.

$3.19, $3.09, $3.09, $2.93, $2.95, $3.09, $2.99, $2.99, $2.95, $2.99, $2.99, $2.97

QUESTION Find the standard deviation for the prices. Explain what this value means in the context of the data.

SOLUTION We show this result two ways. The first way is by hand, which illustrates how to apply Formula 3.2. The second way uses a statistical calculator. The first step is to find the mean. We did this earlier in Example 2, using Formula 3.1, which gave us a mean value of $3.02. We substitute this value for $\bar{x}$ in Formula 3.2.

Table 3.1 shows the first two steps. First we find the deviations (in column 2). Next we square each deviation (in column 3). The numbers are sorted so we can more easily compare the differences.

x	$x - \bar{x}$	$(x - \bar{x})^2$
2.93	−0.09	0.0081
2.95	−0.07	0.0049
2.95	−0.07	0.0049
2.97	−0.05	0.0025
2.99	−0.03	0.0009
2.99	−0.03	0.0009
2.99	−0.03	0.0009
2.99	−0.03	0.0009
3.09	0.07	0.0049
3.09	0.07	0.0049
3.09	0.07	0.0049
3.19	0.17	0.0289

▲ TABLE 3.1

The sum of the squared deviations—the sum of column 3—is 0.0676. Dividing this by 11 (because $n - 1 = 12 - 1 = 11$), we get 0.006145455. The last step is to take the square root of this. The result is our standard deviation:

$$s = \sqrt{0.006145455} = 0.07839$$

To recap:

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} = \sqrt{\frac{0.0676}{12 - 1}} = \sqrt{\frac{0.0676}{11}} = \sqrt{0.006145455}$$

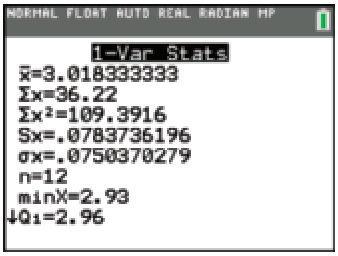
$$= 0.07839, \text{ or about 8 cents}$$

When you are doing these calculations, your final result will be more accurate if you do not round any of the intermediate results. For this reason, it is far easier, and more accurate, to use a statistical calculator or statistical software to find the standard deviation. Figure 3.9 shows the standard deviation as $Sx = 0.0783736196$, which we round to about 0.08 dollar (or, if you prefer, 8 cents). Note that the value reported by the calculator in Figure 3.9 is not exactly the same as the figure we obtained by hand. In part, this is because we used an approximate value for the mean in our hand calculations.

CONCLUSION The standard deviation is about 8 cents, or $0.08. Therefore, at most of these gas stations, the price of a gallon of gas is within 8 cents of $3.02.

TRY THIS! Exercise 3.19

Tech



▲ **FIGURE 3.9** The standard deviation is denoted "Sx" in the TI calculator output.

One reason why we suggest using statistical software rather than the formulas we present is that we nearly always look at data using several different statistics and approaches. We nearly always begin by making a graph of the distribution. Usually the next step is to calculate a measure of the center and then a measure of the spread. It does not make sense to have to enter the data again every time you want to examine them; it is much better to enter them once and use the functions on your calculator (or software).

SNAPSHOT THE STANDARD DEVIATION OF A SAMPLE

WHAT IS IT? ▶ A numerical summary.

WHAT DOES IT DO? ▶ Measures the spread of a distribution of a sample of data.

HOW DOES IT DO IT? ▶ It measures the typical distance of the observations from the mean.

HOW IS IT USED? ▶ To measure the amount of variability in a sample when the distribution is fairly symmetric.

Variance, a Close Relative of the Standard Deviation

Another way of measuring spread—a way that is closely related to the standard deviation—is the variance. The **variance** is simply the standard deviation squared, and it is represented symbolically by s^2.

Formula 3.3: $$\text{Variance} = s^2 = \frac{\sum(x - \bar{x})^2}{n - 1}$$

In Example 5, the standard deviation of the concentration of particulate matter in the cities in our sample was 2.6 micrograms per cubic meter. The variance is therefore $2.6 \times 2.6 = 6.76$ micrograms squared per cubic meter squared. The standard deviation in daily high temperatures in Provo is 21 degrees, so the variance is $21 \times 21 = 441$ degrees squared.

For most applications, the standard deviation is preferred over the variance. One reason is that the units for the variance are always squared (degrees squared in the last paragraph), which implies that the units used to measure spread are different from the units used to measure center. The standard deviation, on the other hand, has the same units as the mean.

SECTION 3.2

What's Unusual? The Empirical Rule and *z*-Scores

Finding the standard deviation and the mean is a useful way to compare different samples and to compare observations from one sample with those in another sample.

The Empirical Rule

The **Empirical Rule** is a rough guideline, a rule of thumb, that helps us understand how the standard deviation measures variability. This rule says that if the distribution is unimodal and symmetric, then

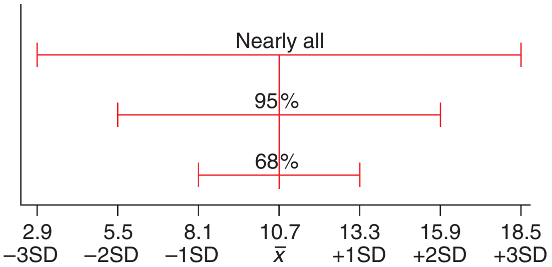
- Approximately 68% of the observations (roughly two-thirds) will be within one standard deviation of the mean.
- Approximately 95% of the observations will be within two standard deviations of the mean.
- Nearly all the observations will be within three standard deviations of the mean.

When we say that about 68% of the observations are within one standard deviation of the mean, we mean that if we count the observations that are between the mean minus one standard deviation and the mean plus one standard deviation, we will have counted about 68% of the total observations.

The Empirical Rule is illustrated in Figure 3.10 in the context of the data on particulate matter in 333 U.S. cities, introduced in Example 3. Suppose we did not have access to the actual data and knew only that the distribution is unimodal and symmetric, that the mean particulate matter is 10.7 micrograms per cubic meter, and that the standard deviation is 2.6 micrograms per cubic meter. The Empirical Rule predicts that about 68% of the cities will fall between 8.1 micrograms per cubic meter $(10.7 - 2.6 = 8.1)$ and 13.3 micrograms per cubic meter $(10.7 + 2.6 = 13.3)$.

The Empirical Rule predicts that about 95% of the cities will fall within two standard deviations of the mean, which means that about 95% of the cities will be between 5.5 and 15.9 micrograms per cubic meter $(10.7 - (2 \times 2.6) = 5.5$ and $10.7 + (2 \times 2.6) = 15.9)$. Finally, nearly all cities, according to the Empirical Rule, will be between 2.9 and 18.5 micrograms per cubic meter. This is illustrated in Figure 3.10.

KEY POINT In a relatively large collection of observations, if the distribution is unimodal and roughly symmetric, then about 68% of the observations are within one standard deviation of the mean; about 95% are within two standard deviations of the mean, and almost all observations are within three standard deviations of the mean. Not all unimodal, symmetric distributions are the same, so your actual outcomes might differ from these values, but the Empirical Rule works well enough in a surprisingly large number of situations.



▶ **FIGURE 3.10** The Empirical Rule predicts how many observations we will see within one standard deviation of the mean (68%), within two standard deviations of the mean (95%), and within three standard deviations of the mean (almost all).

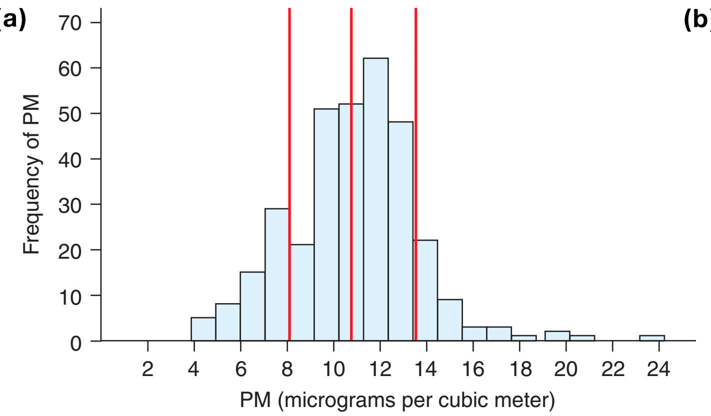
EXAMPLE 7 Comparing the Empirical Rule to Actual Smog Levels

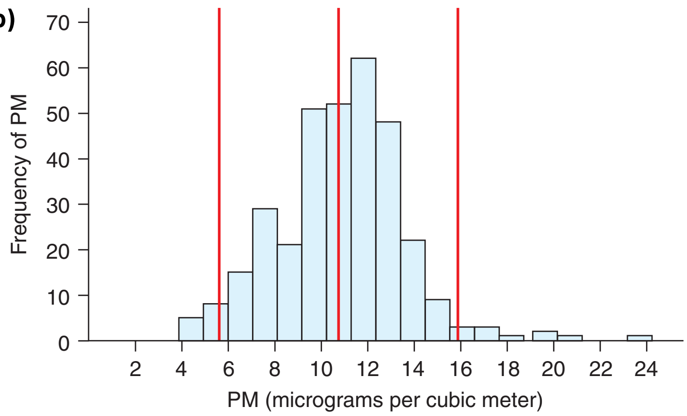
Because the distribution of levels of particulate matter (PM) in 333 U.S. cities is roughly unimodal and symmetric, the Empirical Rule predicts that about 68% of the cities will have particulate matter levels between 8.1 and 13.3 micrograms per cubic meter, about 95% of the cities will have PM levels between 5.5 and 15.9 micrograms per cubic meter, and nearly all will have PM levels between 2.9 and 18.5 micrograms per cubic meter.

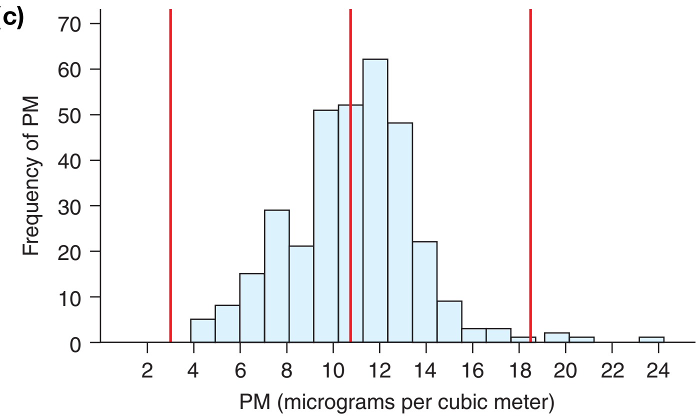
QUESTION Figure 3.11 shows the actual histograms for the distribution of PM levels for these 333 cities. The location of the mean is indicated, as well as the boundaries for points within one standard deviation of the mean (a), within two standard deviations of the mean (b), and within three standard deviations of the mean (c). Using these figures, compare the actual number of cities that fall within each boundary to the number predicted by the Empirical Rule.

SOLUTION From Figure 3.11a, by counting the heights of the bars between the two boundaries, we find that there are about $20 + 50 + 52 + 60 + 48 = 230$ cities that actually lie between these two boundaries. (No need to count very precisely; we're after approximate numbers here.) The Empirical Rule predicts that about 68% of the cities, or $0.68 \times 333 = 226$ cities, will fall between these two boundaries. The Empirical Rule is pretty accurate in this case.

From Figure 3.11b, we count that there are about $15 + 30 + 20 + 50 + 52 + 60 + 48 + 20 + 8 + 2 = 305$ cities within two standard deviations of the mean. The Empirical Rule predicts that about 95% of the cities, or $0.95 \times 333 = 316$ cities, will fall between these two boundaries. So again, the Empirical Rule is not too far off.

(a)


(b)


(c)


▲ **FIGURE 3.11** **(a)** Boundaries are placed one standard deviation below the mean (8.1) and one standard deviation above the mean (13.3). The middle line in all three figures indicates the mean (10.7). **(b)** The Empirical Rule predicts that about 95% of observations will fall between these two boundaries (5.5 and 15.9 micrograms per cubic meter). **(c)** The Empirical Rule predicts that nearly all observations will fall between these two boundaries (2.9 and 18.5 micrograms per cubic meter).

Finally, from Figure 3.11c, we clearly see that all but a few cities (about 4 or 5), are within three standard deviations of the mean. We summarize in a table.

CONCLUSION

Interval	Empirical Rule Prediction	Actual Number
within 1SD of mean	226	230
within 2SD of mean	316	305
within 3SD of mean	nearly all	all but 5

TRY THIS! Exercise 3.27

EXAMPLE 8 Temperatures in San Francisco

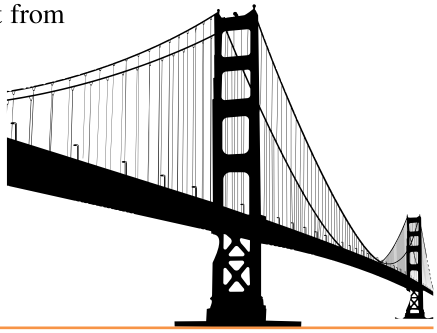
We have seen that the mean daily high temperature in San Francisco is 65 degrees Fahrenheit and that the standard deviation is 8 degrees.

QUESTION Using the Empirical Rule, decide whether it is unusual in San Francisco to have a day when the maximum temperature is colder than 49 degrees.

SOLUTION One way of answering this question is to find out about how many days in the year we would expect the high temperature to be colder than 49 degrees. The Empirical Rule says that about 95% of the days will have temperatures within two standard deviations of the average—that is, within $2 \times 8 = 16$ degrees of 65 degrees. Therefore, on most days the temperature will be between $65 - 16 = 49$ degrees and $65 + 16 = 81$ degrees. Because only (approximately) 5% of the days are outside this range, we know that days either warmer or cooler than this are rare. According to the Empirical Rule, which assumes a roughly symmetric distribution, about half the days outside the range—2.5%—will be colder, and half will be warmer. This shows that about 2.5% of 365 days (roughly 9 or 10 days) should have a maximum temperature colder than 49 degrees.

CONCLUSION A maximum temperature colder than 49 degrees is fairly unusual for San Francisco. According to the Empirical Rule, only about 2.5% of the days should have maximum temperatures below 49 degrees, or about 9 or 10 days out of a 365-day year. Of course, the Empirical Rule is only a guide. In this case, we can compare with the data to see just how often these cooler days occurred.

Your concept of "unusual" might be different from ours. The main idea is that "unusual" is rare, and in selecting two standard deviations, we chose to define temperatures that occur 2.5% of the time or less as rare and therefore unusual. You might very reasonably set a different standard for what you wish to consider unusual.



TRY THIS! Exercise 3.29

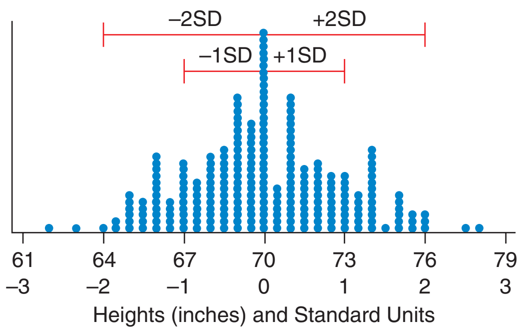
z-Scores: Measuring Distance from Average

The question "How unusual is this?" is perhaps the statistician's favorite question. (It is just as popular as "Compared to what?") Answering this question is complicated because the answer depends on the units of measurement. Eighty-four is a big value if we are measuring a man's height in inches, but it is a small value if we are measuring his weight in pounds. Unless we know the units of measurement and the objects being measured, we can't judge whether a value is big or small.

One way around this problem is to change the units to standard units. **Standard units** measure a value relative to the sample rather than with respect to some absolute measure. A measurement converted to standard units is called a ***z*-score**.

Visualizing z-Scores Specifically, a standard unit measures how many standard deviations away an observation is from the mean. In other words, it measures a distance, but instead of measuring in feet or miles, it counts the number of standard deviations. A measurement with a *z*-score of 1.0 is one standard deviation above the mean. A measurement with a *z*-score of −1.4 is 1.4 standard deviations below the mean.

Figure 3.12 shows a dotplot of the heights (in inches) of 247 adult men. The average height is 70 inches, and the standard deviation is 3 inches (after rounding). Below the dotplot is a ruler that marks off how far from average each observation is, measured in terms of standard deviations. The average height of 70 inches is marked as 0 because 70 is zero standard deviations away from the mean. The height of 76 is marked as 2 because it is two standard deviations above the mean. The height of 67 is marked as −1 because it is one standard deviation *below* the mean.



◀ **FIGURE 3.12** A dotplot of heights of 247 men marked with *z*-scores as well as heights in inches.

We would say that a man from this sample who is 73 inches tall has a *z*-score of 1.0 standard unit. A man who is 67 inches tall has a *z*-score of −1.0 standard unit.

Using z-Scores in Context *z*-Scores enable us to compare observations in one group with those in another, even if the two groups are measured in different units or under different conditions. For instance, some students might choose their math class on the basis of which professor they think is an easier grader. So if one student gets a 65 on an exam in a hard class, how do we compare her score to that of a student who gets a 75 in an easy class? If we converted to standard units, we would know how far above (or below) the average each test score was, so we could compare these students' performances.

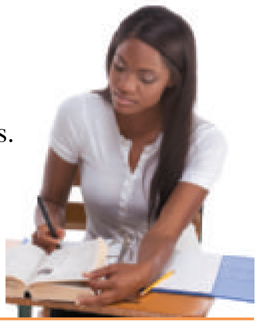
EXAMPLE 9 Exam Scores

Maria scored 80 out of 100 on her first stats exam in a course and 85 out of 100 on her second stats exam. On the first exam, the mean was 70 and the standard deviation was 10. On the second exam, the mean was 80 and the standard deviation was 5.

QUESTION On which exam did Maria perform better when compared to the whole class?

SOLUTION On the first exam, Maria is 10 points above the mean because 80 − 70 is 10. Because the standard deviation is 10 points, she is one standard deviation above the mean. In other words, her *z*-score for the first exam is 1.0.

On the second exam, she is 5 points above average because 85 − 80 is 5. Because the standard deviation is 5 points, she is one standard deviation above the mean. In other words, her *z*-score is again 1.0.



CONCLUSION The second exam was a little easier; on average, students scored higher, and there was less variability in the scores. But Maria scored one standard deviation above average on both exams, so she did equally well on both when compared to the whole class.

TRY THIS! Exercise 3.33

Calculating the z-Score It's straightforward to convert to *z*-scores when the result is a whole number, as in the last few examples. More generally, to convert a value to its *z*-score, first subtract the mean. Then divide by the standard deviation. This simple recipe is summarized in Formula 3.4 which we present in words and symbols:

Formula 3.4a: $$z = \frac{(\text{value} - \text{mean})}{\text{standard deviation}}$$

Formula 3.4b: $$z = \frac{(x - \bar{x})}{s}$$

Let's apply this to the data shown in Figure 3.12. What is the *z*-score of a man who is 75 inches tall? Remember that the mean height is 70 inches and the standard deviation is 3 inches. Formula 3.4 says first to subtract the mean height.

$$75 - 70 = 5 \text{ inches}$$

Next divide by the standard deviation:

$$5/3 = 1.67 \text{ (rounding off to two decimal digits.)}$$

$$z = \frac{x - \bar{x}}{s} = \frac{75 - 70}{3} = \frac{5}{3} = 1.67$$

This person has a *z*-score of 1.67. In other words, we would say that he is 1.67 standard deviations taller than average.

EXAMPLE 10 Daily Temperatures

The mean daily high temperature in San Francisco is 65 degrees F, and the standard deviation is 8 degrees. On one day, the high temperature is 49 degrees.

QUESTION What is this temperature in standard units? Assuming that the Empirical Rule applies, does this seem unusual?

SOLUTION

$$z = \frac{x - \bar{x}}{s} = \frac{49 - 65}{8} = \frac{-16}{8} = -2.00$$

CONCLUSION This is an unusually cold day. From the Empirical Rule, we know that 95% of *z*-scores are between –2 and 2 standard units, so it is fairly unusual to have a day as cold as or colder than this one.

TRY THIS! Exercise 3.35

SNAPSHOT THE *z*-SCORE

WHAT IS IT? ▶ A standardized observation.

WHAT DOES IT DO? ▶ Converts a measurement into standard units.

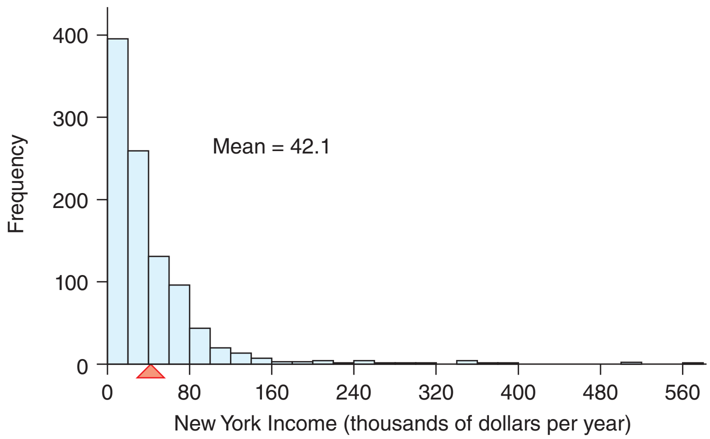
HOW DOES IT DO IT? ▶ By measuring how many standard deviations away a value is from the sample mean.

HOW IS IT USED? ▶ To compare values from different groups, such as two exam scores from different exams, or to compare values measured in different units, such as inches and pounds.

SECTION 3.3

Summaries for Skewed Distributions

As we noted earlier, for a skewed distribution, the center of balance is not a good way of measuring a "typical" value. Another concept of center, which is to think of the center as being the location of the *middle* of a distribution, works better in these situations. You saw one example of this in Figure 3.3, which showed that the mean baseball player salary was quite a bit higher than what a majority of the players actually earned. Figure 3.13 shows another example of a strongly right-skewed distribution. This is the distribution of incomes of 1000 New York State residents, drawn from a survey done by the U.S. government in 2011. The mean income of $42,101 is marked with a triangle. However, note that the mean doesn't seem to match up very closely with what we think of as typical. The mean seems to be too high to be typical. In fact, a majority (about 69%) of residents earn less than this mean amount.



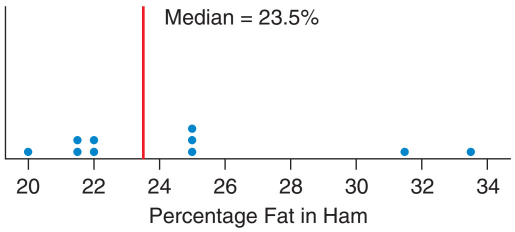
◀ **FIGURE 3.13** The distribution of annual incomes for a collection of New York State residents is right-skewed. Thus the mean is somewhat greater than what many people would consider a typical income. The location of the mean is shown with a triangle.

In skewed distributions, the mean can often be a poor measure of "typical." Instead of using the mean and standard deviation in these cases, we measure the center and spread differently.

The Center as the Middle: The Median

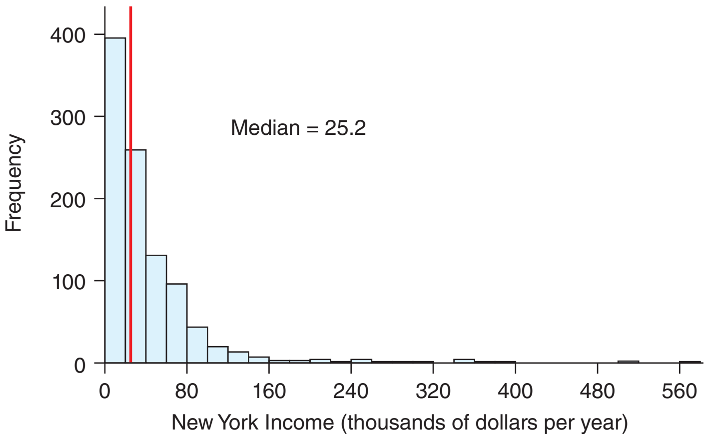
The median provides an alternative way of determining a typical observation. The **median** of a sample of data is the value that would be right in the middle if you were to sort the data from smallest to largest. The median cuts a distribution down the middle, so about 50% of the observations are below it and about 50% are above it.

Visualizing the Median One of the authors found herself at the grocery store, trying to decide whether to buy ham or turkey for sandwiches. Which is more healthy? Figure 3.14 shows a dotplot of the percentage of fat for each of ten types of sliced ham. The vertical line marks the location of the median at 23.5%. Note that five observations lie below the median and five lie above it. The median cuts the distribution exactly in half. In Example 12, you'll see how the median percentage of fat in sliced turkey compares to that in sliced ham.



▶ **FIGURE 3.14** A dotplot of percentage fat from ham has a median of 23.5%. This means that half the observations are below 23.5% and half are above it.

Finding the median for the distribution of incomes of the 1000 New York State residents, shown again in Figure 3.15, is slightly more complicated because we have many more observations. The median (shown with the red vertical line) cuts the total area of the histogram in half. The median is at $25,200, and very close to 50% of the observations are below this value and just about 50% are above it.

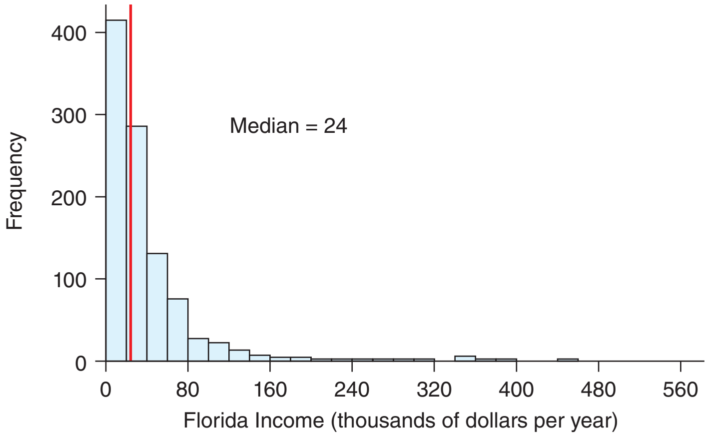


▶ **FIGURE 3.15** The distribution of incomes of New York State residents, with the median indicated by a vertical line. The median has about 50% of the observations above it and about 50% below it.

KEY POINT The median is the value that cuts a distribution in half. The median value represents a "typical" value in a data set.

The Median in Context The median is used for the same purpose as the mean: to give us a typical value of a set of data. Knowing the typical value of one group helps us to compare it to another. For example, as we've seen, the typical median income of this sample of New York State residents is $25,200. How does the typical income in New York compare to the typical income in Florida? A representative sample of 1000 Florida residents (where many New Yorkers go to retire) has a median income of $24,000, which is only slightly less than the median income in New York (Figure 3.16).

The typical person in our data set of New York incomes makes more than the typical person in our Florida data set, as measured by the median. Because the median for New York is $25,200, we know that more than half of New York residents in the sample make more than the median Florida salary of $24,000.



◀ **FIGURE 3.16** Distribution of incomes of a selection of Florida residents (in thousands of dollars). The median income is $24,000.

The median is often reported in news stories when the discussion involves variables with distributions that are skewed. For example, you may hear reports of "median housing costs" and "median salaries."

Calculating the Median To calculate the value of the median, follow these steps:

1. Sort the data from smallest to largest.
2. If the set contains an odd number of observed values, the median is the middle observed value.
3. If the set contains an even number of observed values, the median is the average of the two middle observed values. This places the median precisely halfway between the two middle values.

EXAMPLE 11 Twelve Gas Stations

The prices of a gallon of regular gas at 12 Austin, Texas, gas stations in October 2013 (see Example 6) were

$3.19, $3.09, $3.09, $2.93, $2.95, $3.09, $2.99, $2.99, $2.95, $2.99, $2.99, $2.97

QUESTION Find the median price for a gallon of gas and interpret the value.

SOLUTION First, we sort the data from smallest to largest.

2.93, 2.95, 2.95, 2.97, 2.99, 2.99, 2.99, 2.99, 3.09, 3.09, 3.09, 3.19

Because the data set contains an even number of observations (12), the median is the average of the two middle observations, the sixth and seventh: 2.99 and 2.99.

2.93, 2.95, 2.95, 2.97, 2.99, 2.99, 2.99, 2.99, 3.09, 3.09, 3.09, 3.19
Med

CONCLUSION The median is $2.99, which is the typical price of a gallon of gas at these 12 gas stations.

TRY THIS! Exercise 3.41a

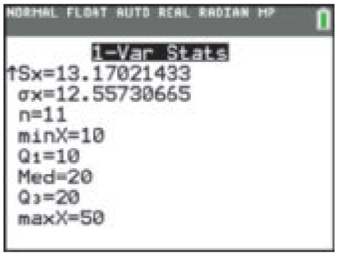
Example 12 demonstrates how to find the median in a sample with an odd number of observations.

EXAMPLE 12 Sliced Turkey

Figure 3.14 showed that the median percentage of fat from the various brands of sliced ham for sale at a grocery store was 23.5%. How does this compare to the median percentage of fat for the turkey? Here are the percentages of fat for the available brands of sliced turkey:

14, 10, 20, 20, 40, 20, 10, 10, 20, 50, 10

QUESTION Find the median percentage of fat and interpret the value.



▲ **FIGURE 3.17** Some TI-84 output for the percentage of fat in the turkey.

SOLUTION The data are sorted and displayed below. Because we have 11 observations, the median is the middle observation, 20.

10 10 10 10 14 20 20 20 20 40 50
Med

CONCLUSION The median for the turkey is 20% fat. Thus the typical percentage of fat for these types of sliced turkey is 20%. This is (slightly) less than that for the typical sliced ham, which has 23.5% fat. Figure 3.17 provides TI-84 output that confirms our calculation.

TRY THIS! Exercise 3.43

SNAPSHOT THE MEDIAN OF A SAMPLE

WHAT IS IT? ▶ A numerical summary.

WHAT DOES IT DO? ▶ Measures the center of a distribution.

HOW DOES IT DO IT? ▶ It is the value that has roughly the same number of observations above it and below it.

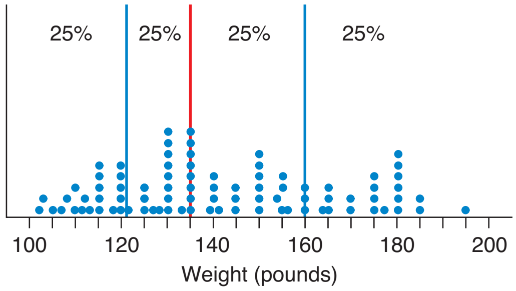
HOW IS IT USED? ▶ To measure the typical value in a data set, particularly when the distribution is skewed.

Measuring Variability with the Interquartile Range

The standard deviation measures how spread out observations are with respect to the mean. But if we don't use the mean, then it doesn't make sense to use the standard deviation. When a distribution is skewed and you are using the median to measure the center, an appropriate measure of variation is called the interquartile range. The **interquartile range (IQR)** tells us, roughly, how much space the middle 50% of the data occupy.

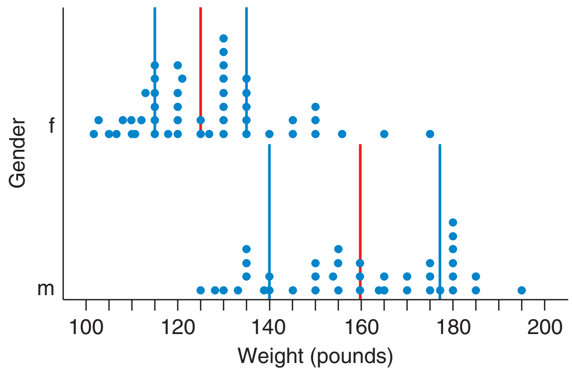
Visualizing the IQR To find the IQR, we cut the distribution into four parts with roughly equal numbers of observations. The distance taken up by the middle two parts is the interquartile range.

The dotplot in Figure 3.18 shows the distribution of weights for a class of introductory statistics students. The vertical lines slice the distribution into four parts so that each part has about 25% of the observations. The IQR is the distance between the first "slice" (at about 121 pounds) and the third slice (at about 160 pounds). This is an interval of 39 pounds ($160 - 121 = 39$).



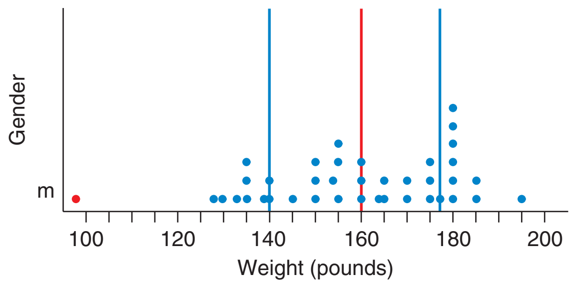
◀ **FIGURE 3.18** The distribution of weights (in pounds) of students in a class is divided into four sections so that each section has roughly 25% of the observations. The IQR is the distance between the outer vertical lines (at 121 pounds and 160 pounds).

Figure 3.19 shows distributions for the same students, but this time the weights are separated by gender. The vertical lines are located so that about 25% of the data are below the leftmost line, and about 25% are above the rightmost line. This means that about half of the data lie between these two boundaries. The distance between these boundaries is the IQR. You can see that the IQR for the males, about 38 pounds, is much larger than the IQR for the females, which is about 20 pounds. The females have less variability in their weights.



◀ **FIGURE 3.19** The dotplot of Figure 3.18 with weights separated by gender (the women on top). The men have a much larger interquartile range than the women.

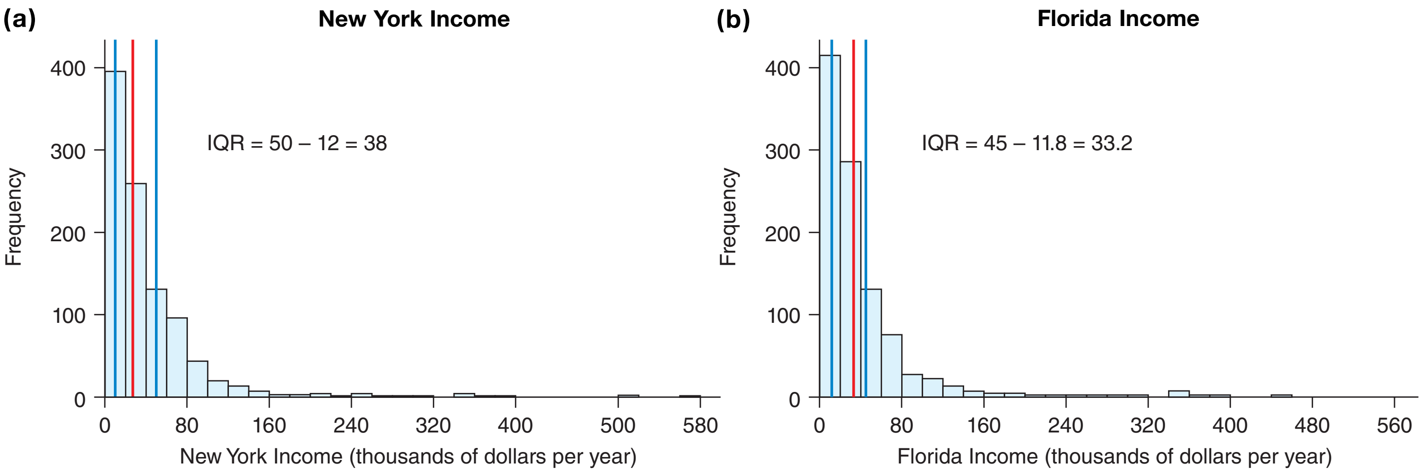
The IQR focuses only on the middle 50% of the data. We can change values outside of this range without affecting the IQR. Figure 3.20 shows the men's weights, but this time we've changed one of the men's weights to be very small. The IQR is still the same as in Figure 3.19.



◀ **FIGURE 3.20** The men's weights are given with a fictitious point (in red) below 100. Moving an extremely large (or small) value does not change the interquartile range.

The Interquartile Range in Context The IQR for the incomes of the New Yorkers in the data set previously shown in Figure 3.15 is $38,000, as shown in Figure 3.21 at the top of the next page. This tells us that the middle 50% of people in our data set had incomes that varied by as much as $38,000. Compare this to the incomes from Florida, which have an IQR of $33,150. There is less variability among the Floridians; Floridians are more similar (at least in terms of their incomes).

An IQR of $38,000 for New Yorkers seems like a pretty large spread. However, considered in the context of the entire distribution (see Figure 3.21), which includes



▲ **FIGURE 3.21** **(a)** The distribution of incomes for New Yorkers. **(b)** The distribution of income for Floridians. In both figures, vertical bars are drawn to divide the distribution into four areas, each with about 25% of the observations. The IQR is the distance between the outer vertical lines, and it is wider for the New York incomes. The red line indicates the median income.

incomes near \$0 and as large as nearly \$600,000, the IQR looks fairly small. The reason is that lots of people (half of our data set) have incomes in this fairly narrow interval.

> **Details**
>
> **Software and Quartiles**
> Different software packages don't always agree with each other on the values for Q1 and Q3, and therefore they might report different values for the IQR. The reason is that several different accepted methods exist for calculating Q1 and Q3. So don't be surprised if the software on your computer gives different values from your calculator.

Calculating the Interquartile Range Calculating the interquartile range involves two steps. First, you must determine where to "cut" the distribution. These points are called the **quartiles**, because they cut the distribution into quarters (fourths). The **first quartile (Q1)** has roughly one-fourth, or 25%, of the observations at or below it. The **second quartile (Q2)** has about 50% at or below it; actually, Q2 is just another name for the median. The **third quartile (Q3)** has about 75% of the observations at or below it. The second step is the easiest: To find the interquartile range you simply find the interval between Q3 and Q1—that is, Q3 − Q1.

To find the quartiles:

- First find the median, which is also called the second quartile, Q2. The median cuts the data into two regions.
- The first quartile (Q1) is the median of the lower half of the sorted data. (Do not include the median observation in the lower half if you started with an odd number of observations.)
- The third quartile (Q3) is the median of the upper half of the sorted data. (Again, do not include the median itself if your full set of data has an odd number of observations.)

Formula 3.5: $\text{Interquartile Range} = Q3 - Q1$

EXAMPLE 13 Heights of Children

A group of eight children have the following heights (in inches):

48.0, 48.0, 53.0, 53.5, 54.0, 60.0, 62.0, and 71.0

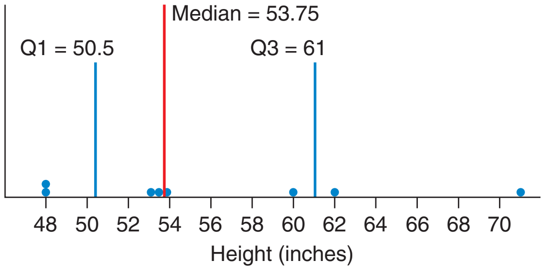
They are shown in Figure 3.22.

QUESTION Find the interquartile range for the distribution of the children's heights.

SOLUTION As before, we first explain how to do the calculations by hand and then show the output of technology.

(a)

(b)



▲ **FIGURE 3.22** **(a)** Eight children sorted by height. **(b)** A dotplot of the heights of the children. The median and quartiles are marked with vertical lines. Note that two dots, or 25% of the data, appear in each of the four regions.

First we find Q2 (the median). Note that the data are sorted and that there are four observed values below the median and four observed values above the median.

Duncan	Charlie	Grant	Aidan		Sophia	Seamus	Cathy	Drew
48	48	53	53.5		54	60	62	71
				Q2				
				53.75				

To find Q1, examine the numbers below the median and find the median of them, as shown.

Duncan	Charlie		Grant	Aidan
48	48		53	53.5
		Q1		
		50.50		

To find Q3, examine the numbers above the median and find the median of them.

Sophia	Seamus		Cathy	Drew
54	60		62	71
		Q3		
		61.00		

Together, these values are

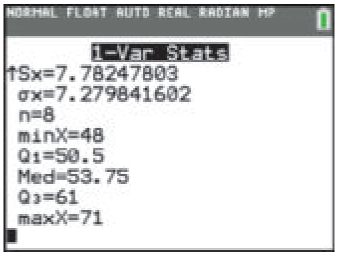
Duncan	Charlie		Grant	Aidan		Sophia	Seamus		Cathy	Drew
48	48		53	53.5		54	60		62	71
		Q1			Q2			Q3		
		50.50			53.75			61.00		

Here’s how we calculated the values:

$$Q1 = \frac{48 + 53}{2} = \frac{101}{2} = 50.50 \qquad \text{(Halfway between Charlie and Grant)}$$

$$Q2 = \frac{53.5 + 54}{2} = \frac{107.5}{2} = 53.75 \qquad \text{(Halfway between Aidan and Sophia)}$$

$$Q3 = \frac{60 + 62}{2} = \frac{122}{2} = 61.00 \qquad \text{(Halfway between Seamus and Cathy)}$$



▲ **FIGURE 3.23** Some output of a TI-84 for eight children's heights.

Tech

The last step is to subtract:

$$\text{IQR} = \text{Q3} - \text{Q1} = 61.00 - 50.50 = 10.50$$

Figure 3.22b shows the location of Q1, Q2, and Q3. Note that 25% of the data (two observations) lie in each of the four regions created by the vertical lines.

Figure 3.23 shows the TI-84 output. The TI-84 does not calculate the IQR directly; you must subtract Q3 – Q1 yourself. The IQR is $61 - 50.5 = 10.5$, which is the same as the IQR done by hand above.

CONCLUSION The interquartile range of the heights of the eight children is 10.5 inches.

TRY THIS! Exercise 3.45b

Finding the Range, Another Measure of Variability

Another measure of variability is similar to the IQR but much simpler. The **range** is the distance spanned by the entire data set. It is very simple to calculate: It is the largest value minus the smallest value.

Formula 3.6: Range = maximum − minimum

For the heights of the eight children (Example 13), the range is $71.0 - 48.0 = 23.0$ inches.

The range is useful for a quick measurement of variability because it's very easy to calculate. However, because it depends on only two observations—the largest and the smallest—it is very sensitive to any peculiarities in the data. For example, if someone makes a mistake when entering the data and enters 710 inches instead of 71 inches, then the range will be very wrong. The IQR, on the other hand, depends on many observations and is therefore more reliable.

SNAPSHOT THE INTERQUARTILE RANGE

WHAT IS IT? ▶ A numerical summary.

WHAT DOES IT DO? ▶ It measures the spread of the distribution of a data set.

HOW DOES IT DO IT? ▶ It computes the distance taken up by the middle half of the sorted data.

HOW IS IT USED? ▶ To measure the variability in a sample, particularly when the distribution is skewed.

SECTION 3.4

Comparing Measures of Center

Which should you choose, the mean (accompanied by the standard deviation) or the median (with the IQR)? These pairs of measures have different properties, so you need to choose the pair that's best for the data you're considering. Our primary goal is to choose a value that is a good representative of the typical values of the distribution.

Look at the Shape First

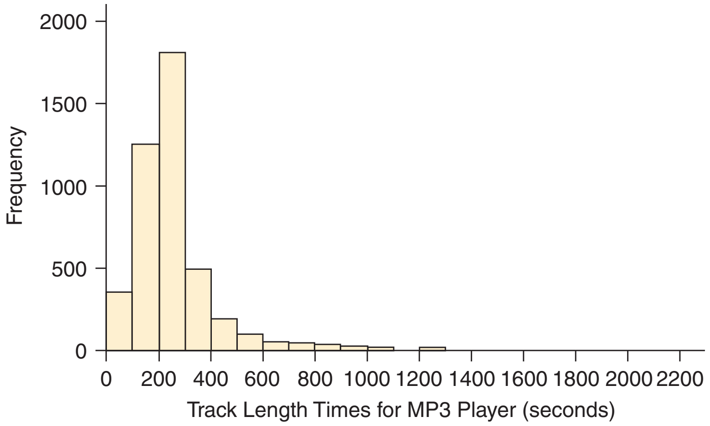
This decision begins with a picture. The shape of the distribution will determine which measures are best for communicating the typical value and the variability in the distribution.

EXAMPLE 14 MP3 Song Lengths

One of the authors created a data set of the songs on his mp3 player. He wants to describe the distribution of song lengths.

QUESTION What measures should he use for the center and spread: the mean (250.2 seconds) with the standard deviation (152.0 seconds) or the median (226 seconds) with the interquartile range (117 seconds)? Interpret the appropriate measures. Refer to the histogram in Figure 3.24.

SOLUTION Before looking at the histogram, you should think about what shape you expect the graph to be. No song can be shorter than 0 seconds. Most songs on the radio are around 4 minutes (240 seconds), with some a little longer and some a little shorter. However, a few songs, particularly classical tracks, are quite long, so we might expect the distribution to be right-skewed. This suggests that the median and IQR are the best measures to use.



◀ **FIGURE 3.24** The distribution of lengths of songs (in seconds) on the mp3 player of one of the authors.

Figure 3.24 confirms that, as we predicted, the distribution is right-skewed, so the median and interquartile range would be the best measures to use.

CONCLUSION The median length is 226 seconds (roughly 3 minutes and 46 seconds), and the interquartile range is 117 seconds (close to 2 minutes). In other words, the typical track on the author's mp3 player is about 4 minutes, but there's quite a bit of variability, with the middle 50% of the tracks differing by about 2 minutes.

TRY THIS! Exercise 3.53

Sometimes, you don't have the data themselves, so you can't make a picture. If so, then you must deduce a shape for the distribution that is reasonable and choose the measure of center on the basis of this deduction.

When a distribution is right-skewed, as it is with the mp3 song lengths, the mean is generally larger than the median. You can see this in Figure 3.24; the right tail means the balancing point must be to the right of the median. With the same reasoning, we can see that in a left-skewed distribution, the mean is generally less than the median. In a symmetric distribution, the mean and median are approximately the same.

KEY POINT In a symmetric distribution, the mean and median are approximately the same. In a right-skewed distribution, the mean tends to be greater than the median, and in a left-skewed distribution, the mean tends to be less than the median.

The Effect of Outliers

Looking Back

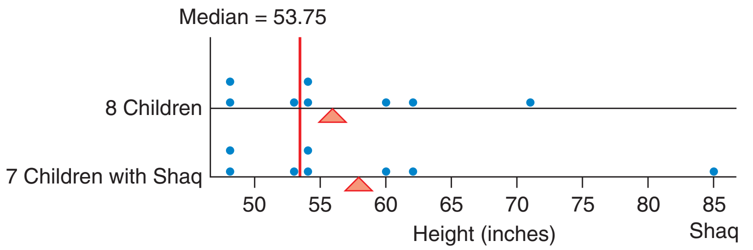
Outliers
Recall from Chapter 2 that an outlier is an extremely large or small observation relative to the bulk of the data.

Even when a distribution is mostly symmetric, the presence of one or more outliers can have a large effect on the mean. Usually, the median is a more representative measure of center when an outlier is present.

The average height of the eight children from Example 13 was 56.2 inches. Imagine that we replace the tallest child (who is 71 inches tall) with basketball player Shaquille O'Neal, whose height is 85 inches. Our altered data set is now

48, 48, 53, 53.5, 54, 60, 62, and 85 inches

In order to keep the balance of the data, the mean has to shift higher. The mean of this new data set is 57.9 inches—over 1 inch higher. The median of the new data set, however, is the same: 53.75 inches, as shown in Figure 3.25.



▶ **FIGURE 3.25** The effect of changing the tallest child's height into Shaquille O'Neal's height. Note that the mean (shown with triangles) changes, but the median (the vertical line) stays the same.

When outliers are present, the median is a good choice for a measure of center. In technical terms, we say that the median is **resistant to outliers**; it is not affected by the size of an outlier and does not change even if a particular outlier is replaced by an even more extreme value.

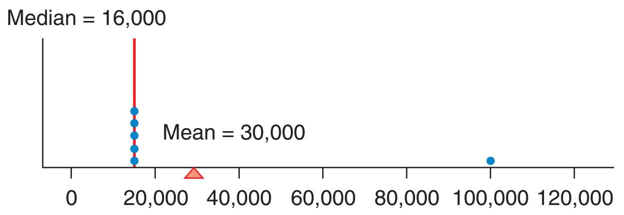
KEY POINT The median is resistant to outliers; it is not affected by the size of the outliers. It is therefore a good choice for a measure of the center if the data contain outliers and you want to reduce their effect.

EXAMPLE 15 Fast Food

A (very small) fast-food restaurant has five employees, all of whom work full-time for $7 per hour. Each employee's annual income is about $16,000 per year. The owner, on the other hand, makes $100,000 per year.

QUESTION Find both the mean and the median. Which would you use to represent the typical income at this business—the mean or the median?

SOLUTION Figure 3.26 shows a dotplot of the data. The mean income is $30,000, and the median is $16,000. Nearly all of the employees earned less than the mean amount!



▶ **FIGURE 3.26** Dotplot of salaries for five employees and their boss at a fast-food company.

CONCLUSION Given the choice between mean and median, the median is better at showing the typical income.

Why are the mean and median so different? Because the owner's salary of $100,000 is an outlier.

TRY THIS! Exercise 3.67

Looking Back

Bimodality
Recall from Chapter 2 that a mode is represented by a major mound in a graph (such as a histogram) of a single numerical variable. A bimodal distribution has two major mounds.

Many Modes: Summarizing Center and Spread

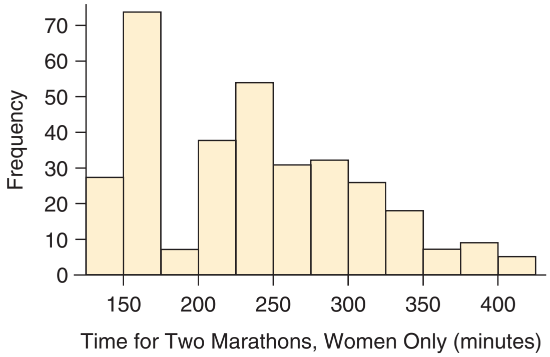
What should you do if the distribution is bimodal or has several modes? Unfortunately, the answer is "It's complicated."

You learned in Chapter 2 that multiple modes in a graphical display of a distribution sometimes indicate that the data set combines different groups. For example, perhaps a data set containing heights includes both men and women. The distribution could very well be bimodal, because we're combining two groups of people who differ markedly in terms of their heights. In this case, and in many other contexts, it is more useful to separate the groups and report summary measures separately for each group. If we know which observations belong to the men and which to the women, then we can separate the data and compute the mean height for men separately from the mean height for women.

For example, Figure 3.27 shows a histogram of the finishing times of female marathon runners. The most noticeable feature of this distribution is that there appear to be two modes. When confronted with this situation, a natural question to ask is "Are two different groups of runners represented in this data set?"

Caution

Don't Get Modes from Computer Output
Avoid using your computer to calculate the modes. For data sets with many numerical observations, many software programs give meaningless values for the modes. For the data shown in Figure 3.27, for example, most software packages would report the location of five different modes, none of them corresponding to the high points on the graph.



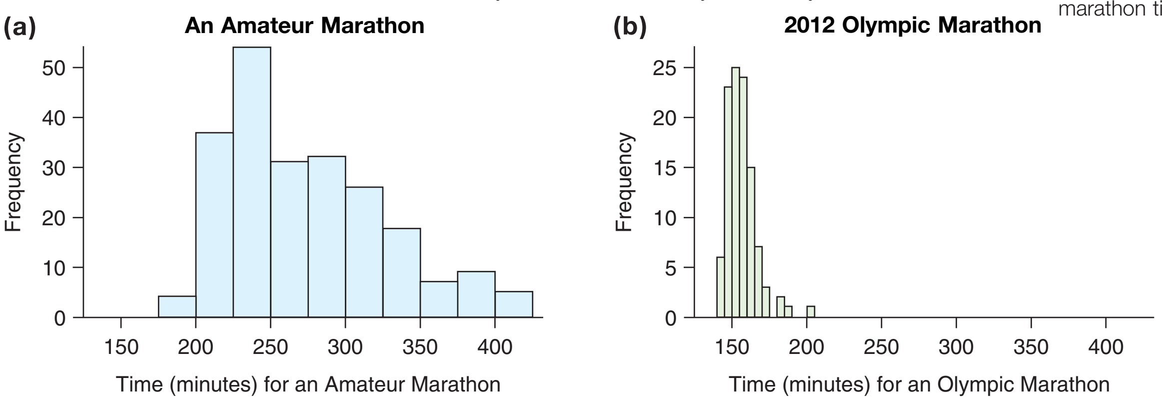
◀ **FIGURE 3.27** Marathon times reported for two groups: amateur and Olympic athletes. Note the two modes. Only women runners were included.

As it turns out, the answer is "yes." Table 3.2 shows the first few lines of the data set. From this, we see that the times belong to runners from two different events. One event was the 2012 Olympics, which includes the best marathoners in the world. The other event was an amateur marathon in Portland, Oregon, in 2013.

Time (minutes)	Event
185.1	Olympics
202.2	Olympics
214.5	Amateur
215.7	Amateur

▲ **TABLE 3.2** Four lines of marathon times for women.

Figure 3.28 shows the data separately for each of these events. We could now compute measures for center and spread separately for the Olympic and amateur events. However, you will sometimes find yourself in situations where a bimodal distribution occurs but it *does* make sense to compute a single measure of center. We can't give you advice for what to do in all situations. Our best advice is always to ask, "Does my summary make sense?"



▲ **FIGURE 3.28** Women's times for a marathon, separated into two groups: **(a)** amateur athletes and **(b)** Olympic athletes.

Caution

Average
A guy comes across a statistician who is standing with one foot in a pot of boiling water and one foot frozen in a block of ice. "Isn't that painful?" the man asks. "Maybe so," says the statistician, "but on average it feels just right." Beware of applying the mean to situations where it will not provide a "typical" measure!

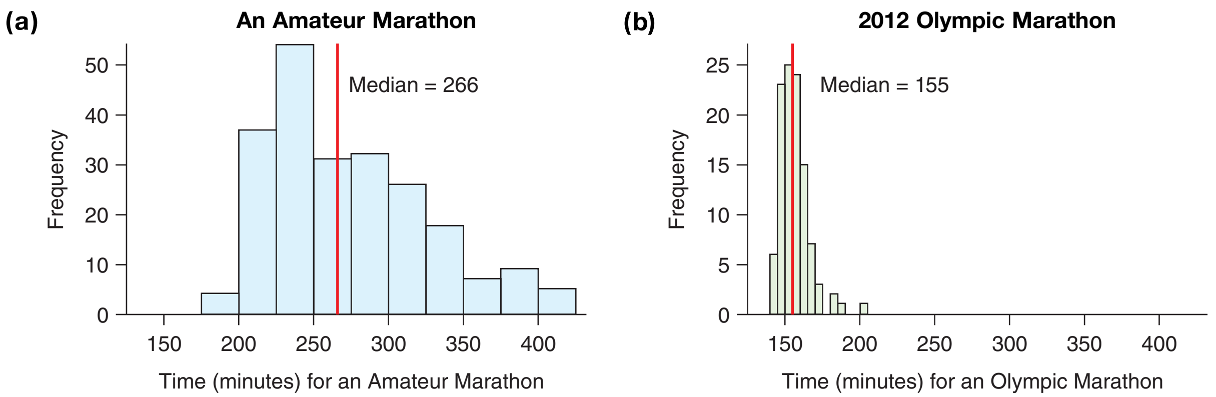
Comparing Two Groups with Different-Shaped Distributions

Note that the times for the amateur runners in Figure 3.28a are fairly symmetric, which suggests we should use the mean to report the typical finishing time. But the Olympic times are right-skewed, which suggests we should compute the median. Which do you think is the better measure for comparing these two groups, the mean or the median?

When comparing two distributions, you should always use the same measures of center and spread for both distributions. Otherwise, the comparison is not valid.

EXAMPLE 16 Marathon Times, Revisited

In Figure 3.27 we lumped all of the marathon runners' finishing times into one group. But in fact, our data set had a variable that told us which time belonged to an Olympic runner and which to an amateur runner, so we could separate the data into groups. Figure 3.28 shows the same data, separated by groups.



▲ **FIGURE 3.29** Women's times for a marathon, separated into two groups: **(a)** amateur athletes and **(b)** Olympic athletes. The median is shown for each group.

QUESTION Typically, which group has the fastest finishing times?

SOLUTION The distribution of Olympic runners is right-skewed, so the median would be the best measure. Although the distribution of amateur runners is relatively symmetric, we report the median because we want to compare the typical running time of the amateurs to the typical running time of the Olympic runners. The median of the women Olympic runners is 154.8 minutes (about 2.6 hours), and the median for the amateur women runners is 240.0 minutes (about 4 hours). Figure 3.29 shows the location of each group's median running time.

CONCLUSION The typical woman Olympic runner finishes the marathon considerably faster: a median time of 154.8 minutes (about 2.6 hours) compared to 240.0 minutes (about 4 hours) for the amateur athlete.

TRY THIS! Exercise 3.69

KEY POINT When you are comparing groups, if any one group is strongly skewed or has outliers, it is usually best to compare the medians and interquartile ranges for all groups.

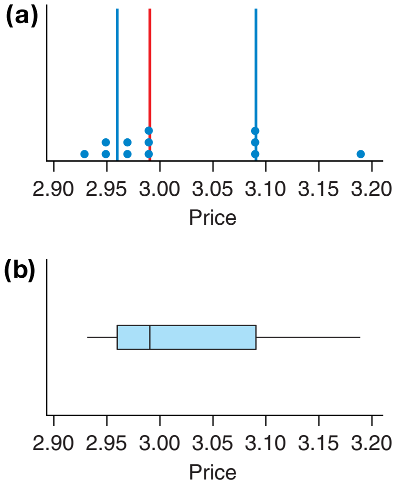
SECTION 3.5

Using Boxplots for Displaying Summaries

Boxplots are a useful graphical tool for visualizing a distribution, especially when comparing two or more groups of data. A boxplot shows us the distribution divided into fourths. The left edge of the box is at the first quartile (Q1) and the right edge is at the third quartile (Q3). Thus the middle 50% of the sorted observations lie inside the box. Therefore, the length of the box is proportional to the IQR.

A vertical line inside the box marks the location of the median. Horizontal lines, called whiskers, extend from the ends of the box to the smallest and largest values, or nearly so. (We'll explain soon.) Thus the entire length of the boxplot spans most, or all, of the range of the data.

Figure 3.30 compares a dotplot (with the quartiles marked with vertical lines) and a boxplot for the price of gas at stations in Austin, Texas, as discussed in Examples 6 and 11.



▲ **FIGURE 3.30** Two views of the same data: **(a)** A dotplot with Q1, Q2, and Q3 indicated and **(b)** a boxplot for the price of regular, unleaded gas at stations in Austin, Texas.

Unlike many of the graphics used to visualize data, boxplots are relatively easy to draw by hand, assuming that you've already found the quartiles. Still, most of the time you will use software or a graphing calculator to draw the boxplot. Most software packages produce a variation of the boxplot that helps identify observations that are extremely large or small compared to the bulk of the data.

These extreme observations are called potential outliers. Potential outliers are different from the outliers we discussed in Chapter 2, because sometimes points that look extreme in a boxplot are not that extreme when shown in a histogram or dotplot. They are called *potential* outliers because you should consult a histogram or dotplot of the distribution before deciding whether the observation is too extreme to fit the pattern of the distribution. (Remember, whether or not an observation is an outlier is a subjective decision.)

Potential outliers are identified by this rule: They are observations that are a distance of more than 1.5 interquartile ranges below the first quartile (the left edge of a horizontal box) or above the third quartile (the right edge).

To allow us to see these potential outliers, the whiskers are drawn from the edge of each box to the most extreme observation that is not a potential outlier. This implies that before we can draw the whiskers, we must identify any potential outliers.

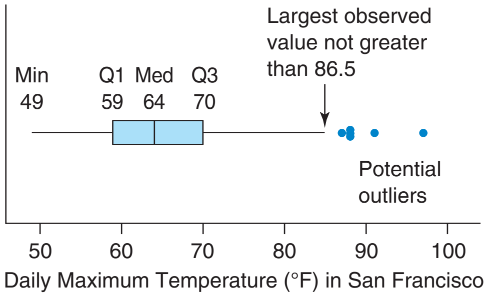
KEY POINT Whiskers in a boxplot extend to the most extreme values that are not potential outliers. Potential outliers are points that are more than 1.5 IQRs from the edges of the box.

Figure 3.31 is a boxplot of temperatures in San Francisco (see Examples 8 and 10). From the boxplot, we can see that

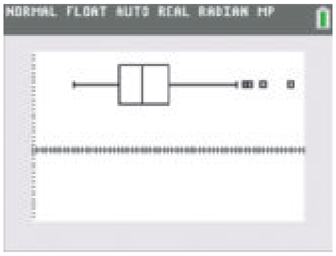
$$\text{IQR} = 70 - 59 = 11$$
$$1.5 \times \text{IQR} = 1.5 \times 11 = 16.5$$
$$\text{Right limit} = 70 + 16.5 = 86.5$$
$$\text{Left limit} = 59 - 16.5 = 42.5$$

Any points below 42.5 or above 86.5 would be potential outliers.

The whiskers go to the most extreme values that are not potential outliers. On the left side of the box, observations smaller than 42.5 would be potential outliers. However, there are no observations that small. The smallest is 49, so the whisker extends to 49.



▶ **FIGURE 3.31** Boxplot of maximum daily temperatures in San Francisco.



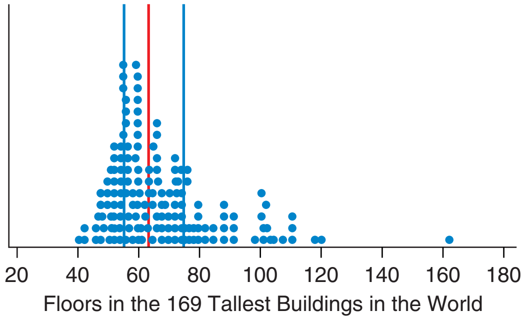
▲ **FIGURE 3.32** TI-84 output for a boxplot of San Francisco temperatures.

Tech

On the right, several values in the data set are larger than 86.5. The whisker extends to the largest temperature that is less than (or equal to) 86.5, and the larger values are shown in Figure 3.31 with dots. These represent days that were unusually warm in San Francisco. Figure 3.32 shows a boxplot made with a TI-84.

EXAMPLE 17 Skyscraping

Figure 3.33 shows the distribution of the 169 tallest buildings in the world as measured by the number of floors in the building. Some summary statistics: Q1 is 55 floors, the median is 63 floors, and Q3 is 74.5 floors. The tallest building is the Burj Khalifa in Dubai, with 162 floors. The shortest is the Kingdom Centre in Saudi Arabia, with 42 floors.



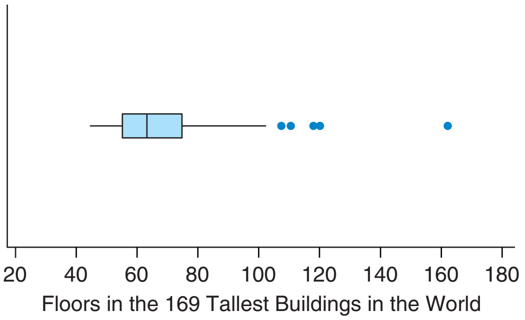
▶ **FIGURE 3.33** Number of floors for the 169 tallest buildings in the world. The vertical lines indicate (from left to right) the first quartile, the median (in red), and the third quartile.

QUESTION Sketch the boxplot. Describe how you determined where to draw the whiskers. Are there outliers? Are the outliers mostly very short buildings or very tall buildings?

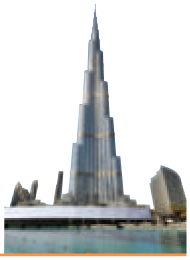
SOLUTION The left edge of the box is at the first quartile, 55 floors, and the right edge is at the third quartile, 74.5 floors. We draw a line inside the box at the median of 63 floors.

Potential outliers on the left side of the box are those more than $1.5 \times$ IQR below Q1. The IQR is $Q3 - Q1 = 74.5 - 55 = 19.5$. Thus $1.5 \times \text{IQR} = 1.5 \times 19.5 = 29.5$. This means that potential outliers on the left must be 29.5 floors less than 55, or $55 - 29.5 = 25.5$, floors or below. The shortest building is 42 floors, so we draw the whisker to the smallest of 42 floors.

On the right, potential outliers are more than 29.5 floors above the third quartile, so any building with more than $74.5 + 29.5 = 104$ floors is a potential outlier. So we draw the right-hand-side whisker to the tallest building that has 104 or fewer floors. From the dotplot this appears to be 104. The remaining buildings we indicate with dots or stars to show that they are potential outliers.



◀ **FIGURE 3.34** Boxplot summarizing the distribution of the number of floors of the world's tallest buildings.



CONCLUSION The boxplot is shown in Figure 3.34.

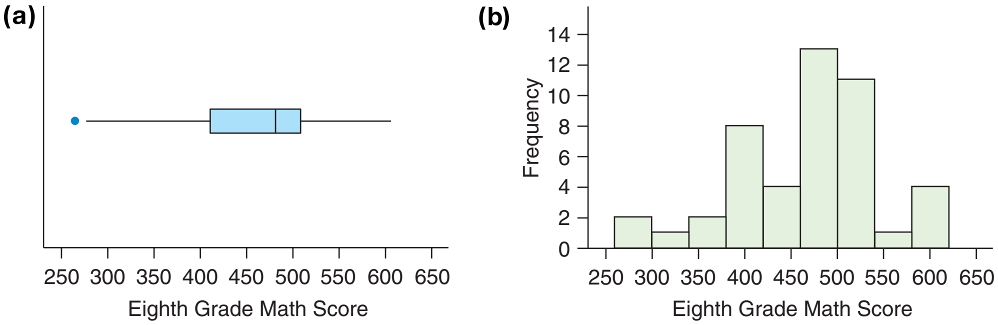
We can see five outliers, and all are tall buildings. None of the outliers are shorter buildings.

TRY THIS! Exercise 3.63

Investigating Potential Outliers

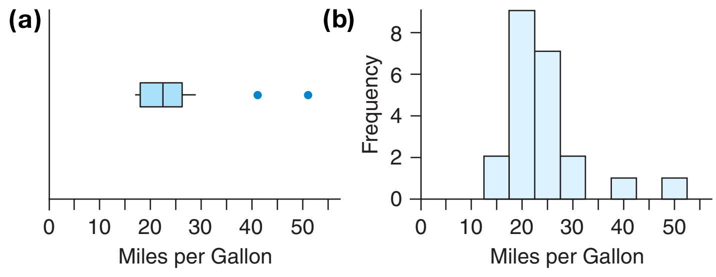
What do you do with potential outliers? The first step is always to investigate. A potential outlier might not be an outlier at all. Or a potential outlier might tell an interesting story, or it might be the result of an error in entering data.

Figure 3.35a is a boxplot of the NAEP international math scores for 42 countries (International math scores 2007). One country (South Africa, as it turns out) is flagged as a potential outlier. However, if we examine a histogram, shown in Figure 3.35b, we see that this outlier is really not that extreme. Most people would not consider South Africa to be an outlier in this distribution, because it is not separated from the bulk of the distribution in the histogram.



◀ **FIGURE 3.35** **(a)** Distribution of international math scores for eighth grade achievement. The boxplot indicates a potential outlier. **(b)** The histogram of the distribution of math scores shows that although South Africa's score of 264 might be the lowest, it is not that much lower than the bulk of the data.

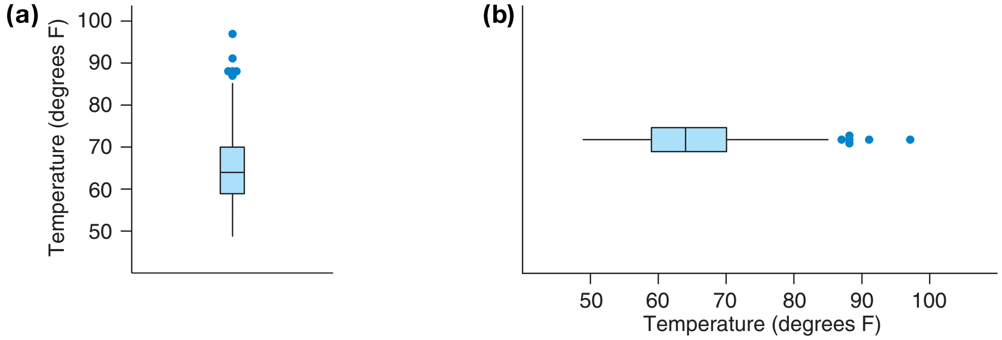
Figure 3.36 shows a boxplot and histogram for the fuel economy (in city driving) of the 2010 model sedans from Ford, Toyota, and GM, in miles per gallon, as listed on their websites. Two potential outliers appear, which are far enough from the bulk of the distribution as shown in the histogram that many people would consider them real outliers. These outliers turn out to be hybrid cars: the Ford Fusion and the Toyota Prius. These hybrids run on both electricity and gasoline, so they deliver much better fuel economy.



▶ **FIGURE 3.36** Distribution of fuel economies for cars from three manufacturers. **(a)** The boxplot identifies two potential outliers. **(b)** The histogram confirms that these points are indeed more extreme than the bulk of the data.

Horizontal or Vertical?

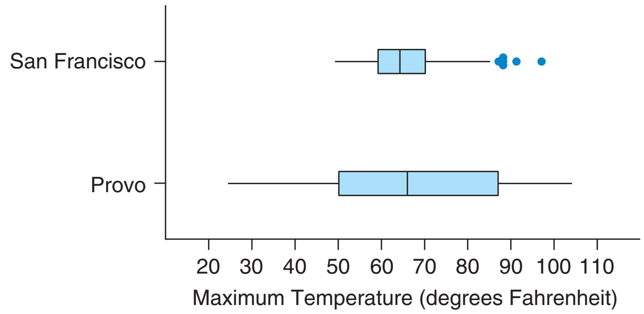
Boxplots do not have to be horizontal. Many software packages (such as Minitab) provide you with the option of making vertical boxplots. (See Figure 3.37a). Which direction you choose is not important. Try both to see which is more readable.



▶ **FIGURE 3.37** **(a)** Default output of Minitab for a boxplot of the San Francisco temperatures. **(b)** Minitab boxplot with the horizontal orientation.

Using Boxplots to Compare Distributions

Boxplots are often a very effective way of comparing two or more distributions. How do temperatures in Provo compare with those in San Francisco? Figure 3.38 shows boxplots of daily maximum temperatures for San Francisco and Provo. At a glance, we can see how these two distributions differ and how they are similar. Both cities have similar typical temperatures (the median temperatures are about the same). Both distributions are fairly symmetric (because the median is in the center of the box, and the boxplots are themselves fairly symmetric). However, the amount of variation in daily temperatures is much greater in Provo than in San Francisco. We can see this easily because the box is wider for Provo's temperatures.

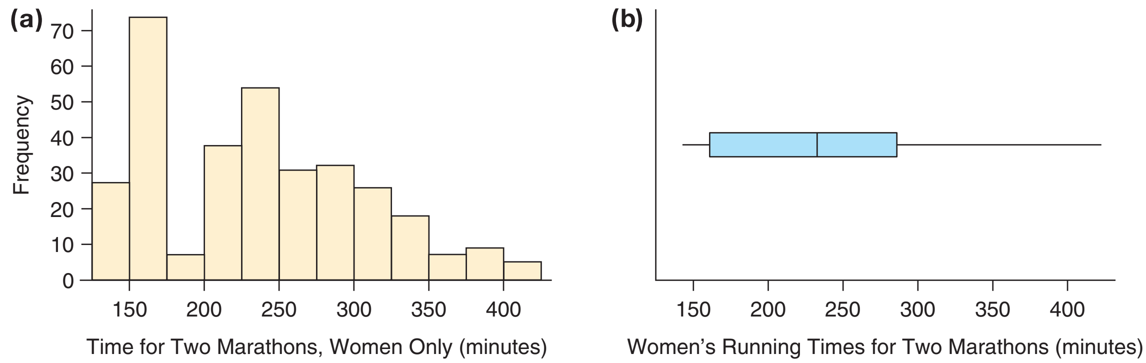


▶ **FIGURE 3.38** Boxplots of daily maximum temperatures in Provo and San Francisco emphasize the difference in variability of temperature in the two cities.

Also note that although both cities do have days that reach about 100 degrees, these days are unusual in San Francisco—they're flagged as potential outliers—but merely fall in the upper 25% for Provo.

Things to Watch for with Boxplots

Boxplots are best used only for unimodal distributions because they hide bimodality (or any multimodality). For example, Figure 3.39a repeats the histogram of marathon running times for two groups of women runners: amateurs and Olympians. The distribution is clearly bimodal. However, the boxplot in part (b) doesn't show us the bimodality. Boxplots can give the misleading impression that a bimodal distribution is really unimodal.



▲ **FIGURE 3.39** **(a)** Histogram of finishing times (in seconds) for two groups of women marathon runners: Olympic athletes and amateurs. The graph is bimodal because the elite athletes tend to run faster, and therefore there is a mode around 160 minutes and another mode around 240 minutes. **(b)** The boxplot hides this interesting feature.

Boxplots should *not* be used for very small data sets. It takes at least five numbers to make a boxplot, so if your data set has fewer than five observations, you can't make a boxplot.

Finding the Five-Number Summary

Boxplots are not really pictures of a distribution in the way that histograms are. Instead, boxplots help us visualize the location of various summary statistics. The boxplot is a visualization of a numerical summary called the **five-number summary**. These five numbers are

the minimum, Q1, the median, Q3, and the maximum

For example, for daily maximum temperatures in San Francisco, the five-number summary is

49, 59, 64, 70, 97

as you can see in Figure 3.31.

Note that a boxplot is not just a picture of the five-number summary. Boxplots always show the maximum and minimum values, but sometimes they also show us potential outliers.

SNAPSHOT **THE BOXPLOT**

WHAT IS IT? ► A graphical summary.

WHAT DOES IT DO? ► Provides a visual display of numerical summaries of a distribution of numerical data.

HOW DOES IT DO IT? ► The box stretches from the first quartile to the third quartile, and a vertical line indicates the median. Whiskers extend to the largest and smallest values that are not potential outliers, and potential outliers are indicated with special marks.

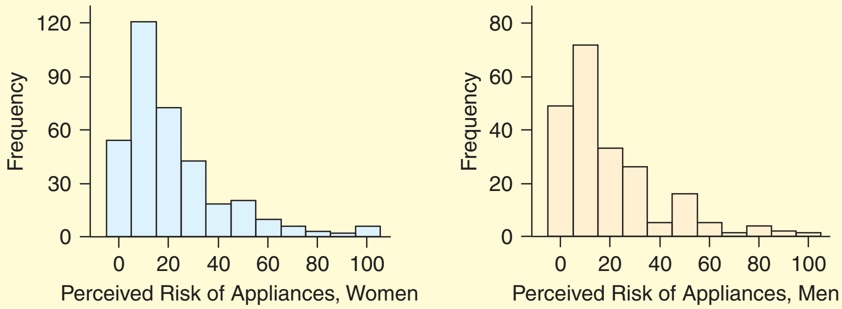
HOW IS IT USED? ► Boxplots are useful for comparing distributions of different groups of data.

CASE STUDY REVISITED

Living in a Risky World

How do the men and women of this study compare when it comes to assigning risk to using a household appliance and to getting an annual X-ray at the doctor's? In Chapter 2 we compared groups graphically, and this is still the first step. But in this chapter, we learned about methods for comparing groups numerically, and this will enable us to be more precise in our comparison.

Our first step is to examine the pictures of the distributions to decide which measures would be most appropriate. (We repeat Figure 3.1.)



▶ **FIGURE 3.1A (repeated)** Histograms showing the distributions of perceived risk of using appliances. The women's data are shown on the left, and the men's data are shown on the right.

	Risk of Appliances	
	Median	**IQR**
Men	10	20
Women	15	25
	Risk of X-rays	
	Mean	**SD**
Men	46.8	20.0
Women	47.8	20.8

▲ **TABLE 3.3** Comparison of perceived risks for men and women.

Risk of Appliances

The histograms for the perceived risk of using appliances (Figure 3.1a) do not show large differences between men and women. We can see that both distributions are right-skewed, and both appear to have roughly the same typical value, although the women's typical value might be slightly higher than the men's. Because the distribution is right-skewed, we compute the medians and IQRs to compare the two groups. (Table 3.3 summarizes these comparisons.) The men assigned a median risk of 10 to using household appliances, and the women assigned a median risk of 15. We see that, first impressions aside, these women tend to feel that using appliances is a riskier activity than do these men. Also, more differences in opinion occurred among these women than among these men. The IQR was 25 for women and 20 for men. Thus the middle 50% of the women varied by as much as 25 points in how risky they saw this activity; there was less variability for the men.

Risk of X-rays

Both distributions for the perceived risk level of X-rays (Figure 3.1b) were fairly symmetric, so it makes sense to compare the two groups using the mean and standard deviation. The mean risk level for men was 46.8 and for women was 47.8. Typically, men and women feel roughly the same about the risk associated with X-rays. The standard deviations are about the same, too: men have a standard deviation of 20.0 and women of 20.8. From the Empirical Rule, we know that a majority (about two-thirds) of men in this sample rated the risk level between 26.8 and 66.8. The majority of women rated it between 27.0 and 68.6.

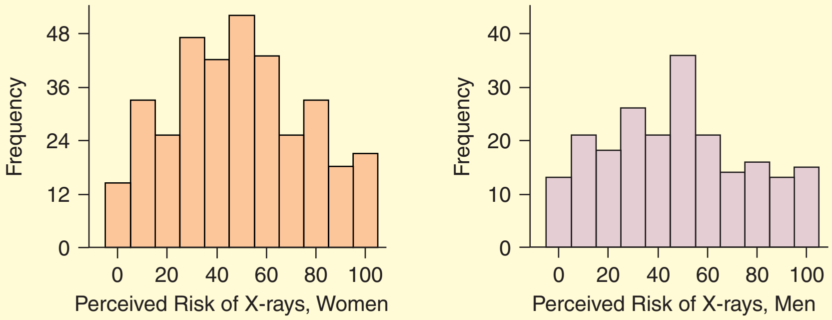
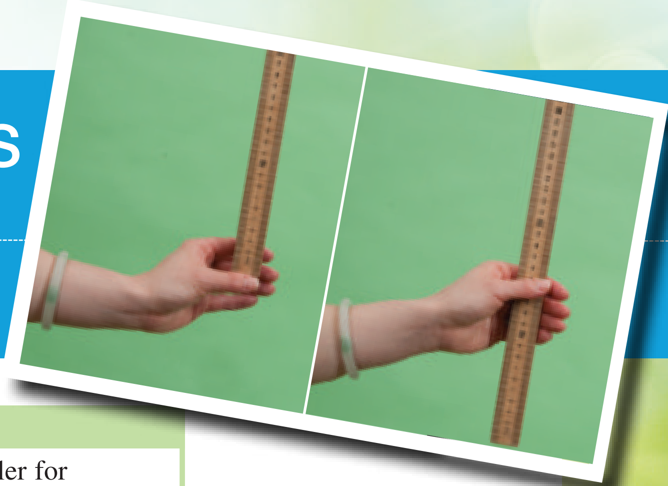


FIGURE 3.1B (repeated) Histograms showing the distributions of perceived risk of X-rays. The women's data are shown on the left, and the men's data are shown on the right.

The comparisons are summarized in Table 3.3.

EXPLORING STATISTICS
CLASS ACTIVITY

Does Reaction Distance Depend on Gender?



GOALS	MATERIALS
Apply the concepts introduced in this chapter to data you have collected in order to compare two groups.	A meter stick or stiff ruler for each group.

ACTIVITY

Work in groups of two or three. One person holds the meter stick vertically, with one hand near the top of the stick, so that the 0-centimeter mark is at the bottom. The other person then positions his or her thumb and index finger about 5 cm apart (2 inches apart) on opposite sides of the meter stick at the bottom. Now the first person drops the meter stick without warning, and the other person catches it. Record the location of the middle of the thumb of the catcher. This is the distance the stick traveled and is called the reaction distance, which is related to reaction time. A student who records a small distance has a fast reaction time, and a student with a larger distance has a slower reaction time. Now switch tasks. Each person should try catching the meter stick twice, and the better (shorter) distance should be reported for each person. Then record the gender of each catcher. Your instructor will collect your data and combine the class results.

BEFORE THE ACTIVITY

1. Imagine that your class has collected data and you have 25 men and 25 women. Sketch the shape of the distribution you expect to see for the men and the distribution you expect to see for the women. Explain why you chose the shape you did.
2. What do you think would be a reasonable value for the typical reaction distance for the women? Do you think it will be different from the typical reaction distance for the men?

AFTER THE ACTIVITY

1. Now that you have actual data, how do the shapes of the distributions for men and women compare to the sketches you made before you collected data?
2. What measures of center and spread are appropriate for comparing men's and women's reaction distances? Why?
3. How do the actual typical reaction distances compare to the values you predicted?
4. Using the data collected from the class, write a short paragraph (a couple of sentences) comparing the reaction distances of men and women. You should also talk about what group you could extend your findings to, and why. For example, do your findings apply to all men and women? Or do they apply only to college students?

CHAPTER REVIEW

KEY TERMS

mean, *108*
average, *110*
standard deviation, *114*
deviation, *116*
variance, *117*
Empirical Rule, *118*
standard units, *121*
z-score, *121*
median, *123*
interquartile range (IQR), *126*
first quartile (Q1), *128*
second quartile (Q2), *128*
third quartile (Q3), *128*
resistant to outliers, *132*
boxplot, *135*
potential outlier, *135*
five-number summary, *139*

LEARNING OBJECTIVES

After reading this chapter and doing the assigned homework problems, you should

- Understand how measures of center and spread are used to describe characteristics of real-life samples of data.
- Understand when it is appropriate to use the mean and standard deviation and when it is better to use the median and interquartile range.
- Understand the mean as the balancing point of the distribution of a sample of data and the median as the point that has roughly 50% of the distribution below it.
- Be able to write comparisons between samples of data in context.

SUMMARY

The first step in any statistical investigation is to make a picture of the distribution of a numerical variable using a dotplot or histogram. Before computing any summary statistics, you must examine a graph of the distribution to determine the shape and whether or not there are outliers. As noted in Chapter 2, you should report the shape, center, and variability of every distribution.

If the shape of the distribution is symmetric and there are no outliers, you can describe the center and spread by using either the median with the interquartile range or the mean with the standard deviation, although it is customary to use the mean and standard deviation.

If the shape is skewed or there are outliers, you should use the median with the interquartile range.

If the distribution is multimodal, try to determine whether the data consist of separate groups; if so, you should analyze these groups separately. Otherwise, see whether you can justify using a single measure of center and spread.

If you are comparing two distributions and one of the distributions is skewed or has outliers, then it is usually best to compare the median and interquartile ranges for both groups.

The choices are summarized in Table 3.4.

Converting observations to z-scores changes the units to standard units, and this enables us to compare individual observations from different groups.

Formulas

Formula 3.1: $$\text{Mean} = \bar{x} = \frac{\sum x}{n}$$

The mean is the measure of center best used if the distribution is symmetric.

Formula 3.2: $$\text{Standard deviation} = s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

The standard deviation is the measure of variability best used if the distribution is symmetric.

Shape	Summaries of Center and Variability
If distribution is bimodal or multimodal	Try to separate groups, but if you cannot, decide whether you can justify using a single measure of center. If a single measure will not work, then report the approximate locations of the modes.
If any group's distribution is strongly skewed or has outliers	Use medians and interquartile ranges for all groups.
If all groups' distributions are roughly symmetric	Use means and standard deviations for all groups.

▲ **TABLE 3.4** Preferred measures to report when summarizing data or comparing two or more groups.

Formula 3.3: $$\text{Variance} = s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

The variance is another measure of variability used if the distribution is symmetric.

Formula 3.4b: $$z = \frac{x - \bar{x}}{s}$$

This formula converts an observation to standard units.

Formula 3.5: Interquartile range $=$ Q3 $-$ Q1

The interquartile range is the measure of variability best used if the distribution is skewed.

Formula 3.6: Range $=$ maximum $-$ minimum

The range is a crude measure of variability.

SOURCES

Carlstrom, L., J. Woodward, P. Arthur, and G. Christina. 2000. Evaluating the simplified conjoint expected risk model: Comparing the use of objective and subjective information. *Risk Analysis* 20(3).

Baseball salaries. 2010. http://www.bizofbaseball.com/index.php?option=com_wrapper&view=wrapper&Itemid=17 (accessed June 13, 2010).

EPA. Particulate matter. http://www.epa.gov/airtrends/factbook.html

Gasoline mileage. 2010. http://www.gm.com and www.fordvehicles.com (accessed June 2010).

International math scores. 2007. http://www.edweek.org/ew/articles/2007/05/02/35air.h26.html

Provo temperatures. 2007. http://www.pgjr.alpine.k12.ut.us/science/james/provo2006.html (accessed August 2007).

San Francisco temperatures. 2007. http://169.237.140.1/calludt.cgi/WXDATAREPORT (accessed August 2007).

SECTION EXERCISES

SECTION 3.1

3.1 Eating A dietician says, "Typically, boys eat more than girls eat daily." What does this statement mean? (Pick the best choice.)

a. The center of the distribution of diet for boys is greater than the center for girls.

b. All boys eat more than all girls.

c. All girls' diets are less varied than all boys' diets.

d. Between boys and girls, boys eat more daily.

3.2 Offices An expert claims that all things being equal, offices without a reception area tend to sell for less than those with a reception area. What does this statement mean? (Pick the best choice.)

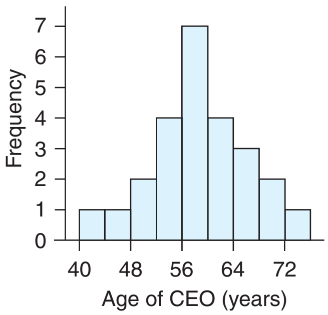
a. There are more offices with a reception area than without.

b. There is more variability in the price of offices without a reception area than in the price of those with it.

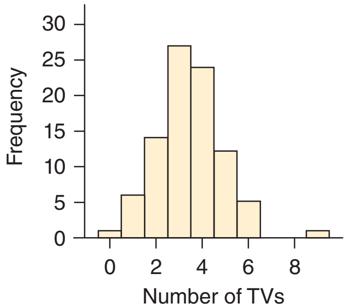
c. The less expensive offices sold do not have a reception areas.

d. The typical price for offices without a reception area is less than the typical price for offices with a reception area.

TRY **3.3 Age of CEOs (Example 1)** The histogram shows the ages of 25 CEOs listed at Forbes.com. Based on the distribution, what is the approximate mean age of the CEOs in this data set? Write a sentence in context (using words in the question) interpreting the estimated mean. The typical CEO is about ________ years old.



3.4 Televisions The histogram shows the number of televisions in the homes of 90 community college students. Judging from the histogram, what is the approximate mean number of televisions in the homes in this collection? Explain.



3.5 Literacy According to indexmundi.com, the rates of literacy in five countries in the Middle East with the lowest rates of literacy are given in the table.

Afghanistan	28
Pakistan	55
Yemen	65
Iraq	79
Syria	84

a. Find and report the mean rate of literacy per country in context: The mean percentage of literates in these five countries is ________. (Report the percentage to the nearest tenth.)

b. Sketch a dotplot of the data and mark the location of the mean. You can use a triangle under the axis (▲, like a fulcrum) to mark the location; it should be at the balance point.

c. Find and report the standard deviation of the rate of literates per country in context; round to the nearest hundredth.

d. Which of these rates is farthest from the mean and therefore contributes most to the standard deviation?

3.6 Literacy According to indexmundi.com, the rates of literacy in five countries in Oceania with the lowest rates of literacy are given in the table.

Papua New Guinea	62
Vanuatu	83
Solomon Islands	84
Palau	92
Fiji	94

a. Find and report the mean rate of literacy per country in context; report the percentage to the nearest tenth.

b. Sketch a dotplot of the data and mark the location of the mean. You can use a triangle under the axis (▲, like a fulcrum) to mark the location; it should be at the balance point.

c. Find and report the standard deviation of the rate of literates in the context; round to the nearest hundredth.

d. Which of these rates is farthest from the mean and therefore contributes most to the standard deviation?

TRY **3.7 Paid Vacation Days (Example 2)** This list represents the numbers of paid vacation days required by law for different countries. (Source: *2009 World Almanac and Book of Facts*)

United States	0
Australia	20
Italy	20
France	30
Germany	24
Canada	10

a. Find the mean, rounding to the nearest tenth of a day. Interpret the mean in this context. Report the mean in a sentence that includes words such as "paid vacation days."

b. Find the standard deviation, rounding to the nearest tenth of a day. Interpret the standard deviation in context.

c. Which number of days is farthest from the mean and therefore contributes most to the standard deviation?

3.8 Age of Prime Ministers This list represents the ages of the first six prime ministers of India when they first assumed office. (Source: www.pmindia.gov.in)

Pandit Jawahar Lal Nehru	58
Gulzari Lal Nanda	66
Lal Bahadur Shastri	63
Indira Gandhi	49
Morarji Desai	81
Charan Singh	72

a. Find the mean age, rounding to the nearest tenth. Interpret the mean in this context.

b. According to a survey, people in the 20th century had an average age of 80 years. How does the mean age of these prime ministers compare to that?

c. Which of the prime ministers listed here had an age that is farthest from the mean and therefore contributes most to the standard deviation?

d. Find the standard deviation, rounding to the nearest tenth.

3.9 Best-Performing Cars In 2013, 66 international motoring journalists selected the 10 best-performing cars of the previous 12 months. The brake horsepowers of the top five cars were 730, 565, 571, 197 and 320. (Source: *The Telegraph*)

a. Find the mean power, rounding to the nearest tenth. The mean of the horsepower of the next five best-performing cars was 402.4. Did the first five cars tend to have more horsepower or less horsepower than the next five cars?

b. Find the standard deviation of horsepower, rounding to the nearest tenth. The standard deviation of the next five cars was 162.07. Did the first five cars tend to have more or less variation than the other five cars?

3.10 Military Coup The tenure of military coups in Argentina has been 2, 3, 3, 1, 7 and 7 years. (Source: Wikipedia)

a. Find and interpret (report in context) the mean time of the coups, rounding to the nearest tenth. The mean time of the military coups in another South American country was 1.3 years. Did the military coups in Argentina tend to be longer than those in the other South American country?

b. Find the standard deviation of the time, rounding to the nearest tenth. The standard deviation of the military coups in the other South American country was 1.5 years. Did the military rule in Argentina tend to have more or less variation than that in another country?

TRY **3.11 Weight Loss (Example 3)** The table shows Minitab descriptive statistics for the weight of some women (weight_f) and men (weight_m), and the self-reported ideal weights for both.

a. Subtract the women's mean weight from their mean ideal weight to find the mean desired weight change. Did the women (as a group) tend to want to lose or to gain weight? How do you know?

b. Subtract the men's mean weight from their mean ideal weight to find the mean desired weight change. Did the men (as a group) tend to want to lose or to gain weight? How do you know?

c. On average, which group wanted the greatest weight change? Compare the mean desired weight loss for women and men.

d. Which group's real weights had more variation as shown by the standard deviations (in the column headed StDev)?

Minitab Statistics: weight_f, weight_m, ideal wt_f, ideal wt_m

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Max
weight_f	26	140.88	22.29	99.0	129.25	133.0	153.75	185
weight_m	13	187.6	38.9	120.0	152.5	200.0	210.0	259
ideal wt_f	26	120.04	11.77	90.0	110.00	120.0	130.00	150
ideal wt_m	13	169.62	23.85	130.0	157.50	170.0	187.50	220

3.12 Education of Fathers and Mothers The table shows Minitab descriptive statistics for the years of education for the fathers and mothers of students in one of the author's statistics classes and in one of her pre-algebra classes. Twelve years is equivalent to a high school education.

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
StatFatherEd	29	8.034	4.625	0.000	3.500	8.000	12.000	16.000
StatMotherEd	29	8.241	4.626	0.000	4.500	8.000	12.000	18.000
PreAlgFatherEd	12	9.00	4.05	2.00	5.00	10.50	12.00	14.00
PreAlgMotherEd	12	10.833	2.691	5.000	8.500	11.500	12.750	14.000

a. Are the means higher for those in pre-algebra or those in statistics?

b. In pre-algebra, is the mean higher for the mothers or the fathers?

c. Which of the four groups has the smallest standard deviation (StDev)?

3.13 Attendance The teachers of a school collected data on the self-reported numbers of sick leaves taken by students in the months of September and October.

September: 8, 3, 2, 1, 6, 4, 1, 8, 2, 2, 2, 5, 7, 1, 8, 9, 4, 3, 2, 6, 8, 1, 3, 9, 8, 1, 1, 12, 9, 10

October: 2, 8, 3, 6, 1, 5, 4, 9, 8, 1, 1, 12, 9, 9, 0, 6, 1, 8,
4, 2, 1, 9, 12, 8, 1, 3, 2, 4, 7, 1

a. Compare the means in a sentence or two.

b. Compare the standard deviation in a sentence or two.

3.14 Working Overtime A statistician asked a group of 60 employees how many hours they have worked overtime during the last two weeks. Half of the employees are white-collar workers and half are blue-collar workers

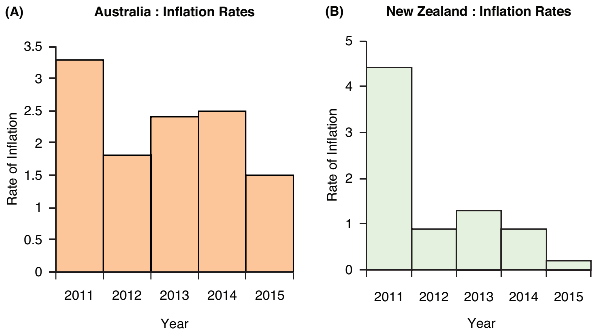
White-collar employees: 20, 10, 13, 10, 11, 12, 10, 15, 8, 5, 7, 9, 8,
6, 4, 20, 15, 11, 14, 12, 11, 12, 4, 9, 8, 4, 7, 8, 4, 6

Blue-collar employees: 22, 22, 19, 11, 19, 11, 13, 12, 10, 12, 7, 9, 8,
4, 9, 20, 20, 16, 14, 19, 10, 15, 14, 12, 10, 6, 7, 8, 8, 4

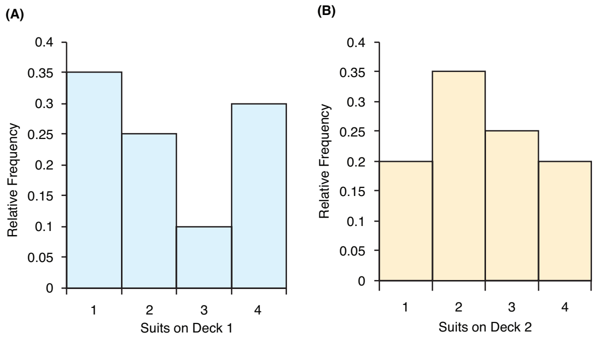
a. Compare the mean in a sentence or two.

b. Compare the standard deviations.

TRY **3.15 Inflation Rates (Example 4)** Look at the two histograms, created from the data taken from data.worldbank.org, and decide whether you think the standard deviation of inflation rates in Australia (A) was larger or smaller than the standard deviation of inflation rates in New Zealand (B). Explain.



3.16 Suits of cards The histograms contain data with a range of 1 to 4. Which group would have the larger standard deviation, group A or group B? Why?



TRY **3.17 Birth Weights (Example 5)** The mean birth weight for U.S. children born at full term (after 40 weeks) is 3462 grams (about 7.6 pounds). Suppose the standard deviation is 500 grams and the shape of the distribution is symmetric and unimodal. (Source: www.babycenter.com)

a. What is the range of birth weights (in grams) of U.S.-born children from one standard deviation below the mean to one standard deviation above the mean?

b. Is a birth weight of 2800 grams (about 6.2 pounds) more than one standard deviation below the mean?

3.18 Birth Length The mean birth length for U.S. children born at full term (after 40 weeks) is 52.2 cm (about 20.6 inches). Suppose the standard deviation is 2.5 cm and the distributions are unimodal and symmetric. (Source: www.babycenter.com)

a. What is the range of birth lengths (in centimeters) of U.S.-born children from one standard deviation below the mean to one standard deviation above the mean?

b. Is a birth length of 54 cm more than one standard deviation above the mean?

TRY **3.19 Ages (Example 6)** A father has five children of ages 2, 3, 5, 8, and 9 years.

a. Calculate the standard deviation of their current ages.

b. Without doing any calculation, indicate whether the standard deviation of the children's ages in the next 15 years will be larger, smaller, or the same as the standard deviation of their current ages. Check your answers by calculating the standard deviation of the ages in 15 years. Explain how adding 15 to each number affects the standard deviation.

c. Find the mean of the children at their current ages.

d. Without doing any calculation, indicate whether the mean of the children's age in the next 15 years will be larger, smaller, or the same as the mean of the current ages. Confirm your answer, and describe how adding 15 to each number affects the mean.

3.20 Pocket Money A father pays pocket money to his daughter on different days as 10, 4, 15, and 12 U.S. dollars (USD). In cents (1 USD = 100 cents), these pocket money values are 1000, 400, 1500, and 1200.

a. Compare the mean in USD with the mean in cents; do not forget the units. Explain what multiplying each number in a data set by 100 does to the mean.

b. Compare the standard deviation in USD and in cents; do not forget the units. Explain what multiplying each number in a data set by 100 does to the standard deviation.

3.21 Cricket In the recent cricket matches, do you think the standard deviation of the average runs scored by all players in a T-20 match would be larger or smaller than the standard deviation of the average runs scored by all players in a test match? Explain.

3.22 IQ Suppose you have a data set with IQ levels of some students who perform well in quizzes and some students who are excellent athletes. Which group of students do you think would have a higher standard deviation of IQ levels? Explain.

3.23 Brain Size The brain size (in hundreds of thousands of pixels) is reported for some men and women in the table. (Source: Willerman, L., Schultz, R., Rutledge, J. N., and Bigler, E. (1991), "In Vivo Brain Size and Intelligence," *Intelligence*, 15, 223–228.)

a. Make two separate stemplots (or dotplots or histograms). Does either data set show strong skew? If so, is the distribution right-skewed or left-skewed?

b. Compare the means.

c. Compare the standard deviations.

Female	Male
8.2	10.0
9.5	10.4
9.3	9.7
9.9	9.0
8.5	9.6
8.3	10.8
8.6	9.2
8.8	9.5
8.7	8.9
8.5	8.9

Female	Male
8.1	9.1
7.9	9.6
8.3	9.4
8.0	10.6
7.9	9.5
8.7	10.0
8.6	8.8
8.3	9.5
9.5	9.3
8.9	9.4

3.24 Happiness A survey on StatCrunch asked people to report their level of happiness from 1 (least happy) to 100 (most happy). The table shows a sample of 20 female and 20 male responses. (Source: StatCrunch Responses to Happiness survey. Owner: Webster West)

a. Make two separate stemplots (or dotplots or histograms). (Don't forget to show empty stems if you are making a stemplot.) Are the distributions skewed? If so, which way?

b. Compare the means.

c. Compare the standard deviations.

Female	Male
35	98
1	85
90	98
95	69
19	84
85	70
80	100
95	3
90	88
86	65

Female	Male
97	4
45	70
64	70
80	99
42	11
50	90
93	90
75	3
82	76
100	12

3.25 Nutrition The percentage of malnutrition prevailing in children is given for urban and rural areas. The data are at this text's website. (Source: www.unicef.org)

a. Compare the mean percentage of malnutrition.

b. Compare the standard deviation of the percentage of malnutrition in children from rural and urban areas.

c. Remove the outliers of 3, 4, and 4 for urban, and compare the means again. What effect did removing the outliers have on the mean?

d. What effect do you think removing the two outliers would have on the standard deviation, and why?

3.26 Professors' Salaries The annual salaries (in Great Britain pounds [GBP]) of 22 professors and 35 assistant professors working in various institutes across the world (for 2012–2013) are given. (Source: Professor Median Salary by Disciplines, Level, and Institution, 2012–2013, accessed via StatCrunch, Owner: statcrunchhelp)

a. Compare the means and standard deviations in context.

b. Remove the outlier of 142,033 GBP for professors, make the comparison again, and comment on the effect of removing the outlier.

Professor?	
NO	Yes
64,414	95,224
64,868	94,432
61,790	93,540
64,903	102,902
60,356	96,312
55,995	87,353
57,819	85,051
75,294	106,568
57,617	86,630
80,078	119,951
67,053	94,164
56,486	87,952
59,849	92,783
91,783	142,033
54,084	82,840
57,811	85,162
57,956	87,861
61,861	96,241

Professor?	
NO	Yes
58,917	88,057
61,646	97,464
56,543	82,996
55,844	88,711
53,802	
60,404	
57,695	
57,299	
59,650	
61,155	
54,138	
67,160	
95,268	
55,270	
55,109	
68,574	
53,624	

SECTION 3.2

TRY **3.27 Violent Crime: West (Example 7)** In 2011, the mean rate of violent crime (per 100,000 people) for the 24 states west of the Mississippi River was 406. The standard deviation was 177. Assume that the distribution of violent crime rates is approximately unimodal and symmetric. *See page 158 for guidance.*

a. Between what two values would you expect to find about 95% of the rates?

b. Between what two values would you expect to find about 68% of the violent crime rates?

c. If a western state had a violent crime rate of 584 crimes per 100,000 people, would you consider this unusual? Explain.

d. If a western state had a violent crime rate of 30 crimes per 100,000 people, would you consider this unusual? Explain.

3.28 Violent Crime: East In 2011, the mean rate of violent crime (per 100,000 people) for the 10 northeastern states was 314, and the standard deviation was 118. Assume the distribution of violent crime rates is approximately unimodal and symmetric.

a. Between what two values would you expect to find about 95% of the rates?

b. Between what two values would you expect to find about 68% of the rates?

c. If a northeastern state had a violent crime rate of 896 crimes per 100,000 people, would you consider this unusual? Explain.

d. If a northeastern state had a violent crime rate of 403 crimes per 100,000 people, would you consider this unusual? Explain.

TRY **3.29 Burglary (Example 8)** In 2014, the mean burglary rate (per 1000 houses) for all states in a country was 978; the standard deviation was 32. Assume the distribution of burglary rates is approximately unimodal and symmetric.

a. Approximately what percentage of states would you expect to have burglary rates between 914 and 1042?

b. Approximately what percentage of states would you expect to have burglary rates between 946 and 1010?

c. If someone guessed that the burglary rate in one of the states was 2031, would you agree that that number was consistent with this data set?

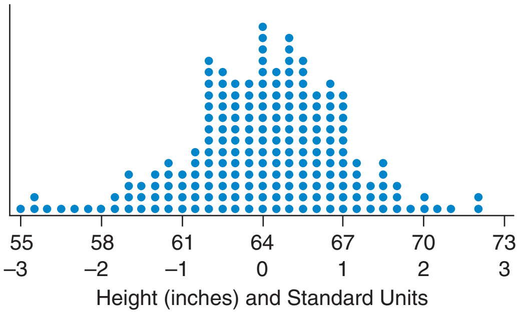
3.30 Burglary In 2014, the mean burglary rate (per 1000 houses) for all cities in a state was 250; the standard deviation was 25. Assume the distribution of burglary rates is approximately unimodal and symmetric.

a. Approximately what percentage of cities would you expect to have burglary rates between 225 and 275?

b. Approximately what percentage of cities would you expect to have burglary rates between 200 and 300?

c. If someone guessed that the burglary rate in one of the cities was 0, would you agree that that number was consistent with this data set?

3.31 Heights and *z*-Scores The dotplot shows heights of college women; the mean is 64 inches (5 feet 4 inches), and the standard deviation is 3 inches.



a. What is the z-score for a height of 58 inches (4 feet 10 inches)?

b. What is the height of a woman with a z-score of 1?

3.32 Heights Refer to the dotplot in the previous question.

a. What is the height of a woman with a z-score of -1?

b. What is the z-score for a woman who is 70 inches tall (5 feet 10 inches)?

TRY **3.33 Unusual Strike (Example 9)** A batsman has a strike rate of 180 and a standard deviation of 12. Which is more unusual, a strike rate of 192 or a strike rate of 156?

3.34 Flight Arrivals Distributions of delays in flight arrival of a particular airline are roughly bell-shaped. The mean delay in arrival time is 12 minutes, and the standard deviation is 4 minutes. Which is more usual, a flight arriving 4 minutes earlier than the mean delay time or a flight arriving 4 minutes after this time? Explain.

3.35 Low-Birth-Weight Babies (Example 10) Babies born weighing 2500 grams (about 5.5 pounds) or less are called low-birth-weight babies, and this condition sometimes indicates health problems for the infant. The mean birth weight for U.S.-born children is about 3462 grams (about 7.6 pounds). The mean birth weight for babies born one month early is 2622 grams. Suppose both standard deviations are 500 grams. Also assume that the distribution of birth weights is roughly unimodal and symmetric. (Source: www.babycenter.com)

a. Find the standardized score (z-score), relative to all U.S. births, for a baby with a birth weight of 2500 grams.

b. Find the standardized score for a birth weight of 2500 grams for a child born one month early, using 2622 as the mean.

c. For which group is a birth weight of 2500 grams more common? Explain what that implies. Unusual z-scores are far from 0.

3.36 Birth Lengths Babies born after 40 weeks gestation have a mean length of 52.2 centimeters (about 20.6 inches). Babies born one month early have a mean length of 47.4 cm. Assume both standard deviations are 2.5 cm and the distributions are unimodal and symmetric. (Source: www.babycenter.com)

a. Find the standardized score (z-score), relative to all U.S. births, for a baby with a birth length of 45 cm.

b. Find the standardized score of a birth length of 45 cm for babies born one month early, using 47.4 as the mean.

c. For which group is a birth length of 45 cm more common? Explain what that means.

3.37 Men's Shoe Sizes Assume that men's shoe sizes have a mean of 7 and a standard deviation of 1.5.

a. What men's shoe size corresponds to a z-score of 1.00?

b. What men's shoe size corresponds to a z-score of -1.50?

3.38 Leaves The number of leaves taken per year by employees of an office has a mean of 50 and a standard deviation of 5.

a. What number of leaves corresponds to a z-score of 1.25?

b. What number of leaves corresponds to a z-score of -1.75?

SECTION 3.3

Note: Reported interquartile ranges will vary depending on technology.

3.39 Name two measures of the center of a distribution, and state the conditions under which each is preferred for describing the typical value of a single data set.

3.40 Name two measures of the variation of a distribution, and state the conditions under which each measure is preferred for measuring the variability of a single data set.

TRY **3.41 Pixar Animated Movies (Example 11)** The ten top-grossing Pixar animated movies for the U.S. box office up to June 2013 are shown in the table on the next page, in millions of dollars. (Source: www.pixar.com)

a. Sort the gross income from smallest (on the left) to largest, and write down the sorted list. Find the median by averaging the two middle numbers. Interpret the median in context.

b. Using the sorted data, find Q1 and Q3. Then find the interquartile range and interpret it in context.

Movie	\$Millions
Toy Story 3	415
Finding Nemo	340
Up	293
Incredible	261
Monsters, Inc.	256
Monsters University	255
Toy Story 2	246
Cars	244
Brave	237
WALL-E	224

3.42 Best-selling Books The numbers of copies sold (in millions) for the top ten best-selling books in 2015–2016 are shown in the table. (Source: www.top101news.com)

Book	Copies (in millions)
A Tale of two Cities	200
Lord of the Rings	150
Le Petit Prince	140
Harry Potter and the Philosopher's Stone	107
And Then There Were None	100
Dream of the Red Chamber	100
The Hobbit	100
She: A Story of Adventure	100
The Lion, the Witch, and the Wardrobe	85
The DaVinci Code	80

a. Find and interpret the median copies for the ten bestsellers.

b. Find and interpret the interquartile range for these books.

TRY **3.43 Pixar Animated Movies Again (Example 12)** Find the median and interquartile range of the top seven Pixar animated movies; refer to Exercise 3.41 for the data.

3.44 Best-selling Books Find the median and interquartile range of the top seven bestsellers; refer to Exercise 3.42 for the data.

TRY **3.45 Happiness (Example 13)** Use the data in Exercise 3.24 on the happiness of men and women.

a. Compare the median happiness level for the men to that for the women by copying the sentence below and filling in the blanks.

The median for the men was _________ and the median for the women was _________, showing that the typical _________ (man or woman) tended to be a bit happier.

b. Compare the interquartile range for the men to that for the women by copying the sentence below and filling in the blanks.

The interquartile range for the men was _________ and the interquartile range for the women was _________, showing more variation in happiness level for the _________ (men or women).

3.46 Nutrition The rate of malnutrition is given for urban areas of 45 countries and rural areas of 30 countries. (Source: www.unicef.org)

a. Compare the median rates of malnutrition for urban and rural areas.

b. Compare the interquartile ranges of the rates of malnutrition for urban and rural areas.

c. Remove the outliers of 3, 4, and 4 for urban areas, and compare the median again. What effect did removing the outliers have on the median?

d. What effect do you think removing the three outliers would have on the interquartile range?

Urban	Rural	Urban	Rural
30	19	23	32
31	33	33	30
3	22	30	35
16	21	22	45
36	38	23	47
28	21	20	42
30	47	38	32
21	36	4	
38	25	27	
24	29	4	
22	37	19	
38	34	33	
31	41	23	
25	42	16	
20	41	21	
21	24	27	
27	32	30	
19	56	26	
10	28	37	
15	27	26	
23	18	13	
22	33	16	
36	34		

SECTION 3.4

3.47 Outliers

a. In your own words, describe to someone who knows only a little statistics how to recognize when an observation is an outlier. What action(s) should be taken with an outlier?

b. Which measure of the center (mean or median) is more resistant to outliers, and what does "resistant to outliers" mean?

3.48 Center and Variation When you are comparing two sets of data, and one set is strongly skewed and the other is symmetric, which measures of the center and variation should you choose for the comparison?

3.49 Mistake A teacher recorded the test score of a student in five different subjects. The maximum score for each subject is 50. As you can see, a mistake was made in recording one entry.

25, 30, 41, 45, 280

When the mistake was corrected by removing the extra 0, will the median scores change? Will the mean change? Explain without doing any calculation.

3.50 Factory Strike In 2010, the laborers of a major factory in Australia went on strike. At that time, the average salary of a laborer was AU$ 21.60 per hour, and the median salary was AU$ 20.00. If you were representing the owners, which summary would you use to convince the public that a strike was not needed? If you were representing the laborers, which would you use? Why was there a discrepancy between the mean and median hourly rates? Explain. (Source: www.payscale.com)

3.51 Heads The graphs show the circumferences of heads for a group of men and a group of women.

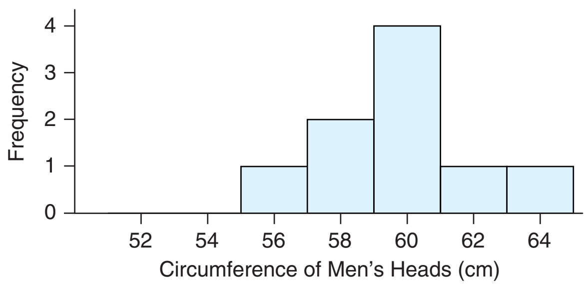
a. If you were describing the men's heads in terms of shape, center, and spread, without comparing them to the women's heads, would you use the mean and standard deviation or the median and interquartile range? Why?

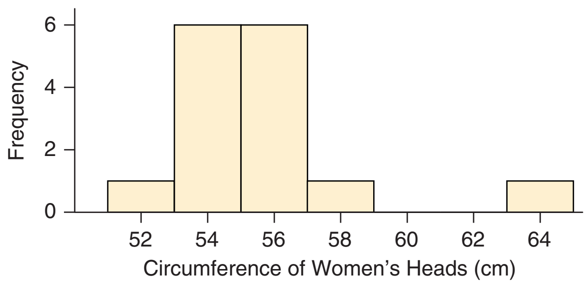
b. If you were describing the women's heads in terms of shape, center, and spread, without comparing them to the men's heads, would you use the mean and standard deviation or the median and interquartile range? Why?

c. If you were comparing the two groups, what measures would you use, and why?

d. In which of the two graphs would the mean and median be close together, and why?

e. In which of the two graphs would the mean and median be farther apart, and which would be larger?





3.52 House Prices The graphs show the house prices (in hundreds of thousands of dollars) in two fictitious towns.

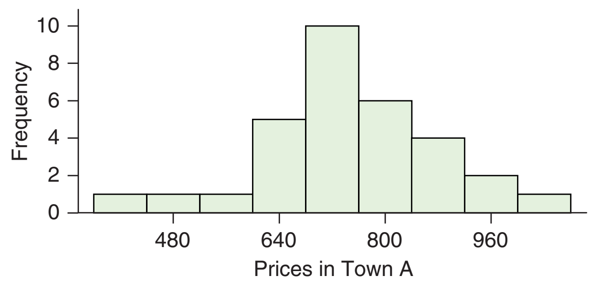
a. If you were describing the prices in Town A in terms of shape, center, and spread, without comparing them to the prices in Town B, would you use the mean and standard deviation or the median and interquartile range? Why?

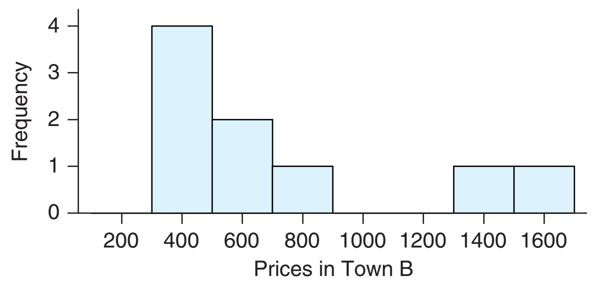
b. If you were describing the prices in Town B in terms of shape, center, and spread, without comparing them to the prices in Town A, would you use the mean and standard deviation or the median and interquartile range? Why?

c. If you were comparing the two groups, what measures would you use, and why?

d. In which of the two graphs are the mean and median closer together, and why?

e. In which of the two graphs are the mean and median farther apart, and which would be larger?





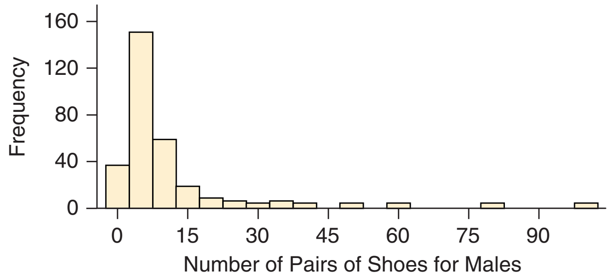
TRY **3.53 Shoes (Example 14)** The histograms show the number of pairs of shoes reported for 300 males and for 300 females. Descriptive statistics are also shown. (Source: StatCrunch Responses to Shoe Survey. Owner: scsurvey)

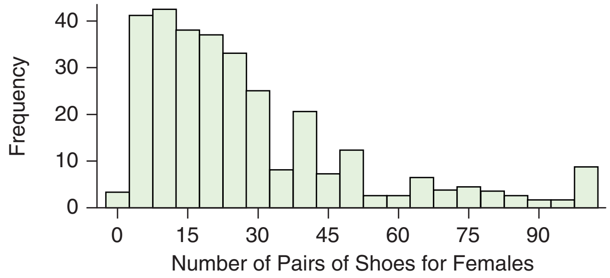
a. Describe the shape of each histogram.

b. Because of the shapes, what measures of the center should be compared, the means or the medians?

c. Because of the shapes, what measure of spread is preferred, the standard deviation or the interquartile range?

d. Compare the centers and spreads in context.

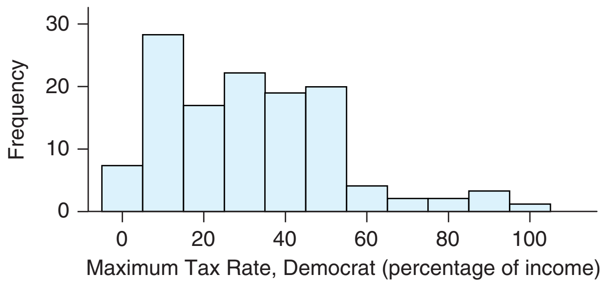


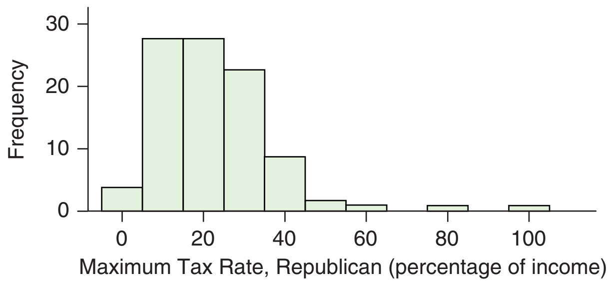


Descriptive Statistics: Female, Male

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Max	Interquartile range
Female	300	27.06	22.19	1.00	11.25	20.00	35.00	100.00	23.75
Male	300	9.753	13.245	1.000	4.000	6.000	10.000	100.000	6.000

3.54 Tax Rates A StatCrunch survey asked people what maximum income tax rate (as a percentage of income) should be allowed and whether they were Republican or Democrat. Compare the two political parties. Compare shapes and appropriate measures of the center and spread.





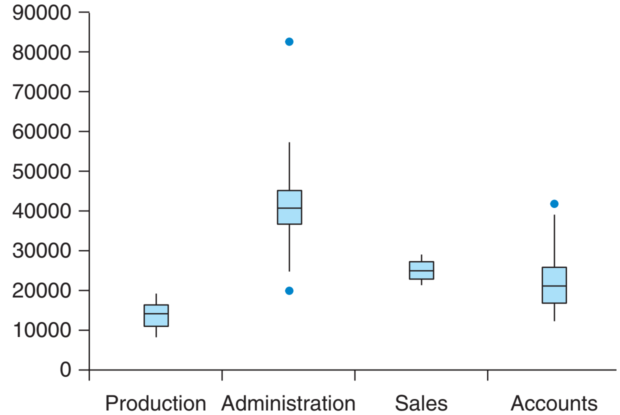
Descriptive Statistics: Democrat, Republican

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
MaxTax_Democrat	125	30.40	21.11	0.00	10.00	30.00	45.00	100.00
MaxTax_Republican	97	21.16	15.35	0.00	10.00	20.00	28.00	100.00

SECTION 3.5

3.55 Departmental Disparity The figure shows the salary for 60 employees of an organization. The organization is divided into four departments—Production (P), Administration (AD), Sales (S), and Accounts (A). In the Administration department, the potential outlier is CEO, and in the Accounts department, the potential outlier is CA.

Why is it best to compare medians and interquartile ranges for these data, rather than comparing means and standard deviations? List the appropriate median salary for each department; for example, the median for the Production department is between 15,000 and 20,000. Also, arrange the departments from lowest interquartile range (on the left) to the highest.

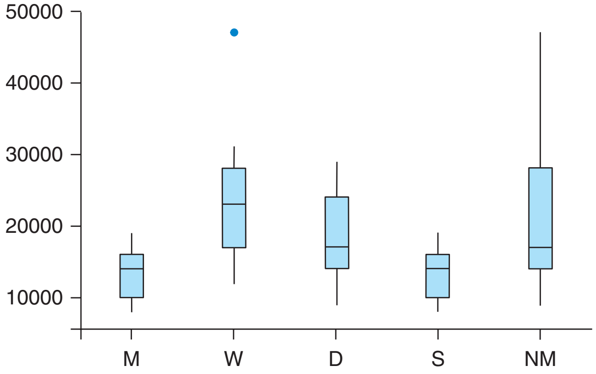


3.56 Health Insurance Rates The boxplot shows the rate of persons per 100,000 who are not covered by health insurance in different countries of the Middle East. The persons are categorized as Married (M), Widowed (W), Divorced (D), Separated (S), and Never Married (NM). (Source: *2015 World Almanac and Book of Facts*)

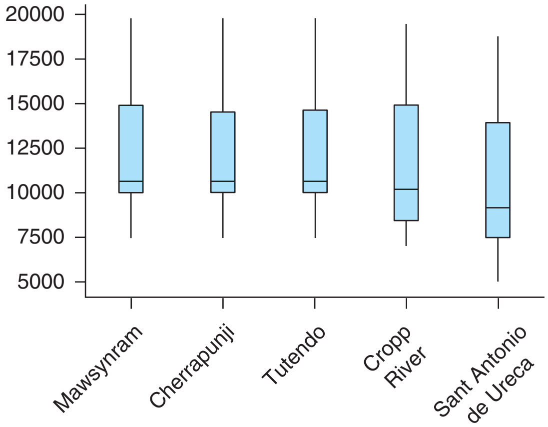
a. List the approximate uninsured rate per 100,000 people for all the categories; for example, the median for M is about 15,000.

b. List the categories from the lowest to highest interquartile range.

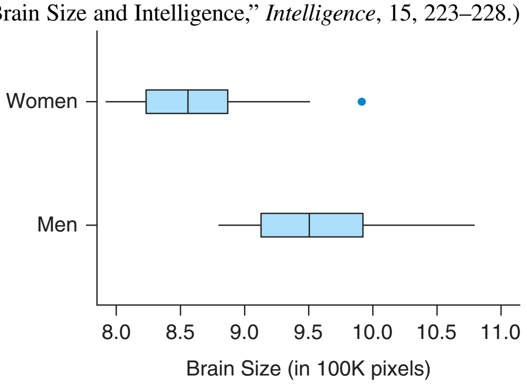
c. Why is the interquartile range a better measure of the variability for these data than the range is?



3.57 Rainfall The boxplot shows rainfall for the five rainiest places in the world. Each place's boxplot was made from 12 rainfall records: the average monthly rainfall over a period of years. Which place tends to be the wettest? Which place has the most variation in rainfall? Compare the rainfall records of places by interpreting the boxplot. If rainfall were the only factor to consider, which place would you choose to live in and why? (Source: *The Daily Telegraph,* August 18, 2014)



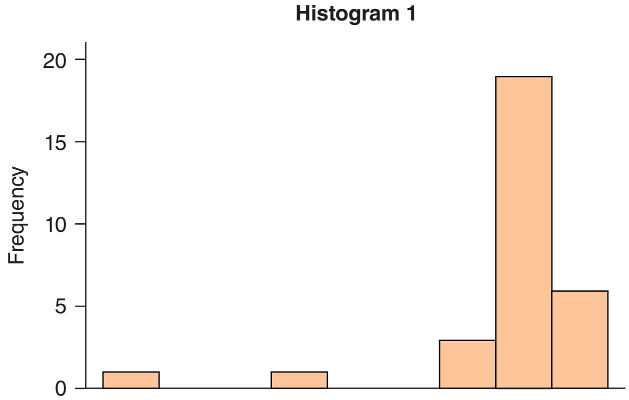
3.58 Brain Size The boxplots show the brain size (in hundreds of thousands of pixels) for 20 men and 20 women. Estimate the numerical values of the medians by using the boxplots. Do these men, or do these women, have greater variation in brain size? Why? (Source: Willerman, L., Schultz, R., Rutledge, J. N., and Bigler, E. (1991), "In Vivo Brain Size and Intelligence," *Intelligence*, 15, 223–228.)

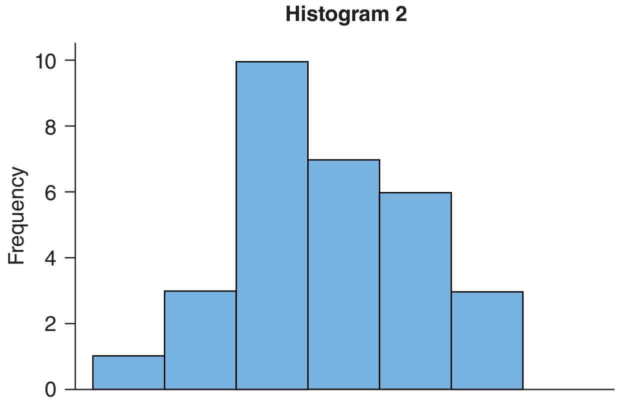


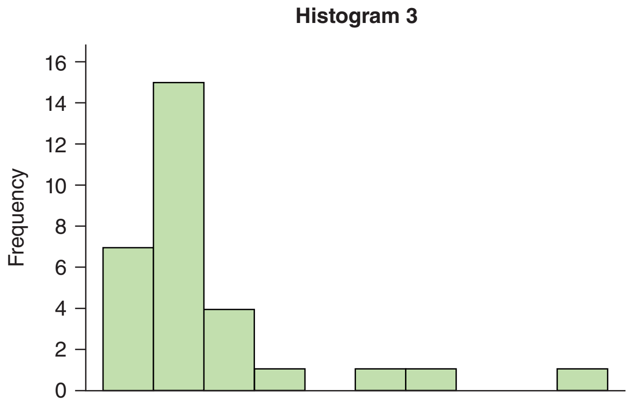
3.59 Matching Boxplots and Histograms

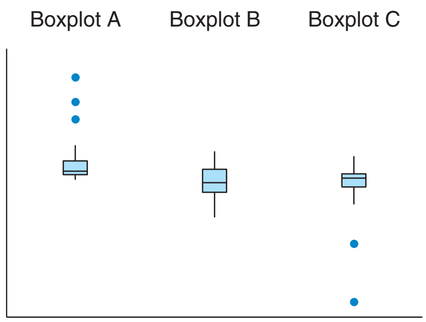
a. Report the shape of each histogram.

b. Match each histogram with the corresponding boxplot (A, B, or C).

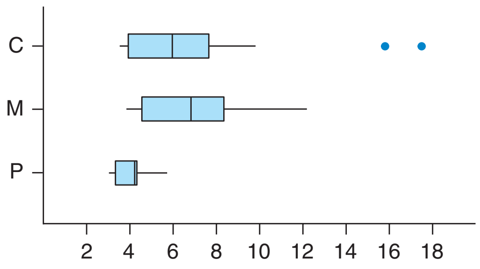




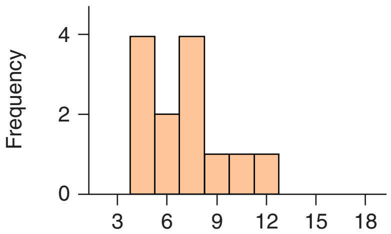




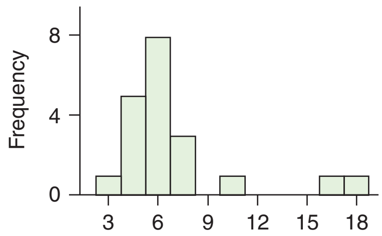
3.60 Matching Boxplots and Histograms Match each of the histograms (X, Y, and Z) with the corresponding boxplot (C, M, or P). Explain your reasoning.



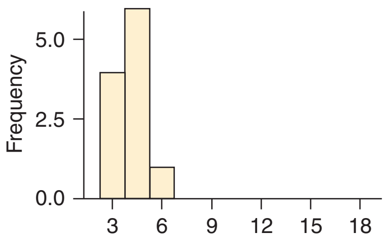
X



Y



Z



3.61 Sleep Time of Animals Data at this text's website show the average amount of time animals sleep per day (in hours). The number listed for humans is 8 hours. You may either make a boxplot with technology using the data at this text's website or make a sketch of the boxplot from the descriptive statistics below. Compare the median with humans' median of 8 hours. There are no potential outliers. (Source: http://faculty.washington.edu/chudler/chasleep.html)

Descriptive Statistics: SleepHours

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
SleepHours	45	10.864	4.707	1.900	7.400	10.800	14.450	19.900

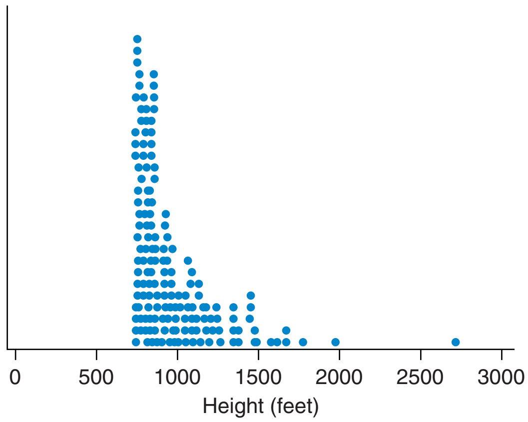
* **3.62 BA Percentage** The data show the percentage of residents with bachelor's degrees in the 50 states and Washington, DC. Make a boxplot of the data. You may either make a boxplot with technology using the data at this text's website or sketch the boxplot using the descriptive statistics below. Washington, D.C. is a potential outlier with a bachelor's rate of 45.7%; there are no other potential outliers. The next highest rate is for Massachusetts with 36.7%.

Descriptive Statistics: Percent

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Percent	51	27.331	5.447	15.300	24.300	25.600	30.800	45.700

3.63 Tall Buildings The dotplot shows the distribution of the world's tallest buildings with respect to their height, in feet. The five-number summary is

745 ft, 810 ft, 883 ft, 1093 ft, 2717 ft

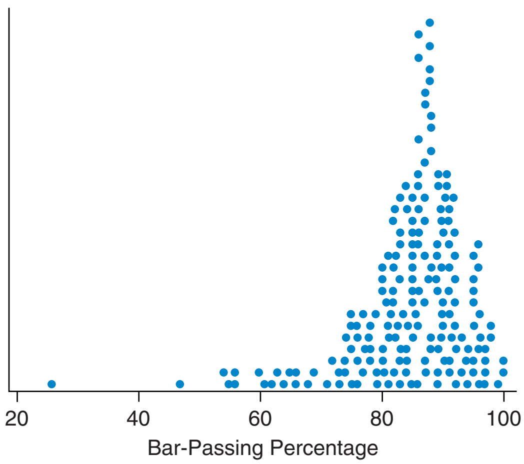


Draw a boxplot. Describe the shape of the distribution.

3.64 Passing the Bar Exam The dotplot shows the distribution of passing rates for the bar exam at 185 law schools in the United States in 2009.

The five number summary is

26, 80, 86, 90, 100



Draw the boxplot and explain how you determined where the whiskers go.

* **3.65 Exam Scores** The five-number summary for a distribution of final exam scores is

40, 78, 80, 90, 100

Explain why it is not possible to draw a boxplot based on this information. (*Hint:* What more do you need to know?)

* **3.66 Exam Scores** The five-number summary for a distribution of final exam scores is

60, 78, 80, 90, 100

Is it possible to draw a boxplot based on this information? Why or why not?

CHAPTER REVIEW EXERCISES

TRY **3.67 Death Row: South (Example 15)** The table shows the numbers of capital prisoners (prisoners on death row) in 2013 in the southern U.S. states. (Source: http://www.deathpenaltyinfo.org)

a. Find the median number of prisoners and interpret (using a sentence in context).

b. Find the interquartile range (showing Q3 and Q1 in the process) to measure the variability in the number of prisoners.

c. What is the mean number of prisoners?

d. Why is the mean so much larger than the median?

e. Why is it better to report the median, instead of the mean, as a typical measure?

State	CapPris
Alabama	198
Arkansas	38
Florida	413
Georgia	97
Kentucky	34
Louisiana	88
Maryland	5
Mississippi	48

State	CapPris
North Carolina	161
Oklahoma	60
South Carolina	53
Tennessee	87
Texas	300
Virginia	11
West Virginia	0

3.68 Death Row: West The table shows the numbers of capital prisoners (prisoners on death row) in 2013 in the western U.S. states. (Source: deathpenaltyinfo.org)

State	CapPris
Alaska	0
Arizona	127
California	727
Colorado	4
Hawaii	0
Idaho	13
Montana	2

State	CapPris
Nevada	79
New Mexico	2
Oregon	37
Utah	9
Washington	8
Wyoming	1

a. Find the median.

b. Find the interquartile range (showing Q3 and Q1 in the process).

c. Find the mean number of capital prisoners.

d. Why is the mean so much larger than the median?

TRY * **3.69 Head Circumference (Example 16)** Following are head circumferences, in centimeters, for some men and women in a statistics class.

Men: 58, 60, 62.5, 63, 59.5, 59, 60, 57, 55

Women: 63, 55, 54.5, 53.5, 53, 58.5, 56, 54.5, 55, 56, 56, 54, 56, 53, 51

Compare the circumferences of the men's and the women's heads. Start with histograms to determine shape; then compare appropriate measures of center and spread, and mention any outliers. *See page 158 for guidance.*

3.70 Heights of Sons and Dads The data at this text's website give the heights of 18 male college students and their fathers, in inches.

a. Make histograms and describe the shapes of the two data sets from the histograms.

b. Fill in the following table to compare descriptive statistics.

	Mean	Median	Standard deviation	Interquartile range
Sons	______	______	______	______
Dads	______	______	______	______

c. Compare the heights of the sons and their dads, using the means and standard deviations.

d. Compare the heights of the sons and their dads, using the medians and interquartile ranges.

e. Which pair of statistics is more appropriate for comparing these samples: the mean and standard deviation or the median and interquartile range? Explain.

3.71 Final Exam Grades The data that follow are final exam grades for two sections of statistics students at a community college. One class met twice a week relatively late in the day; the other class met four times a week at 11 a.m. Both classes had the same instructor and covered the same content. Is there evidence that the performances of the classes differed? Answer by making appropriate plots (including side-by-side boxplots) and reporting and comparing appropriate summary statistics. Explain why you chose the summary statistics that you used. Be sure to comment on the shape of the distributions, the center, and the spread, and be sure to mention any unusual features you observe.

11 a.m. grades: 100, 100, 93, 76, 86, 72.5, 82, 63, 59.5, 53, 79.5, 67, 48, 42.5, 39

5 p.m. grades: 100, 98, 95, 91.5, 104.5, 94, 86, 84.5, 73, 92.5, 86.5, 73.5, 87, 72.5, 82, 68.5, 64.5, 90.75, 66.5

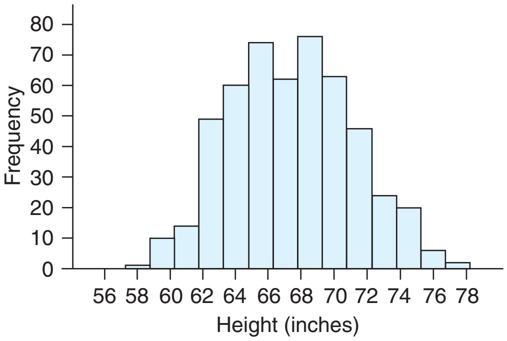
3.72 Speeding Tickets College students Diane Glover and Esmeralda Olguin asked 25 men and 25 women how many speeding tickets they had received in the last three years.

Men: 14 men said they had 0 tickets, 9 said they had 1 ticket, 1 had 2 tickets, and 1 had 5 tickets.

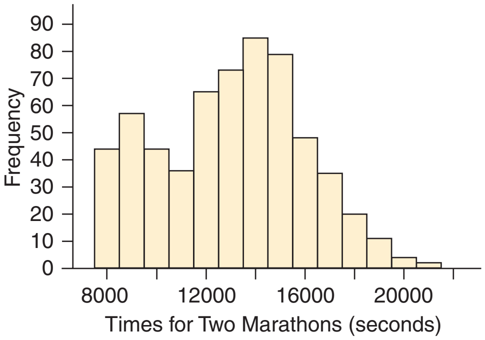
Women: 18 said they had 0 tickets, 6 said they had 1 ticket, and 1 said she had 2 tickets.

Is there evidence that the men and women differed? Answer by making appropriate plots and comparing appropriate summary statistics. Be sure to comment on the shape of the distributions and to mention any unusual features you observe.

3.73 Heights The graph shows the heights for a large group of adults. Describe the distribution, and explain what might cause this shape. (Source: www.amstat.org)



3.74 Marathon Times The histogram of marathon times includes data for men and women and also for both an Olympic marathon and an amateur marathon. Greater values indicate slower runners. (Sources: www.forestcityroadraces.com and www.runnersworld.com)



a. Describe the shape of the distribution.

b. What are two different possible reasons for the two modes?

c. Knowing that there are usually fewer women who run marathons than men, and that more people ran in the amateur marathon than in the Olympic marathon, look at the size of the mounds and try to decide which of the reasons stated in part b is likely to cause this. Explain.

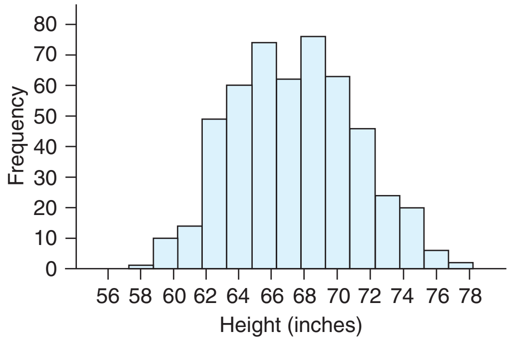
* **3.75 Soda Consumption** A StatCrunch survey asked people about their consumption of soda as a percentage of liquid intake. The data for 130 males and 130 females can be found at this text's website. Compare the two groups with histograms and appropriate measures of center and spread. (Source: StatCrunch: Responses to Soda survey. Owner: scsurvey)

3.76 Holiday Spending A StatCrunch survey asked people how much money they spent for gifts during the holidays. Compare the males and females, using both graphical and numerical summaries. Calculate the numbers from the data at this text's website.

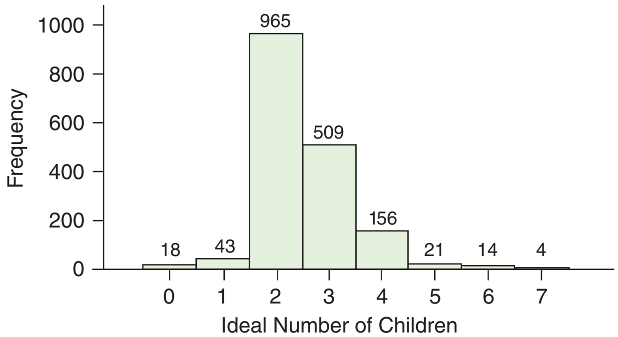
3.77 Heights

a. State an approximate value for the mean height by looking at the graph.

b. Here is a proposed method for finding an approximation for the standard deviation based on the histogram: Find the approximate range and divide by 6. This method comes from the idea that nearly all the data should be within 3 standard deviations of the mean. Use this method to find an approximate standard deviation.



3.78 Ideal Family In 2012, the General Social Survey asked respondents how many children they felt would be in an "ideal" family. The histogram contains the data from 1730 people who responded to the survey.



a. Approximately what is the mean ideal number of children? Explain how you chose this value.

b. What is the approximate value for the median ideal number of children? Describe how you chose this value.

c. Find the mean by completing the work that is started below:

$$\bar{x} = \frac{18(0) + 43(1) + 965(2) + \cdots}{1730}$$

d. Explain how the method in part c is related to the usual method of finding the mean, which has all the raw numbers given, without frequencies.

e. Which is more appropriate to report for these data, the mean or the median? Why?

3.79–3.82 *Construct two sets of numbers with at least five numbers in each set (showing them as dotplots) with the following characteristics:*

3.79 The means are the same, but the standard deviation of one of the sets is larger than that of the other. Report the mean and both standard deviations.

3.80 The means are different, but the standard deviations are the same. Report both means and the standard deviation.

3.81 The mean of set A is larger than that of set B, but the median of set B is larger than the median of set A. Label each dotplot with its mean and median in the correct place.

* **3.82** The standard deviation of set A is larger, but the interquartile range of set B is larger. Report both standard deviations and interquartile ranges.

* **3.83 Population Density** Data were recorded for each of the Egyptian districts: the district, its population (in 2006), its area (in square miles), and the governorate in which it is located (Markaz, Kism, New city, etc.). For each state, find the population density: the number of people divided by the number of square miles. Write a few sentences comparing the distribution of population densities in the different governorates. Support your description with appropriate graphs. (Source: www.citypopulation.de)

* **3.84 Population Increase** Data were recorded for each of the Egyptian districts: the district, its population in 1996, its population in 2006, and the governorate in which it is located (Markaz, Kism, New City, etc.). Find the population increase percentage for each state by applying the following formula:

$$\frac{\text{pop}_{2006} - \text{pop}_{1996}}{\text{pop}_{1996}} \times 100\%$$

Write a few sentences comparing the distribution of percentage population increases for the different governorates. Support your description with appropriate graphs. (Source: www.citypopulation.de)

TRY **3.85 Surfing, Again** College students and surfers Rex Robinson and Sandy Hudson collected data on the self-reported numbers of days surfed in a month for 30 longboard surfers and 30 shortboard surfers.

Longboard: 4, 9, 8, 4, 8, 8, 7, 9, 6, 7, 10, 12, 12, 10, 14, 12, 15, 13, 10, 11, 19, 19, 14, 11, 16, 19, 20, 22, 20, 22

Shortboard: 6, 4, 4, 6, 8, 8, 7, 9, 4, 7, 8, 5, 9, 8, 4, 15, 12, 10, 11, 12, 12, 11, 14, 10, 11, 13, 15, 10, 20, 20

a. Compare the typical number of days surfing for these two groups by placing the correct numbers in the blanks in the following sentence: The median for the longboards was __________ days, and the median for the shortboards was __________ days, showing that those with __________ boards typically surfed more days in this month.

b. Compare the interquartile ranges by placing the correct numbers in the blanks in the following sentence: The interquartile range for the longboards was __________ days, and the interquartile range for the shortboards was __________ days, showing more variation in the days surfed this month for the __________ boards.

3.86 Eating Out, Again College student Jacqueline Loya asked 50 employed students how many times they went out to eat last week. Half of the students had full-time jobs and half had part-time jobs.

Full-time: 5, 3, 4, 4, 4, 2, 1, 5, 6, 5, 6, 3, 3, 2, 4, 5, 2, 3, 7, 5, 5, 1, 4, 6, 7

Part-time: 1, 1, 5, 1, 4, 2, 2, 3, 3, 2, 3, 2, 4, 2, 1,2, 3, 2, 1, 3, 3, 2, 4, 2, 1

a. Using the median values, write a sentence comparing the typical numbers of times the two groups ate out.

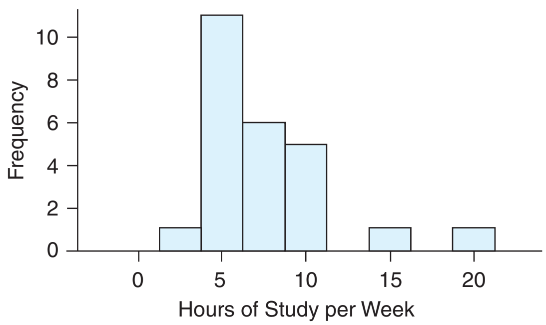
b. Using the interquartile ranges, write a sentence comparing the variability of these two groups.

3.87 Study Hours A group of 50 statistics students, 25 men and 25 women, reported the number of hours per week spent studying statistics.

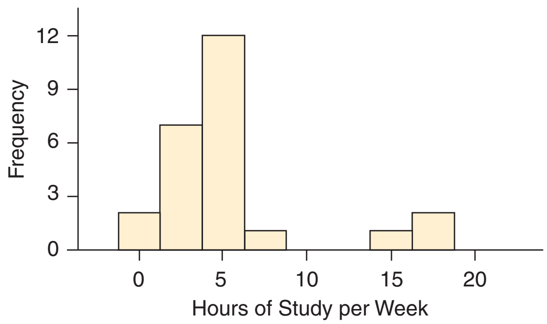
a. Refer to the histograms. Which measure of the center should be compared: the means or the medians? Why?

b. Compare the distributions in context using appropriate measures. (Don't forget to mention outliers, if appropriate). Refer to the Minitab output for the summary statistics.

Hours of Study for Women



Hours of Study for Men



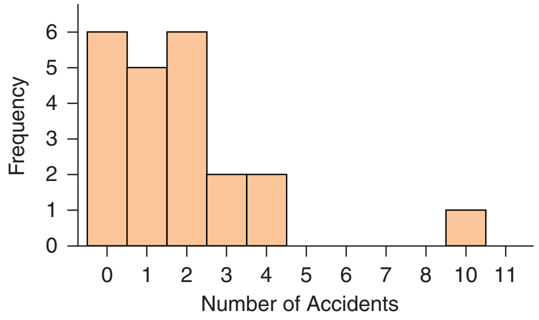
Minitab Statistics: Men, Women

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Men	25	5.20	4.378	1.00	2.50	4.00	5.50	17.00
Women	25	7.52	3.787	2.00	5.00	7.00	9.50	20.00

3.88 Driving Accidents College student Sandy Hudson asked a group of college students the total number of traffic accidents they had been in as drivers. The histograms are shown, and the table displays some descriptive statistics.

Men's Driving Accidents

Women's Driving Accidents



Minitab Statistics: Men, Women

Variable	N	Mean	StDev	Min	Q1	Median	Q3	Max
Men	20	0.850	0.988	0.00	0.00	0.505	2.00	3.00
Women	22	1.864	2.210	0.00	0.00	1.50	2.25	10.00

a. Refer to the histograms. If we wish to compare the typical numbers of accidents for these men and women, should we compare the means or the medians? Why?

b. Write a sentence or two comparing the distributions of numbers of accidents for men and women in context.

3.89 Exam Scores An exam has a mean of 70 and a standard deviation of 10. What exam score corresponds to a z-score of 1.5?

3.90 Girls' Weights The weight-for-age for girls of 8 years is 26 kg with a standard deviation of 4.5 kg. How heavy is an 8-year-old girl with a z-score of -0.4? (Source: www.who.int)

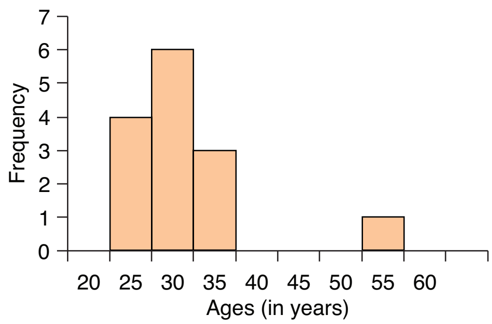
3.91 Salaries The monthly salaries of qualified professionals have a mean of \$50,000 and a standard deviation of \$20,000, while those of semi-qualified professionals have a mean of \$29,000 and a standard deviation of \$3,500. Assuming both types of salaries have distributions that are unimodal and symmetric, which is more unusual: a qualified professional having a salary of \$80,000 or a semi-qualified professional having a salary of \$36,000? Show your work.

3.92 Children's Heights Mrs. Diaz has two children: a three-year-old boy 43 inches tall and a ten-year-old girl 57 inches tall. Three-year-old boys have a mean height of 38 inches and a standard deviation of 2 inches, and ten-year-old girls have a mean height of 54.5 inches and a standard deviation of 2.5 inches. Assume the distributions of boys' and girls' heights are unimodal and symmetric. Which of Mrs. Diaz's children is more unusually tall for his or her age and gender? Explain, showing any calculations you perform. (Source: www.kidsgrowth.com)

3.93 Players' Ages Here are the ages of some players in a team: 22, 25, 27, 26, 29, 28, 24, 31, 34, 29, 30, 34, and 21. The coach's age is 55 and should be included as one of the ages when you do the calculations below. The figure shows a histogram of the data.

a. Describe the distribution of ages by giving the shape and, the numerical value for an appropriate measure of spread, as well as mentioning any outliers.

b. Make a rough sketch (or copy) of the histogram, and mark the approximate locations of the mean and the median. Why are they not at the same location?



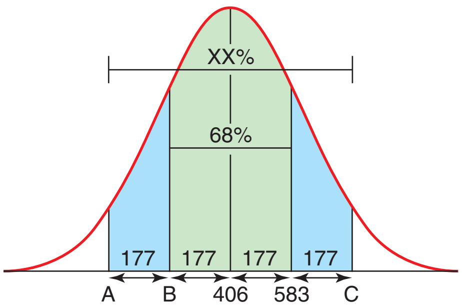
3.94 House Prices The figure, which is from data taken from the *Ventura County Star,* shows a histogram of house prices in Thousand Oaks, California, in 2009. The location of the mean and median are marked with letters. Which is the location of the mean, A or B? Explain why the mean and median are not the same.

GUIDED EXERCISES

3.27 Violent Crime: West In 2011, the mean rate of violent crime (per 100,000 people) for the 24 states west of the Mississippi River was 406, and the standard deviation was 177. Assume the distribution of violent crime rates is approximately unimodal and symmetric. The Empirical Rule says 95% of data should lie within two standard deviations of the mean and 68% of the data should lie within one standard deviation of the mean with unimodal and symmetric data.

QUESTIONS Answer these questions by following the numbered steps.

a. Between which two values would you expect to find about 95% of the violent crime rates?
b. Between which two values would you expect to find about 68% of the violent crime rates?
c. If a western state had a violent crime rate of 584 crimes per 100,000 people, would you consider this unusual? Explain.
d. If a western state had a violent crime rate of 30 crimes per 100,000 people, would you consider this unusual? Explain.



The green area is 68% of the area under the curve. The green and blue areas together shade XX% of the area. The numbers without percentage signs are crime rates, with a mean of 406 and a standard deviation of 177.

Step 1 ▶ Percentage
Reproduce Figure A, which is a sketch of the distribution of the crime rates. What percentage of data should occur within two standard deviations of the mean? Include this number in the figure, where it now says XX.

Step 2 ▶ Why 583?
How was the number 583, shown on the sketch, obtained?

Step 3 ▶ A, B, and C
Fill in numbers for the crime rates for areas A, B, and C.

Step 4 ▶ Boundaries for 95%
Read the answer from your graph:

a. Between which two values would you expect to find about 95% of the rates?

Step 5 ▶ Boundaries for 68%
Read the answer from your graph:

b. Between which two values would you expect to find about 68% of the violent crime rates?

Step 6 ▶ Unusual?

c. If a western state had a violent crime rate of 584 crimes per 100,000 people, would you consider this unusual? Many people would consider any numbers outside your boundaries of A and C to be unusual because such values occur in 5% of the states or fewer.

Step 7 ▶ Unusual?

d. If a western state had a violent crime rate of 30 crimes per 100,000 people, would you consider this unusual? Explain.

3.69 Head Circumference The head circumferences, in centimeters, for some men and women in a statistics class are given.

Men: 58, 60, 62.5, 63, 59.5, 59, 60, 57, 55

Women: 63, 55, 54.5, 53.5, 53, 58.5, 56, 54.5, 55, 56, 56, 54, 56, 53, 51

QUESTION Compare the circumferences of the men's and women's heads by following the numbered steps.

Step 1 ▶ Histograms
Make histograms of the two sets of data separately. (You may use the same horizontal axes—if you want to—so that you can see the comparison easily.)

Step 2 ▶ Shapes
Report the shapes of the two data sets.

Step 3 ▶ Measures to Compare
If either data set is skewed or has an outlier (or more than one), you should compare medians and interquartile ranges for *both* groups. If both data sets are roughly symmetric, you should compare means and standard deviations. Which measures should be compared with these two data sets?

Step 4 ▶ Compare Centers
Compare the centers (means or medians) in the following sentence: The ________ (mean or median) head circumference for the men was ________ cm, and the ________ (mean or median) head circumference for the women was ________ cm. This shows that the typical head circumference was larger for the________.

Step 5 ▶ Compare Variations
Compare the variations in the following sentence: The ________ (standard deviation or interquartile range) for the head circumferences for the men was ________ cm, and the ________ (standard deviation or interquartile range) for the women was ________ cm. This shows that the ________ tended to have more variation, as measured by the ________ (standard deviation or interquartile range).

Step 6 ▶ Outliers
Report any outliers, and state which group(s) they belong to.

Step 7 ▶ Final Comparison
Finally, in a sentence or two, make a complete comparison of head circumferences for the men and the women.

CHECK YOUR TECH

Finding the Standard Deviation of Vacation Days for Several Countries

The Minitab output shown gives the mean and standard deviation of the data set. (Source: *2009 World Almanac and Book of Facts*)

Descriptive Statistics: Days

```
Variable  N   Mean  StDev
Days      6  30.00  10.35
```

Minitab Output

The table at the left reports the mean number of vacation days per year for several countries.

Country	Mean Days
United States	13
Japan	25
Italy	42
France	37
Germany	35
U. K.	28

$$\text{Standard deviation} = s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

QUESTION Find the standard deviation of vacation days (by following the numbered steps) and verify that it is the same as StDev given in the Minitab output.

SOLUTION

Step 1 ▶ Mean
From the standard deviation formula above, you can see that you will need the mean, $\bar{x}$, in order to calculate the standard deviation. Check that the mean is 30 by adding the six numbers, reporting the sum, and dividing by 6.

Step 2 ▶ Table
Fill in the table below.

x	$x - \bar{x}$	$(x - \bar{x})^2$
13	$13 - 30 = -17$	$(-17)^2 = 289$
25		
42	$42 - 30 = 12$	$12^2 = 144$
37		
35		
28		

Step 3 ▶ Sum of Squares
Add the numbers in the last column of your table to get $\sum(x - \bar{x})^2$

Step 4 ▶ Variance
Divide your answer to step 3 by $n - 1$, which is $6 - 1$, or 5, to get the following:

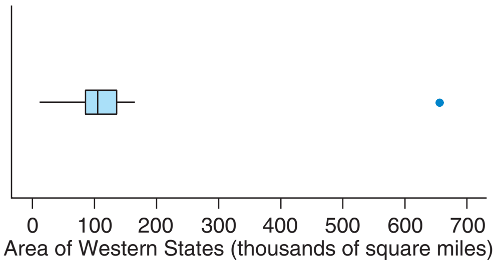
$$\frac{\sum(x-\bar{x})^2}{n-1}$$

Step 5 ▶ Standard Deviation
Finally, to get the standard deviation, take the square root of your answer to step 4. Show a long version of your answer, and then round it to hundredths and compare it with the StDev in the Minitab output.

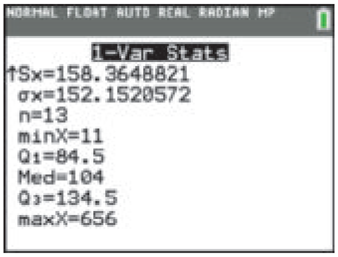
Making a Boxplot of the Area of Western States

State	Area
Alaska	656
Arizona	114
California	164
Colorado	104
Hawaii	11
Idaho	84
Montana	147
Nevada	111
New Mexico	122
Oregon	98
Utah	85
Washington	71
Wyoming	98

The area of western states is given in the table (in thousands of square miles). Verify that the boxplot given in Figure A is correct by following the numbered steps. (Source: *2012 World Almanac and Book of Facts*)



▲ **FIGURE A** Minitab Output



▲ **FIGURE B** TI-84 Output

QUESTION Make a boxplot by hand, using graph paper, a ruler, and a pencil, by following the numbered steps. You may use the TI-84 output reported in Figure B.

SOLUTION

Step 1 ▶ The Axis
Draw a horizontal line (using a ruler or other straight edge) across the page, and label it in equal intervals up to a bit above the highest area (700), and down to a bit below the lowest area (0). Below the numbers, add a label telling what the numbers are.

Step 2 ▶ The Box
Draw vertical lines at the positions of Q1 and Q3, and join them with horizontal lines to make a box. The height of the box is arbitrary.

Step 3 ▶ The Median
Put a vertical line representing the median in the correct position inside the boxplot. Which state is at the median?

Step 4 ▶ The Interquartile Range
Find and report the interquartile range, Q3 − Q1.

Step 5 ▶ Lower Limit and Potential Lower Outliers
Find and report the lower limit by finding

$$Q1 - 1.5\ IQR$$

If there are any points lower than that limit, make separate marks showing they are potential outliers.

Step 6 ▶ Lower Whisker
Draw the lower whisker (horizontal line) from the box to the lowest point that is not a potential outlier. Which state is at the left end of the lower whisker?

Step 7 ▶ Upper Limit and Potential Upper Outliers
Now find and report the upper limit by finding

$$Q3 + 1.5\ IQR$$

If there are any points higher than that, make separate marks showing they are potential outliers. Report the name of the state that is a potential outlier.

Step 8 ▶ Upper Whisker
Draw the upper whisker to the highest point that is not a potential outlier. Report this point and report the name of the state that the point represents.

TechTips

Example

Analyze the data given by finding descriptive statistics and making boxplots. The tables give calories per ounce for sliced ham and turkey. Table 3A shows unstacked data, and Table 3B shows stacked data. We coded the meat types with numerical values (1 for ham and 2 for turkey), but you could also use descriptive terms, such as “ham” and “turkey.”

Ham	Turkey
21	35
25	25
35	25
35	25
25	25
30	25
30	29
35	29
40	23
30	50
	25

▲ TABLE 3A

Cal	Meat
21	1
25	1
35	1
35	1
25	1
30	1
30	1
35	1
40	1
30	1
35	2
25	2
25	2
25	2
25	2
25	2
29	2
29	2
23	2
50	2
25	2

▲ TABLE 3B

TI-84

Enter the unstacked data (Table 3A) into **L1** and **L2**.

For Descriptive Comparisons of Two Groups

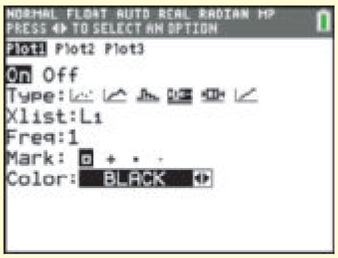
Follow the steps twice, first for **L1** and then for **L2**.

Finding One-Variable Statistics

1. Press **STAT**, choose **CALC** (by using the right arrow on the keypad), and choose **1** (for **1-Var Stats**).
2. Specify **L1** (or the list containing the data) by pressing **2ND**, **1**, and **ENTER**. Then press **ENTER**, **ENTER**.
3. Output: On your calculator, you will need to scroll down using the down arrow on the keypad to see all of the output.

Making Boxplots

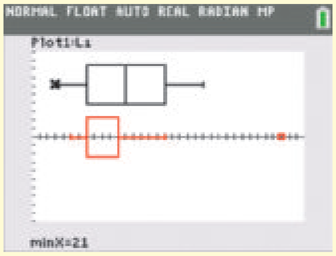
1. Press **2ND**, **STATPLOT**, **4 (PlotsOff)**, and **ENTER** to turn the plots off. This will prevent you from seeing old plots as well as the new ones.
2. Press **2ND**, **STATPLOT**, and **1**.
3. Refer to Figure 3A. Turn on **Plot1** by pressing **ENTER** when **On** is flashing. (**Off** will no longer be highlighted.)



▲ **FIGURE 3A** TI-84 Plot Selection Screen

4. Use the arrows on the keypad to locate the boxplot with outliers (as shown highlighted in Figure 3A), and press **ENTER**. (If you accidentally choose the other boxplot, there will never be any separate marks for potential outliers.)

5. Use the down arrow on the keypad to get to the **XList**. Choose **L$_1$** by pressing **2ND** and **1**.
6. Press **GRAPH**, **ZOOM**, and **9** (**Zoomstat**) to make the graph.
7. Press **TRACE** and move around with the arrows on the keypad to see the numerical labels.



▲ **FIGURE 3B** TI-84 Boxplots

Making Side-by-Side Boxplots

For side-by-side boxplots, turn on a second boxplot (**Plot2**) for data in a separate list, such as **L2**. Then, when you choose **GRAPH**, **ZOOM**, and **9**, you should see both boxplots. Press **TRACE** to see numbers.

Figure 3B shows side-by-side boxplots with the boxplot from the turkey data on the bottom and the boxplot for the ham data on the top.

MINITAB

For Comparisons of Two Groups

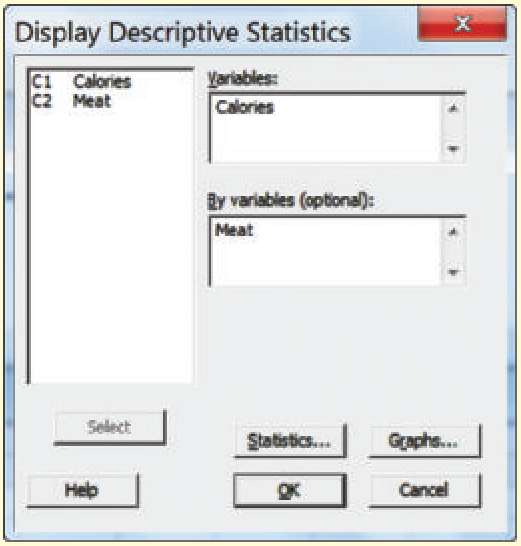
Input the data from this text's website or manually enter the data given. You may use unstacked data entered into two different columns (Table 3A), or you may use stacked data and use descriptive labels such as Ham and Turkey or codes such as 1 and 2 (Table 3B).

Finding Descriptive Statistics: One-Column Data or *Unstacked* Data in Two or More Columns

1. **Stat** > **Basic Statistics** > **Display Descriptive Statistics**.
2. Double click on the column(s) containing the data, such as **Ham** and **Turkey**, to put it (them) in the **Variables** box.
3. Ignore the **By variables (optional)** box.
4. Click on **Statistics**; you can choose what you want to add, such as the interquartile range, then click **OK**.
5. Click **OK**.

Finding Descriptive Statistics: *Stacked and Coded*

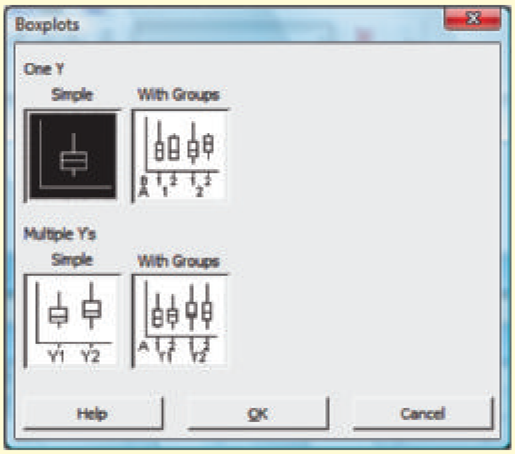
1. **Stat** > **Basic Statistics** > **Display Descriptive Statistics**.
2. See Figure 3C: Double click on the column(s) containing the stack of data, **Calories**, to put it in the **Variables** box.
3. When the **By variables (optional)** box is activated (by clicking in it), double click the column containing the categorical labels, here **Meat**.
4. Click on **Statistics**; you can choose what you want to add, such as the interquartile range. You can also make boxpots by clicking **Graphs**. Click **OK**.
5. Click **OK**.



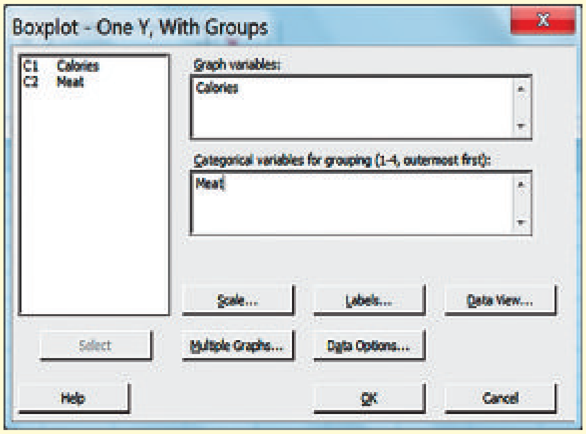
▲ **FIGURE 3C** Minitab Descriptive Statistics Input Screen

Making Boxplots

1. **Graph** > **Boxplot**.
2. For a single boxplot, choose **One Y**, **Simple**, and click **OK**. See Figure 3D.
3. Double click the label for the column containing the data, **Ham** or **Turkey**, and click **OK**.
4. For side-by-side boxplots
   a. If the data are unstacked, choose **Multiple Y's, Simple**, shown in Figure 3D. Then double click both labels for the columns and click **OK**.
   b. If the data are stacked, choose **One Y**, **With Groups** (the top right in Figure 3D) and click **OK**. Then see Figure 3E. Double click on the label for the data stack (such as **Calories**), then click in the **Categorical variables …** box



▲ **FIGURE 3D** Minitab Boxplot Selection Screen

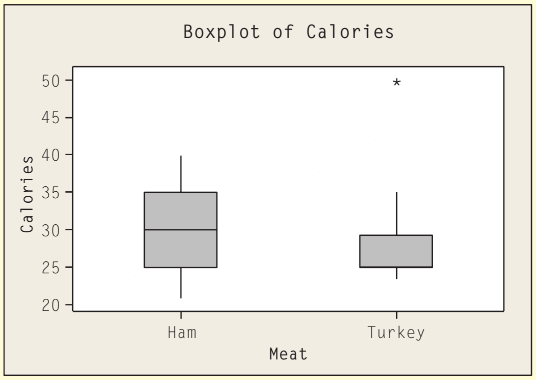


▲ **FIGURE 3E** Minitab Input Screen for Boxplots: One Y, with Groups

and double click the label for codes or words defining the groups, such as **Meat**.

5. Labeling and transposing the boxplots. If you want to change the labeling, double click on what you want to change after the boxplot(s) are made. To change the orientation of the boxplot to horizontal, double click on the *x*-axis and select **Transpose value and category scales**.

Figure 3F Shows Minitab boxplots of the ham and turkey data, without transposition.



▲ **FIGURE 3F** Minitab Boxplots

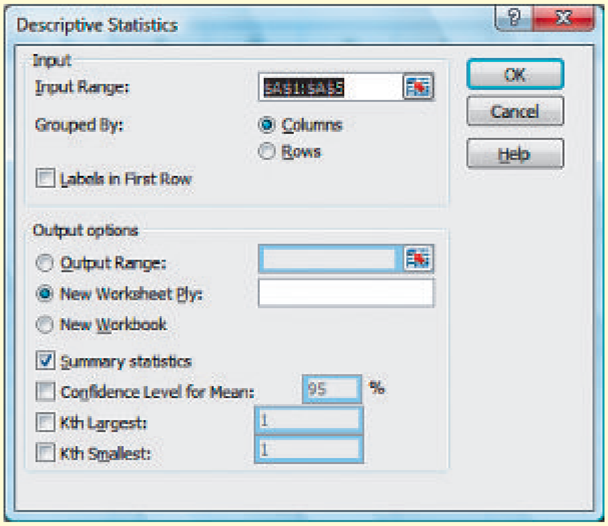
EXCEL

Entering Data

Input the unstacked data from this text's website or enter them manually. You may have labels in the first row such as **Ham** and **Turkey**.

Finding Descriptive Statistics

1. Click **DATA**, **Data Analysis,** and **Descriptive Statistics**.
2. See Figure 3G: In the dialogue screen, for the **Input Range** highlight the cells containing the data (one column only) and then Click **Summary Statistics**. If you include the label in the top cell such as A1 in the **Input Range**, you need to check **Labels in First Row**. Click **OK**.



▲ **FIGURE 3G** Excel Input Screen for Descriptive Statistics

Comparing Two Groups

To compare two groups, use unstacked data and do the above analysis twice. (Choosing **Output Range** and selecting appropriate cells on the same sheet makes comparisons easier.)

Boxplots (Requires XLSTAT Add-in)

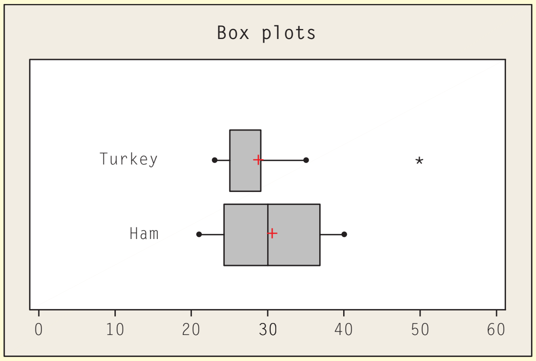
1. Click **ADD-INS**, **XLSTAT**, **Visualizing data**, and **Univariate plots.**
2. When the box under **Quantitative data** is activated, drag your cursor over the column containing the data, including the label at the top such as **Ham**.
3. Click **Charts(1).**
4. Click **Box plots, Outliers** and choose **Horizontal** (or **Vertical**).
5. Click **OK** and **Continue**. See step 6 in the side-by-side instructions.

Side-by-side Boxplots

Use unstacked data with labels in the top row.

1. Click **ADD-INS**, **XLSTAT**, **Visualizing data**, and **Univariate plots**.
2. When the box under **Quantitative data** is activated, drag your cursor over all the columns containing the unstacked data, including labels at the top such as **Ham** and **Turkey**.
3. Click **Charts(1)**
4. Click **Box plots, Group plots, Outliers,** and choose **Horizontal** (or **Vertical**).
5. Click **OK** and **Continue**.
6. When you see the small labels **Turkey** and **Ham**, you may drag them to where you want them, and you can increase the font size.

Figure 3H shows boxplots for the ham and turkey data. The red crosses give the locations of the two means.



▲ **FIGURE 3H** XLSTAT Boxplots

STATCRUNCH

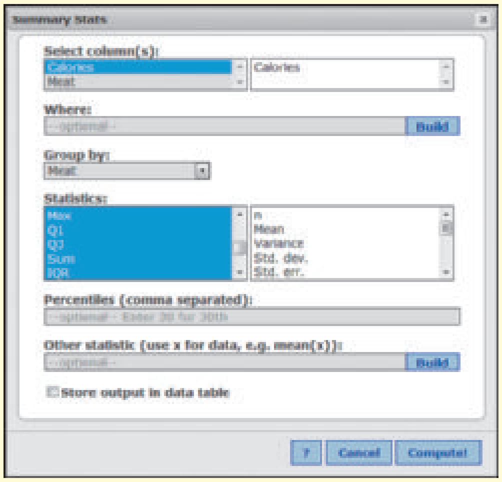
For Comparisons of Two Groups

Input the data from this text's website or enter the data manually. You may use stacked or unstacked data.

Finding Summary Statistics (*stacked* or *unstacked* data)

1. **Stat > Summary Stats > Columns**
2. Refer to Figure 3I. If the data are *stacked* put the stack (here, **Calories**) in the big box on the right and put the code (here, **Meat**) in the rectangle labeled **Group by**.

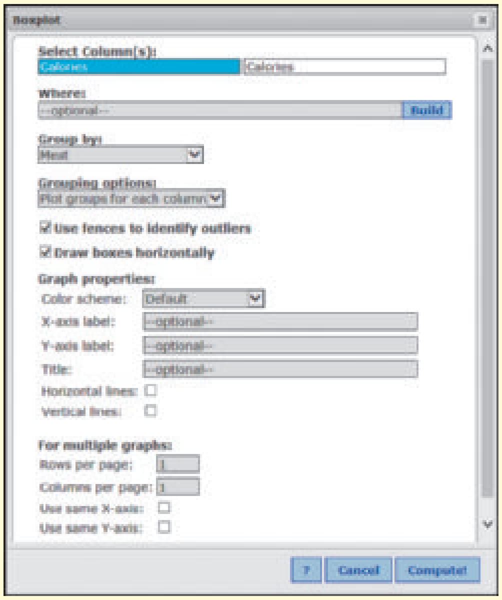
   If the data are *unstacked* put both lists into the large box on the right.
3. To include IQR in the output, go to the **Statistics:** box, click on **n** and drag down to **IQR**.
4. Click **Compute!** to get the summary statistics.



▲ **FIGURE 3I** StatCrunch Input Screen for Summary Statistics

Making Boxplots (*stacked* or *unstacked data*)

1. **Graph > Boxplot**
2. *Unstacked* data: Select the columns to be displayed. A boxplot for each column will be included in a single graph.
3. *Stacked* and coded data: Refer to Figure 3J:

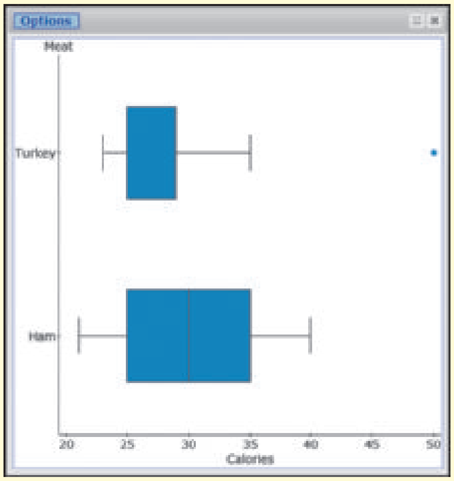


▲ **FIGURE 3J** StatCrunch Input Screen for Boxplots

Put the one column with all the data into the upper box, and then select the column of codes or categories to put in the **Group by:** small rectangle.

4. Check **Use fences to identify outliers** to make sure that the outliers show up as separate marks. You may also check **Draw boxes horizontally** if that is what you want.
5. Click **Compute!**
6. To copy your graph, click **Options** and **Copy**, and paste it into a document.

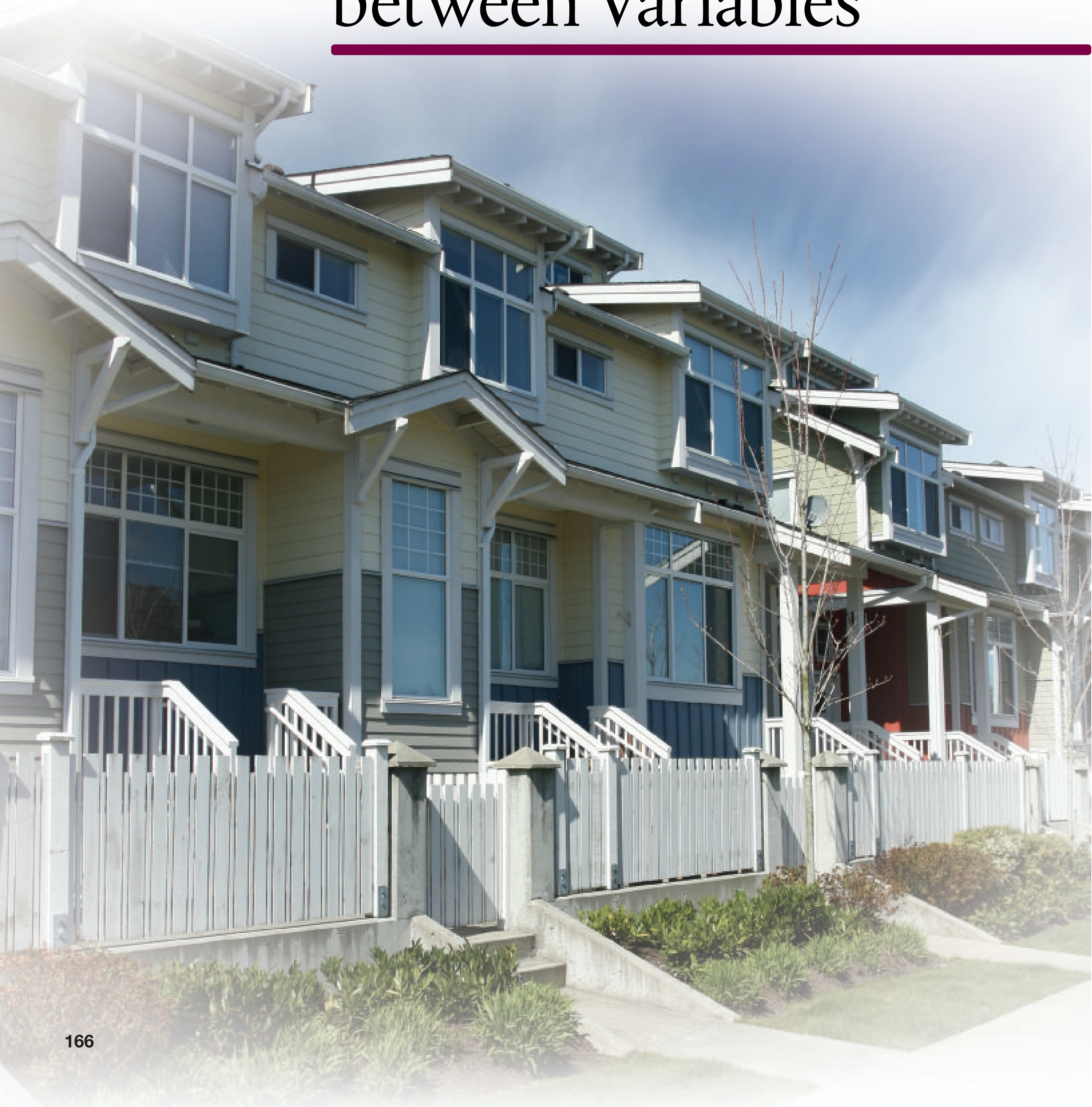
Figure 3K shows boxplots of the ham and turkey data; the top box comes from the turkey data (calories per ounce).



▲ **FIGURE 3K** StatCrunch Boxplots

4

Regression Analysis: Exploring Associations between Variables



THEME

Relationships between two numerical variables can be described in several ways, but, as always, the first step is to understand what is happening visually. When it fits the data, the linear model can help us understand how average values of one variable vary with respect to another variable, and we can use this knowledge to make predictions about one of the variables on the basis of the other.

The website www.Zillow.com can estimate the market value of any home. You need merely type in an address. Zillow estimates the value of a home even if the home has not been on the market for many years. How can this site come up with an estimate for something that is not for sale? The answer is that Zillow takes advantage of associations between the value and other easily observed variables—size of the home, selling price of nearby homes, number of bedrooms, and so on.

What role does genetics play in determining basic physical characteristics, such as height? This question fascinated nineteenth-century statistician Francis Galton (1822–1911). He examined the heights of thousands of father-son pairs to determine the nature of the relationship between these heights. If a father is 6 inches taller than average, how much taller than average will his son be? How certain can we be of the answer? Will there be much variability? If there's a lot of variability, then perhaps factors other than the father's genetic material play a role in determining height.

Associations between variables can be used to predict as-yet-unseen observations. You might think that estimating the value of a piece of real estate and understanding the role of genetics in determining height are unrelated. However, both take advantage of associations between two numerical variables. They use a technique called regression, invented by Galton, to analyze these associations.

As in previous chapters, graphs play a major role in revealing patterns in data, and graphs become even more important when we have two variables, not just one. For this reason, we'll start by using graphs to visualize associations between two numerical variables, and then we'll talk about quantifying these relationships.

CASE STUDY

Catching Meter Thieves

Parking meters are an important source of revenue for many cities. Collecting the money from these meters is no small task, particularly in a large city. In the 1970s, New York City collected the money from its meters using several different private contractors and also some of its own employees. In 1978, city officials became suspicious that employees from one of the private contractors, Brink's Inc., were stealing some of the money. Several employees were later convicted of this theft, and the city also wanted its money back, so it sued Brink's.

But how could the city tell how much money had been stolen? Fortunately, the city had collected data. For each month, it knew how much money its own employees had collected from parking meters and how much the private contractors (excluding Brink's) had collected. If there was a relationship between how much the city collected and how much the honest private contractors had collected, then that information could be used to predict about how much the city should have received from Brink's. City officials could then compare the predicted amount with the amount they actually collected from Brink's. (*Source*: De Groot et al. 1986)

At the end of this chapter, you'll see how a technique called linear regression can take advantage of patterns in the data to estimate how much money had been stolen.

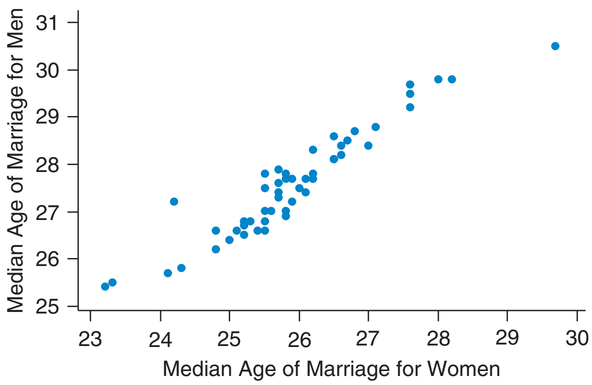
SECTION 4.1

Visualizing Variability with a Scatterplot

At what age do men and women first marry? How does this vary among the 50 states in the United States? These are questions about the relationship between two variables: age at marriage for women, and age at marriage for men. The primary tool for examining two-variable relationships, when both variables are numerical, is the **scatterplot**. In a scatterplot, each point represents one observation. The location of the point depends on the values of the two variables. For example, we might expect that states where men marry later would also have women marrying at a later age. Figure 4.1 shows a scatterplot of these data, culled from U.S. Census data. Each point represents a state (and one point represents Washington, D.C.) and shows us the typical age at which men and women marry in that state. The two points in the lower left corner represent Idaho and Utah, where the typical woman first marries at about 23 years of age, the typical man at about age 25. The point in the upper right corner represents Washington, D.C.

▶ **FIGURE 4.1** A scatterplot of typical marrying ages for men and women in the United States. The points represent the 50 states and the District of Columbia.

Tech



! Caution

About the Lower Left Corner
In scatterplots we do not require that the lower left corner be (0, 0). The reason is that we want to zoom in on the data and not show a substantial amount of empty space.

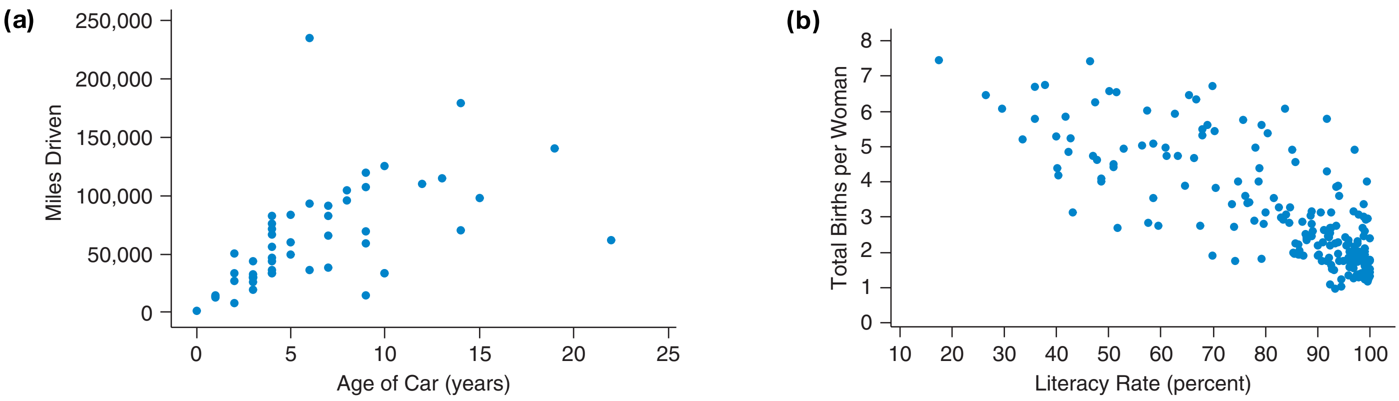
When examining histograms (or other pictures of distributions for a single variable), we look for center, spread, and shape. When studying scatterplots, we look for **trend** (which is like center), **strength** (which is like spread), and **shape** (which is like, well, shape). Let's take a closer look at these characteristics.

Recognizing Trend

The trend of an association is the general tendency of the scatterplot as you scan from left to right. Usually trends are either increasing (uphill, /) or decreasing (downhill, \), but other possibilities exist. Increasing trends are called **positive associations** (or **positive trends**), and decreasing trends are **negative associations** (or **negative trends**).

Figure 4.2 shows examples of positive and negative trends. Figure 4.2a reveals a positive trend between the age of a used car and the miles it was driven (mileage). The positive trend matches our common sense: We expect older cars to have been driven farther, because generally, the longer a car is owned, the more miles it travels. Figure 4.2b shows a negative trend—the birthrate of a country against that country's literacy rate. The negative trend suggests that countries with higher literacy rates tend to have a lower rate of childbirth.

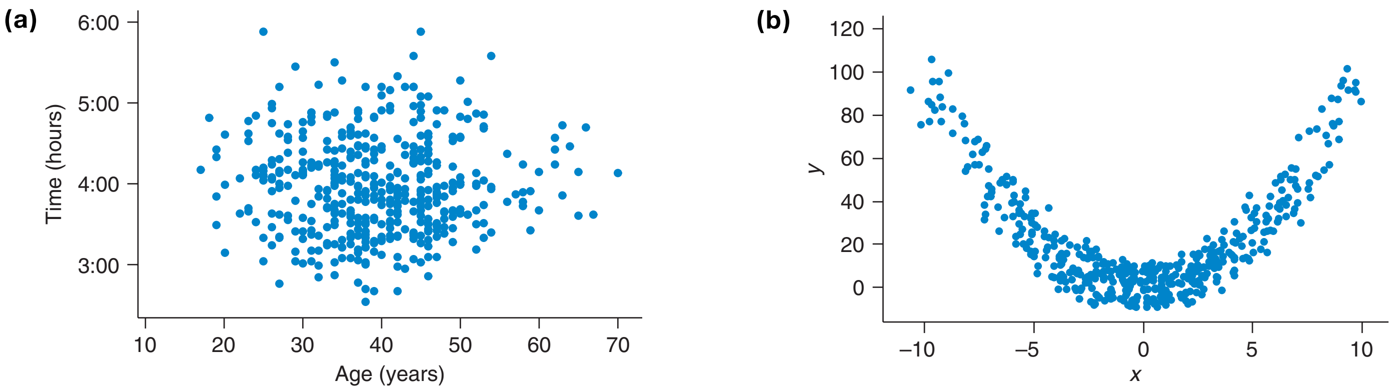
Sometimes, the absence of a trend can be interesting. For example, running a marathon requires considerable training and endurance, and we might expect that the speed at which a person runs a marathon would be related to his or her age. But Figure 4.3a shows no trend at all between the ages of runners in a marathon (in Forest



▲ **FIGURE 4.2** Scatterplots with **(a)** positive trend and **(b)** negative trend. (Sources: (a) the authors; (b) United Nations [UN], Statistics Division, http://unstats.un.org)

City, Canada) and their times. The lack of a trend means that no matter what age group we examine, the runners have about the same times as any other age group, and so we conclude that at least for this group of elite runners, age is not associated with running speed in this marathon.

Figure 4.3b shows simulated data reflecting an association between two variables that cannot be easily characterized as positive or negative—for smaller *x*-values the trend is negative (\), and for larger *x*-values it is positive (/).



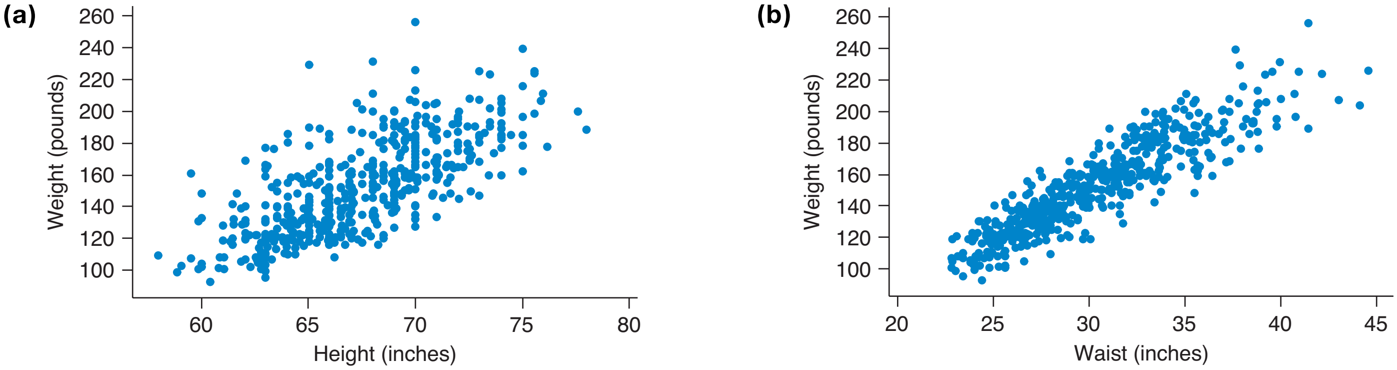
▲ **FIGURE 4.3** Scatterplots with **(a)** no trend and **(b)** a changing trend. (Sources: (a) Forest City Marathon, http://www.forestcityroadraces.com/; (b) simulated data)

Seeing Strength of Association

Weak associations result in a large amount of scatter in the scatterplot. A large amount of scatter means that points have a great deal of spread in the vertical direction. This vertical spread makes it somewhat harder to detect a trend. Strong associations have little vertical variation.

Figure 4.4 on the next page enables us to compare the strengths of two associations. Figure 4.4a shows the association between height and weight for a sample of active adults. Figure 4.4b involves the same group of adults, but this time we examine the association between waist size and weight. Which association is stronger?

The association between waist size and weight is the stronger one (Figure 4.4b). To see this, in Figure 4.4a, consider the data for people who are 65 inches tall. Their weights vary anywhere from about 120 pounds to 230 pounds, a range of 110 pounds. If you were using height to predict weight, you could be off by quite a bit. Compare



▲ **FIGURE 4.4** Part **(a)** This graph shows a relatively weaker association. **(b)** This graph shows a stronger association, because the points have less vertical spread. (Heinz et al. 2003)

this with the data in Figure 4.4b for people with a waist size of 30 inches. Their weights vary from about 120 pounds to 160 pounds, only a 40-pound range. The association between waist size and weight is stronger than that between height and weight because there is less vertical spread, so tighter predictions can be made. If you had to guess someone's weight, and could ask only one question before guessing, you'd do a better job if you asked about the person's waist size than if you asked about his or her height.

Labeling a trend as strong, very strong, or weak is a subjective judgment. Different statisticians might have different opinions. Later in this section, we'll see how we can measure strength with a number.

Identifying Shape

The simplest shape for a trend is **linear**. Fortunately, linear trends are quite common in nature. Linear trends always increase (or decrease) at the same rate. They are called linear because the trend can be summarized with a straight line. Scatterplots of linear trends often look roughly football-shaped, as shown in Figure 4.4a, particularly if there is some scatter and there are a large number of observations. Figure 4.5 shows a linear trend from data provided by Google Trends. Each point represents a week between January 2006 and December 2009. The numbers measure the volume of Google searches for the term *vampire* or *zombie*. Figure 4.5 shows that a positive, linear association exists between the volume of searches for *vampire* and the volume of searches for *zombie*. We've added a straight line to the scatterplot to highlight the linear trend.

Not all trends are linear; in fact, a great variety of shapes can occur. But don't worry about that for now: All we want to do is classify trends as either linear or not linear.

▶ **FIGURE 4.5** A line has been inserted to emphasize the linear trend.

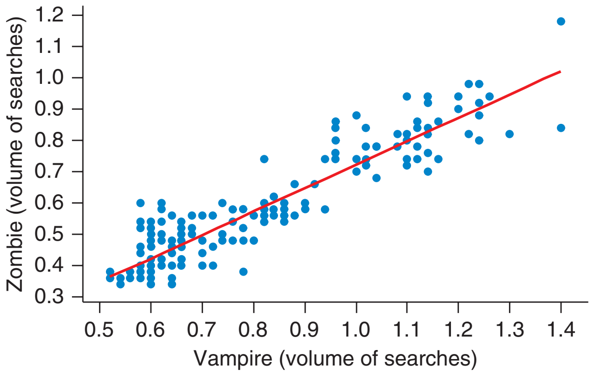
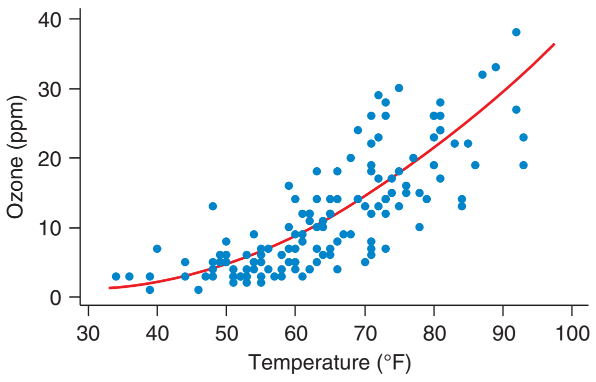


Figure 4.6 shows the relationship between levels of the pollutant ozone in the air (measured in parts per million, or ppm) and air temperature (degrees Fahrenheit) in Upland, California, near Los Angeles, over the course of a year. The trend is fairly flat at first and then becomes steeper. For temperatures less than about 55 degrees, ozone levels do not vary all that much. However, higher temperatures (above 55 degrees) are associated with much greater ozone levels. The curved line superimposed on the graph shows the nonlinear trend.



◀ **FIGURE 4.6** Ozone (ppm) is associated with temperature (degrees Fahrenheit) in a nonlinear way. (Source: Breiman, L. From the R earth package. See Faraway 2005.)

Nonlinear trends are more difficult to summarize than linear trends. This text does not cover nonlinear trends. Although our focus is on linear trends, it is very important that you first examine a scatterplot to be sure that the trend is linear. If you apply the techniques in this chapter to a nonlinear trend, you might reach disastrously incorrect conclusions!

KEY POINT When examining associations, look for the trend, the strength of the trend, and the shape of the trend.

Writing Clear Descriptions of Associations

Good communication skills are vital for success in general, and preparing you to clearly describe patterns in data is an important goal of this text. Here are some tips to help you describe two-variable associations.

- A written description should always include (1) trend, (2) shape, and (3) strength (not necessarily in that order) and should explain what all of these mean in (4) the context of the data. You should also mention any observations that do not fit the general trend.

Example 1 demonstrates how to write a clear, precise description of an association between numerical variables.

EXAMPLE 1 Age and Mileage of Used Cars

Figure 4.2a on page 169 displays an association between the age and mileage of a sample of used cars.

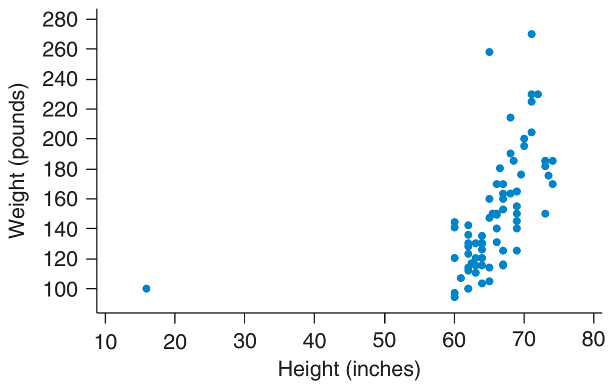
QUESTION Describe the association.

SOLUTION The association between the age and mileage of used cars is positive and linear. This means that older cars tend to have greater mileage. The association is moderately strong; some scatter is present, but not enough to hide the shape of the relationship. There is one exceptional point: One car is only about 6 years old but has been driven many miles.

TRY THIS! Exercise 4.5

The description in Example 1 is good because it mentions trend (a "positive" association), shape ("linear"), and strength ("moderately strong") and does so in context ("older cars tend to have greater mileage").

- It is very important that your descriptions be precise. For example, it would be wrong to say that older cars have greater mileage. This statement is not true of every car in the data set. The one exceptional car (upper left corner of the plot) is relatively new (about 6 years old) but has a very high mileage (about 250,000 miles). Some older cars have relatively few miles on them. To be precise, you could say older cars *tend* to have higher mileage. The word *tend* indicates that you are describing a trend that has variability, so the trend you describe is not true of all individuals but instead is a characteristic of the entire group.
- When writing a description of a relationship, you should also mention unusual features, such as outliers, small clusters of points, or anything else that does not seem to be part of the general pattern. Figure 4.7 includes an outlier. These data are from a statistics class in which students reported their weights and heights. One student wrote the wrong height.



▶ **FIGURE 4.7** A fairly strong, positive association between height and weight for a statistics class. One student reported the wrong height. (Source: R. Gould, UCLA, Department of Statistics)

SECTION 4.2

Measuring Strength of Association with Correlation

The **correlation coefficient** is a number that measures the strength of the linear association between two numerical variables—for example, the relationship between people's heights and weights. We can't emphasize enough that the correlation coefficient *makes sense only if the trend is linear and if both variables are numerical.*

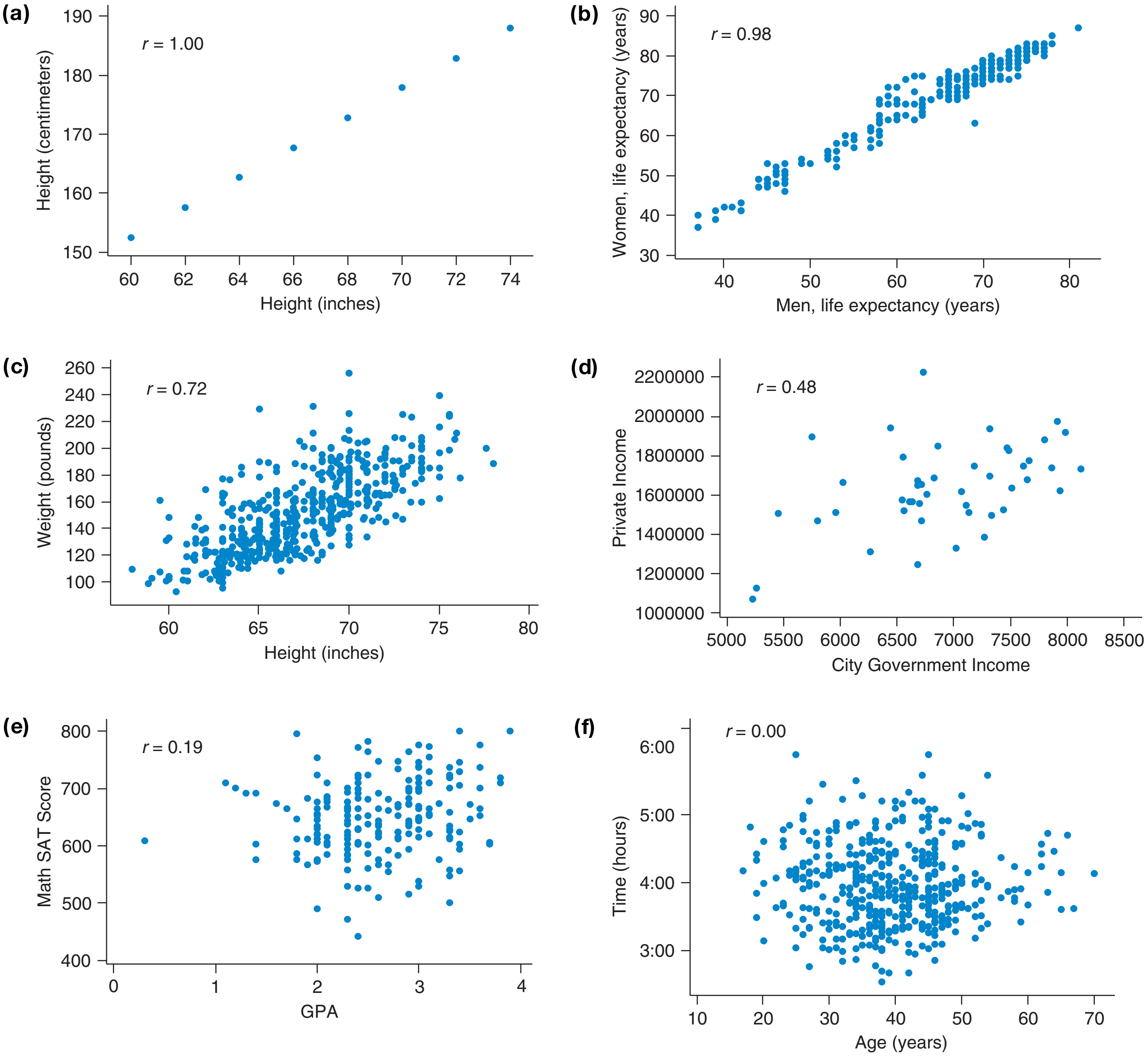
The correlation coefficient, represented by the letter r, is always a number between -1 and $+1$. Both the value and the sign (positive or negative) of r have information we can use. If the value of r is close to -1 or $+1$, then the association is very strong; if r is close to 0, the association is weak. If the value of the correlation coefficient is positive, then the trend is positive; if the value is negative, the trend is negative.

Visualizing the Correlation Coefficient

Figure 4.8 presents a series of scatterplots that show associations of gradually decreasing strength. The strongest linear association appears in Figure 4.8a; the points fall exactly along a line. Because the trend is positive and perfectly linear, the correlation coefficient is equal to 1.

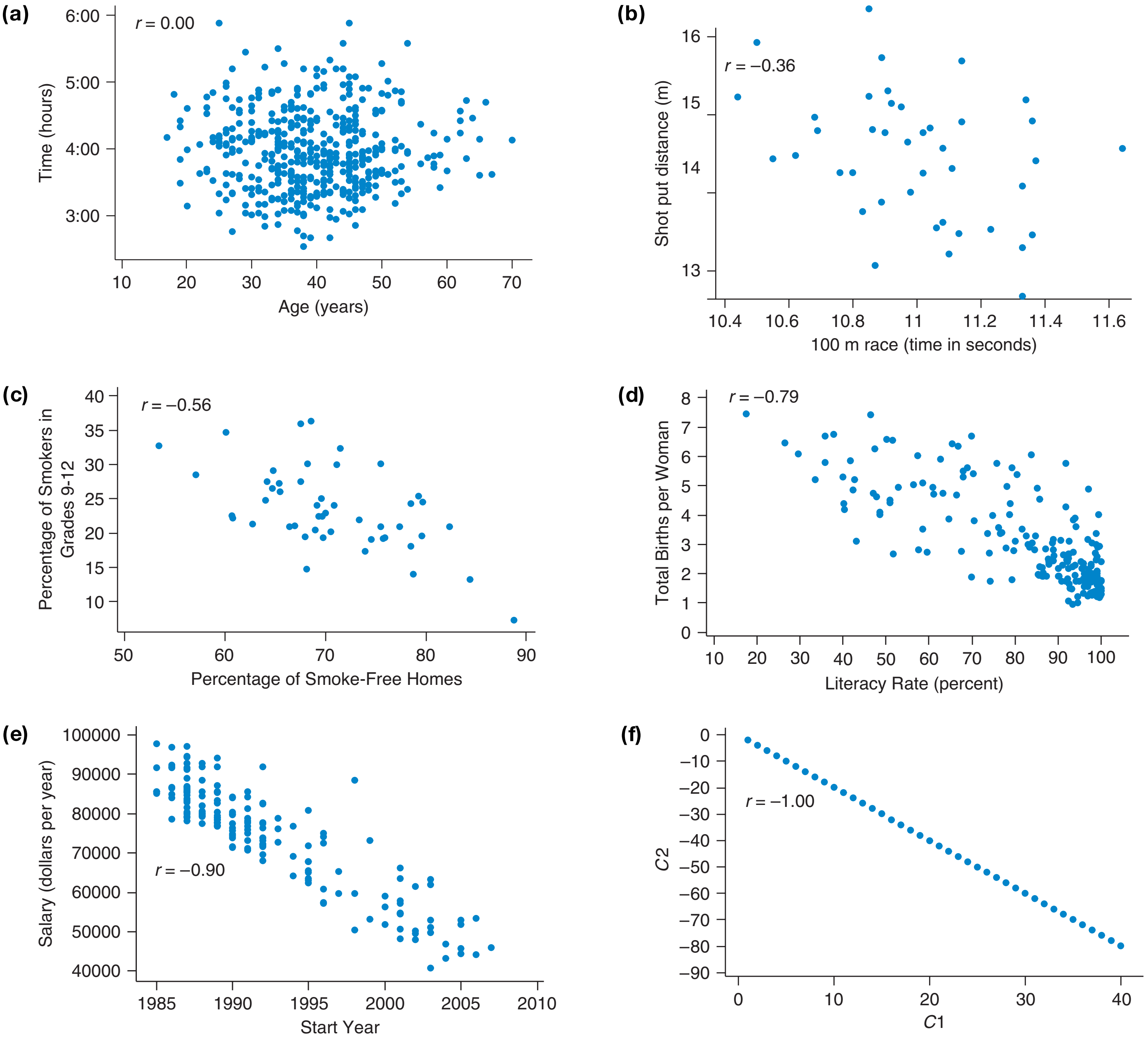
The next scatterplot, Figure 4.8b, shows a slightly weaker association. The points are more spread out vertically. We can see a linear trend, but the points do not fall exactly along a line. The trend is still positive, so the correlation coefficient is also positive. However, the value of the correlation coefficient is less than 1 (it is 0.98).

The remaining scatterplots show weaker and weaker associations, and their correlation coefficients gradually decrease. The last scatterplot, Figure 4.8f, shows no



▲ **FIGURE 4.8** Scatterplots with gradually decreasing positive correlation coefficients. (Sources: (a) simulated data; (b) www.overpopulation.com; (c) Heinz, 2003; (d) Bentow and Afshartous; (e) Fathom™ Sample Document, Educational Testing Service [ETS] validation study; (f) Forest City Marathon) association at all between the two variables, and the correlation coefficient has a value of 0.00.

The next set of scatterplots (Figure 4.9) starts with that same marathon data (having a correlation coefficient of 0.00), and the negative correlations gradually get stronger. The last figure has a correlation coefficient of −1.00.



▲ **FIGURE 4.9** Scatterplots with increasingly negative correlations. (Sources: (a) Forest City Marathon; (b) Data from R package FactoMineR; (c) Centers for Disease Control; (d) UN statistics; (e) Minitab Student 12 file "Salary," adjusted for inflation; (f) simulated data)

The Correlation Coefficient in Context

The correlation between the number of Google searches for *zombie* and *vampire* is $r = 0.924$. If we are told, or already know, that the association between these variables is linear, then we know that the trend is positive and strong. The fact that the correlation is close to 1 means that there is not much scatter in the scatterplot.

College admission offices sometimes report correlations between students' Scholastic Aptitude Test (SAT) scores and their first-year GPAs. If the association is linear and the correlation is high, this justifies using the SAT to make admissions decisions, because a high correlation would indicate a strong association between SAT scores and academic performance. A positive correlation means that students who score above average on the SAT tend to get above-average grades. Conversely, those who score below average on the SAT tend to get below-average grades. Note that we're careful to say "tend to." Certainly, some students with low SAT scores do very well, and some with high SAT scores struggle to pass their classes. The correlation coefficient does not tell us about individual students; it tells us about the overall trend.

Caution

Linearity Is Important
The correlation coefficient is interpretable only for linear associations.

More Context: Correlation Does Not Mean Causation!

Quite often, you'll hear someone use the correlation coefficient to support a claim of cause and effect. For example, one of the authors once read that a politician wanted to close liquor stores in a city because there was a positive correlation between the number of liquor stores in a neighborhood and the amount of crime.

As you learned in Chapter 1, we can't form cause-and-effect conclusions from observational studies. If your data came from an observational study, it doesn't matter how strong the correlation is. Even a correlation near 1 is not enough for us to conclude that changing one variable (closing down liquor stores) will lead to a change in the other variable (crime rate).

A positive correlation also exists between the number of blankets sold in Canada per week and the number of brush fires in Australia per week. Are brush fires in Australia caused by cold Canadians? Probably not. The correlation is likely to be the result of weather. When it is winter in Canada, people buy blankets. When winter is happening in Canada, summer is happening in Australia (which is located in the Southern Hemisphere), and summer is brush-fire season.

What, then, can we conclude from the fact that the number of liquor stores in a neighborhood is positively correlated with the crime rate in that neighborhood? Only that neighborhoods with a higher-than-average number of liquor stores typically (but not always) have a higher-than-average crime rate.

If you learn nothing else from this book, remember this: No matter how tempting, do *not* conclude that a cause-and-effect relationship between two variables exists just because there is a correlation, no matter how close to $+1$ or -1 that correlation might be!

KEY POINT Correlation does not imply causation.

Finding the Correlation Coefficient

The correlation coefficient is best determined through the use of technology. We calculate a correlation coefficient by first converting each observation to a *z*-score, using the appropriate variable's mean and standard deviation. For example, to find the correlation coefficient that measures the strength of the linear relation between weight and height, we first convert each person's weight and height to *z*-scores. The next step is to multiply the observations' *z*-scores together. If both are positive or both negative—meaning that both *z*-scores are above average or both are below average—then the product is a positive number. In a strong positive association, most of these products are positive values. In a strong negative association, however, observations above average on one variable tend to be below average on the other variable. In this case, one *z*-score is negative and one positive, so the product is negative. Thus, in a strong negative association, most *z*-score products are negative.

Looking Back

***z*-Scores**
Recall that *z*-scores show how many standard deviations a measurement is from the mean. To find a *z*-score from a measurement, first subtract the mean and then divide the difference by the standard deviation.

To find the correlation coefficient, add the products of z-scores together and divide by $n - 1$ (where n is the number of observed pairs in the sample). In mathematical terms, we get

$$\textbf{Formula 4.1: } r = \frac{\sum z_x z_y}{n - 1}$$

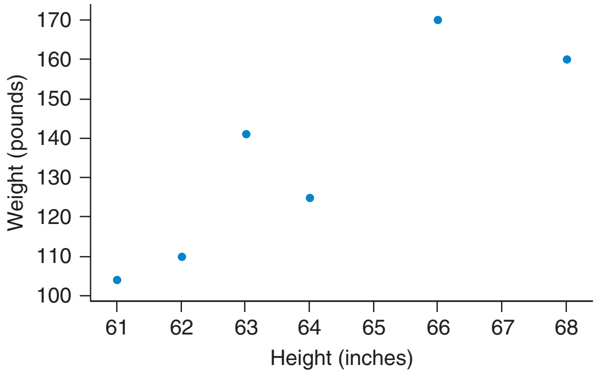
The following example illustrates how to use Formula 4.1 in a calculation.

EXAMPLE 2 Heights and Weights of Six Women

Figure 4.10a shows the scatterplot for heights and weights of six women.

QUESTION Using the data provided, find the correlation coefficient of the linear association between heights (inches) and weights (pounds) for these six women.

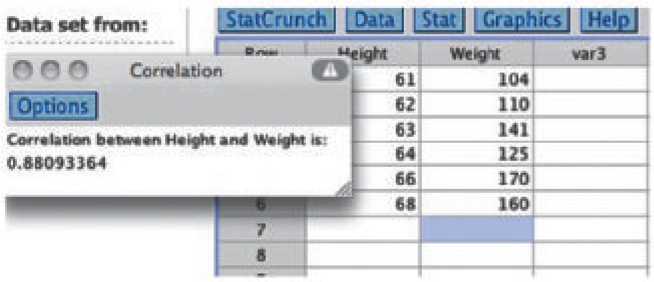
Heights	61	62	63	64	66	68
Weights	104	110	141	125	170	160



▶ **FIGURE 4.10a** Scatterplot showing heights and weights of six women.

SOLUTION Before proceeding, we verify that the conditions hold. Figure 4.10a suggests that a straight line is an acceptable model; a straight line through the data might summarize the trend, although this is hard to see with so few points.

Next, we calculate the correlation coefficient. Ordinarily, we use technology to do this, and Figure 4.10b shows the output from StatCrunch, which gives us the value $r = 0.88093364$.



▶ **FIGURE 4.10b** StatCrunch, like all statistical software, lets you calculate the correlation between any two columns of numerical data you choose.

Because the sample size is small, we confirm this output using Formula 4.1. It is helpful to go through the steps of this calculation to better understand how the correlation coefficient measures linear relationships between variables.

The first step is to calculate average values of height and weight and then determine the standard deviation for each.

$$\text{For the height: } \bar{x} = 64 \text{ and } s_x = 2.608$$

$$\text{For the weight: } \bar{y} = 135 \text{ and } s_y = 26.73$$

Next we convert all of the points to pairs of z-scores and multiply them together. For example, for the woman who is 68 inches tall and weighs 160 pounds,

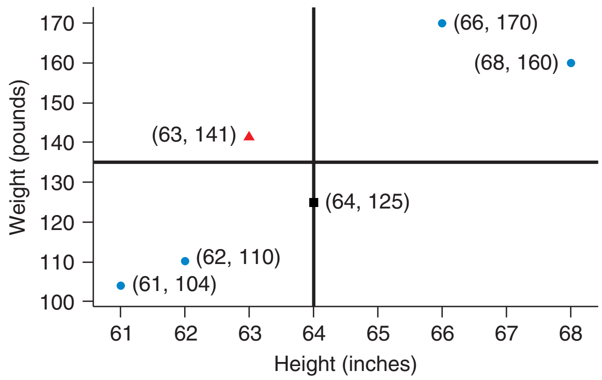
$$z_x = \frac{x - \bar{x}}{s_x} = \frac{68 - 64}{2.608} = \frac{4}{2.608} = 1.53$$

$$z_y = \frac{y - \bar{y}}{s_y} = \frac{160 - 135}{26.73} = \frac{25}{26.73} = 0.94$$

The product is

$$z_x \times z_y = 1.53 \times 0.94 = 1.44$$

Note that this product is positive and shows up in the upper right quadrant in Figure 4.10c.



◀ **FIGURE 4.10c** The same scatterplot as in 4.10a, but with the plot divided into quadrants based on average height and weight. Points represented with blue circles contribute a positive value to the correlation coefficient (positive times positive is positive, or negative times negative equals a positive). The red triangle represents an observation that contributes negatively (a negative z-score times a positive z-score is negative), and the black square contributes nothing because one of the z-scores is 0.

Figure 4.10c can help you visualize the rest of the process. The two blue circles in the upper right portion represent observations that are above average in both variables, so both z-scores are positive. The two blue circles in the lower left region represent observations that are below average in both variables; the products of the two negative z-scores are positive, so they add to the correlation. The red triangle has a positive z-score for weight (it is above average) but a negative z-score for height, so the product is negative. The black square is a point that makes no contribution to the correlation coefficient. This person is of average height, so her z-score for height is 0.

The correlation between height and weight for these six women comes out to be about 0.881.

CONCLUSION The correlation coefficient for the linear association of weights and heights of these six women is $r = 0.881$. Thus, there is a strong positive correlation between height and weight for these women. Taller women tend to weigh more.

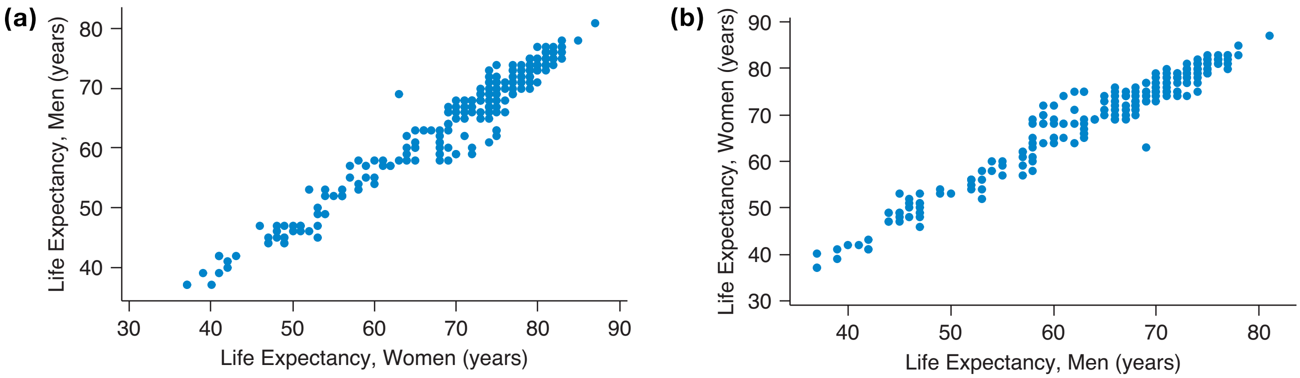
TRY THIS! Exercise 4.21a

Understanding the Correlation Coefficient

The correlation coefficient has a few features you should know about when interpreting a value of r or deciding whether you should compute the value.

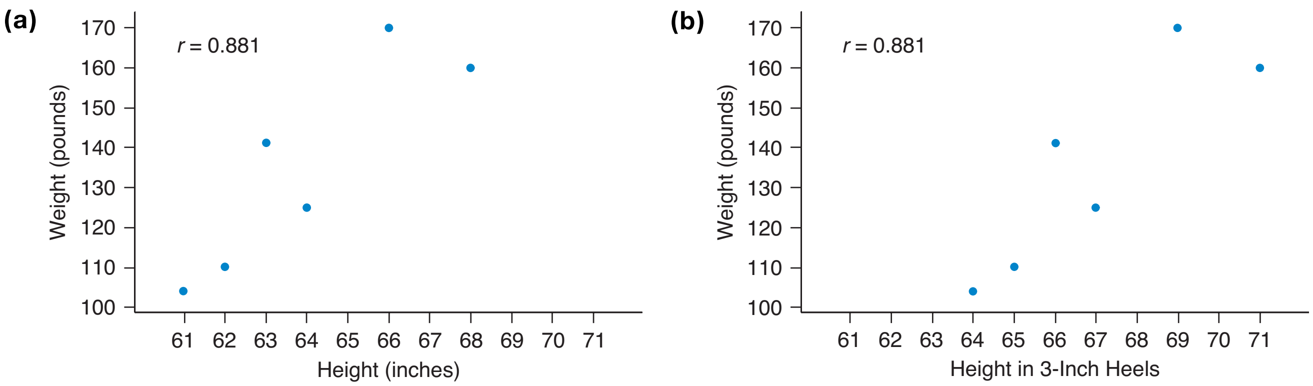
- *Changing the order of the variables does not change r.* This means that if the correlation between life expectancy for men and women is 0.977, then the correlation between life expectancy for women and men is also 0.977. This makes sense because the correlation measures the strength of the linear relationship between *x* and *y*, and that strength will be the same no matter which variable gets plotted on the horizontal axis and which on the vertical.

Figure 4.11a and b have the same correlation; we've just swapped axes.



▲ **FIGURE 4.11** Scatterplots showing the relationship between men's and women's life expectancies for various countries. **(a)** Women's life expectancy is plotted on the *x*-axis. **(b)** Men's life expectancy is plotted on the *x*-axis. (Sources: http://www.overpopulation.com and Fathom™ Sample Documents)

- *Adding a constant or multiplying by a positive constant does not affect r.* The correlation between the heights and weights of the six women in Example 2 was 0.881. What would happen if all six women in the sample had been asked to wear 3-inch platform heels when their heights were measured? Everyone would have been 3 inches taller. Would this have changed the value of *r*? Intuitively, you should sense that it wouldn't. Figure 4.12a shows a scatterplot of the original data, and Figure 4.12b shows the data with the women in 3-inch heels.



▲ **FIGURE 4.12** **(a)** A repeat of the scatterplot of height and weight for six women. **(b)** The same women in 3-inch heels. The correlation remains the same.

We haven't changed the strength of the relationship. All we've done is shift the points on the scatterplot 3 inches to the right. But shifting the points doesn't change the relationship between height and weight. We can verify that the correlation is unchanged by looking at the formula. The heights will have the same *z*-scores both before and after the women put on the shoes; since everyone "grows" by the same amount, everyone is still the same number of standard deviations away from the average, which also "grows" by 3 inches. As another example, if science found a way to add 5 years to the life expectancy of men in all countries in the world, the correlation between life expectancies for men and women would still be the same.

More generally, we can add a constant (a fixed value) to all of the values of one variable, or of both variables, and not affect the correlation coefficient.

For the very same reason, we can multiply either or both variables by positive constants without changing *r*. For example, to convert the women's heights from inches to feet, we multiply their heights by 1/12. Doing this does not change how strong the association is; it merely changes the units we're using to measure height. Because the strength of the association does not change, the correlation coefficient does not change.

- *The correlation coefficient is unitless.* Height is measured in inches and weight in pounds, but *r* has no units because the *z*-scores have no units. This means that we will get the same value for correlation whether we measure height in inches, meters, or fathoms.
- *Linear, linear, linear.* We've said it before, but we'll say it again: We're talking only about linear relationships here. The correlation can be misleading if you do not have a linear relationship. Figure 4.13a through d illustrate the fact that different nonlinear patterns can have the same correlation. All of these graphs have $r = 0.817$, but the graphs have very different shapes. The take-home message is that the correlation alone does not tell us much about the shape of a graph. We must also know that the relationship is linear to make sense of the correlation.

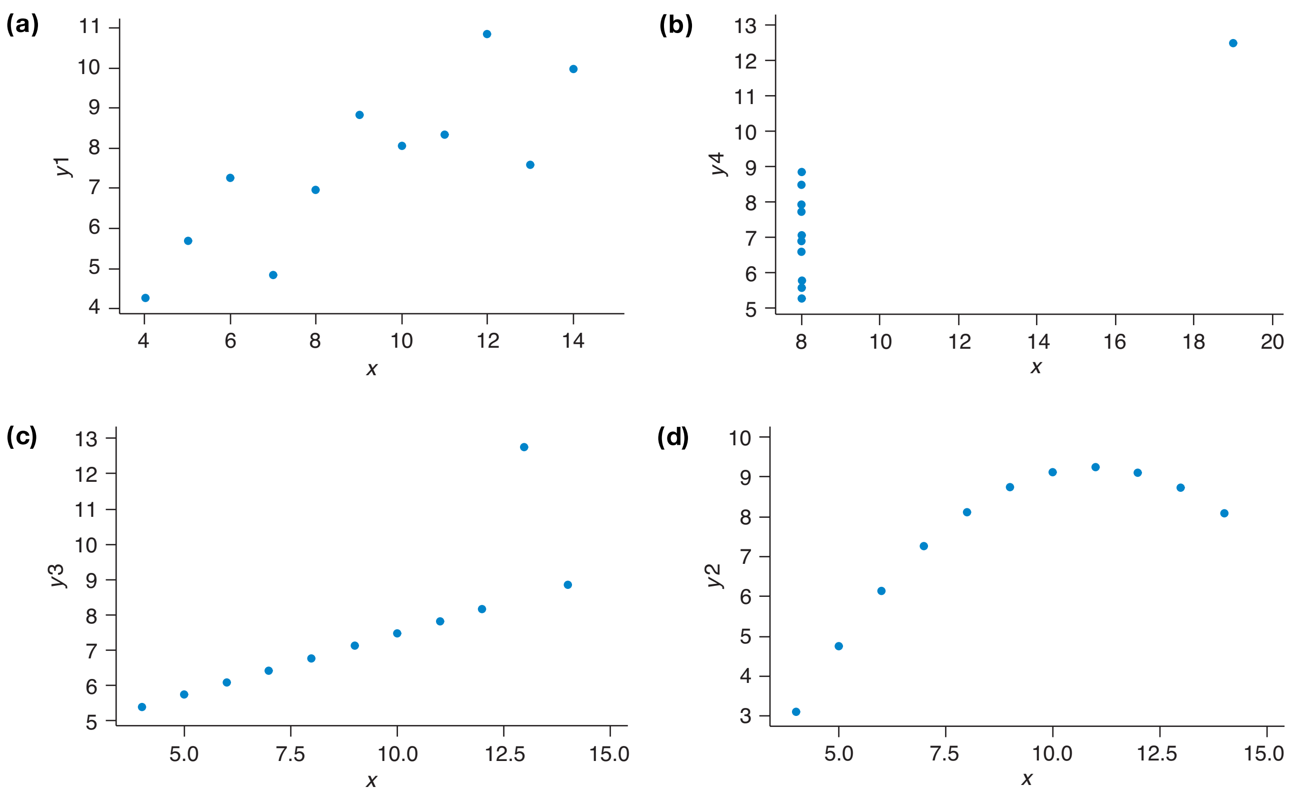
  Remember: *Always* make a graph of your data. If the trend is nonlinear, the correlation (and, as you'll see in the next section, other statistics) can be very misleading.

! Caution

Correlation Coefficient and Linearity

A value of *r* close to 1 or −1 does *not* tell you that the relationship is linear. You must check visually; otherwise, your interpretation of the correlation coefficient might be wrong.

KEY POINT The correlation coefficient does not tell you whether an association is linear. However, if you already know that the association is linear, then the correlation coefficient tells you how strong the association is.



▲ **FIGURE 4.13 (a–d)** Four scatterplots with the same correlation coefficient of 0.817 have very different shapes. The correlation coefficient is meaningful only if the trend is linear. (Source: Anscombe, F. 1973)

SNAPSHOT THE CORRELATION COEFFICIENT

WHAT IS IT? ► Correlation coefficient.

WHAT DOES IT DO? ► Measures the strength of a linear association.

HOW DOES IT DO IT? ► By comparing z-scores of the two variables. The products of the two z-scores for each point are averaged.

HOW IS IT USED? ► The sign tells us whether the trend is positive (+) or negative (−). The value tells us the strength. If the value is close to 1 or −1, then the points are tightly clustered about a line; if the value is close to 0, then there is no linear association.

Note: The correlation coefficient can be interpreted only with linear associations.

SECTION 4.3

Modeling Linear Trends

How much more do people tend to weigh for each additional inch in height? How much value do cars lose each year as they age? Are home run hitters good for their teams? Can we predict how much space a book will take on a bookshelf just by knowing how many pages are in the book? It's not enough to remark that a trend exists. To make a prediction based on data, we need to measure the trend and the strength of the trend.

To measure the trend, we're going to perform a bit of statistical sleight of hand. Rather than interpret the data themselves, we will substitute a model of the data and interpret the model. The model consists of an equation and a set of conditions that describe when the model will be appropriate. Ideally, this equation is a very concise and accurate description of the data; if so, the model is a good fit. When the model is a good fit to the data, any understanding we gain about the model accurately applies to our understanding of the real world. If the model is a bad fit, however, then our understanding of real situations might be seriously flawed.

The Regression Line

The **regression line** is a tool for making predictions about future observed values. It also provides us with a useful way of summarizing a linear relationship. Recall from Chapter 3 that we could summarize a sample distribution with a mean and a standard deviation. The regression line works the same way: It reduces a linear relationship to its bare essentials and enables us to analyze a relationship without being distracted by small details.

Review: Equation of a Line The regression line is given by an equation for a straight line. Recall from algebra that equations for straight lines contain a ***y* intercept** and a **slope**. The equation for a straight line is

$$y = mx + b$$

The letter m represents the slope, which tells how steep the line is, and the letter b represents the y-intercept, which is the value of y when $x = 0$.

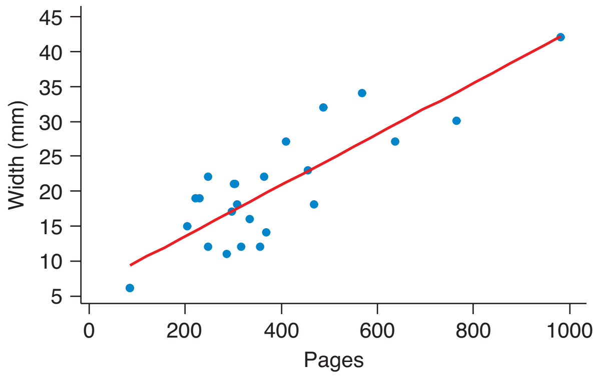
Statisticians write the equation of a line slightly differently and put the intercept first; they use the letter a for the intercept and b for the slope and write

$$y = a + bx$$

We often use the names of variables in place of *x* and *y* to emphasize that the regression line is a model about two real-world variables. We will sometimes write the word *predicted* in front of the *y*-variable to emphasize that the line consists of predictions for the *y*-variable, not actual values. A few examples should make this clear.

Visualizing the Regression Line Can we know how wide a book is on the basis of the number of pages in the book? A student took a random sample of books from his shelf, measured the width of the spine (in millimeters, mm), and recorded the number of pages. Figure 4.14 illustrates how the regression line captures the basic trend of a linear association between width of the book and the number of pages for this sample. The equation for this line is

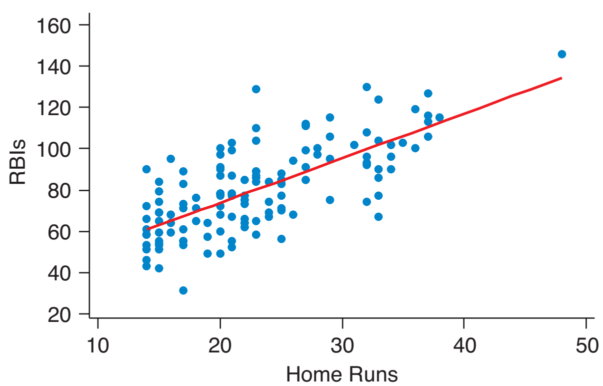
$$\text{Predicted Width} = 6.22 + 0.0366 \text{ Pages}$$



◀ **FIGURE 4.14** The regression line summarizes the relationship between the width of the book and the number of pages for a small sample of books. (Source: Onaga, E. 2005, UCLA, Department of Statistics)

In baseball, two numbers used to measure how good a batter is are the number of runs batted in (RBI) and the number of home runs. (A home run occurs when the batter rounds all three bases with one hit and scores a run. An RBI occurs when a player is already on base, and a batter hits the ball far enough for the on-base runner to score a run.) Some baseball fans believe that players who hit a lot of home runs might be exciting to watch but are not that good for the team. (Some believe that home run hitters tend to strike out more often; this takes away scoring opportunities for the team.) Figure 4.15 shows the relationship between the number of home runs and RBIs for the best home run hitters in the major leagues in the 2008 season. The association seems fairly linear, and the regression line can be used to predict how many RBIs a player will score, given the number of home runs hit. The data suggest that players who score a large number of home runs also tend to score a large number of points through RBIs.

$$\text{Predicted RBIs} = 30.46 + 2.16 \text{ Home Runs}$$



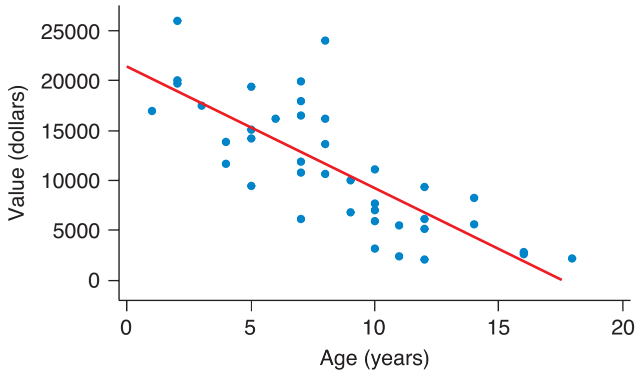
◀ **FIGURE 4.15** The regression line summarizes the relationship between RBIs and home runs for the top home run hitters in the 2008 major league baseball season. (Source: www.baseball1.com)

Regression in Context Suppose you have a 10-year-old car and want to estimate how much it is worth. One of the more important uses of the regression line is to make predictions about what y-values can be obtained for particular x-values. Figure 4.16 suggests that the relationship between age and value is linear, and the regression line that summarizes this relationship is

$$\text{Predicted Value} = 21375 - 1215\ \text{Age}$$

We can use this equation to predict approximately how much a 10-year-old car is worth:

$$\begin{aligned}\text{Predicted Value} &= 21375 - 1215 \times 10 \\ &= 27375 - 12150 \\ &= 9225\end{aligned}$$



▶ **FIGURE 4.16** The regression line summarizes the relationship between the value of a car, according to the Kelley Blue Book, and the car's age for a small sample of students' cars. (Source: C. Ryan 2006)

The regression line predicts that a 10-year-old car will be valued at about $9225. As we know, many factors other than age affect the value of a car, and perhaps with more information we might make a better prediction. However, if the only thing we know about a car is its age, this may be the best prediction we can get. It is also important to keep in mind that this sample comes from one particular group of students in one particular statistics class and is not representative of all used cars on the market in the United States.

Using the regression line to make predictions requires certain assumptions. We'll go into more detail about these assumptions later, but for now, just use common sense. This predicted value of $9225 is useful only if our data are valid. For instance, if all the cars in our data set are Toyotas, and our 10-year-old car is a Chevrolet, then the prediction is probably not useful.

EXAMPLE 3 Book Width

A college instructor with far too many books on his shelf is wondering whether he has room for one more. He has about 20 mm of space left on his shelf, and he can tell from the online bookstore that the book he wants has 598 pages. The regression line is

$$\text{Predicted Width} = 6.22 + 0.0366\ \text{Pages}$$

QUESTION Will the book fit on his shelf?

SOLUTION Assuming that the data used to fit this regression line are representative of all books, we would predict the width of the book corresponding to 598 pages to be

$$\begin{aligned}\text{Predicted Width} &= 6.22 + 0.0366 \times 598 \\ &= 6.22 + 21.8868 \\ &= 28.1068\ \text{mm}\end{aligned}$$

CONCLUSION The book is predicted to be 28 mm wide. Even though the actual book width is likely to differ somewhat from 28 mm, it seems that the book will probably not fit on the shelf.

TRY THIS! Exercise 4.29

Common sense tells us that not all books with 598 pages are exactly 28 mm wide. There is a lot of variation in the width of a book for a given number of pages.

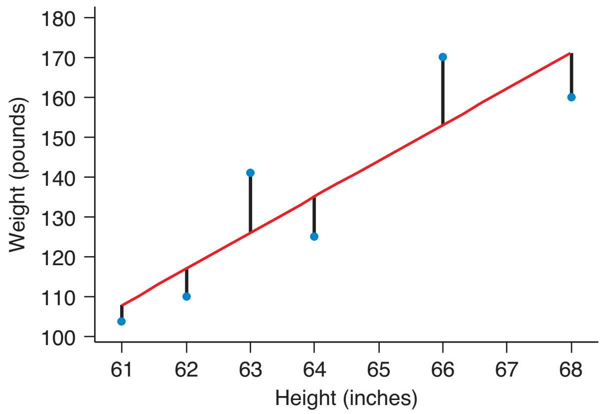
Finding the Regression Line In almost every case, we'll use technology to find the regression line. However, it is important to know how the technology works, and to be able to calculate the equation when we don't have access to the full data set.

To understand how technology finds the regression line, imagine trying to draw a line through the scatterplots in Figures 4.14 through 4.16 to best capture the linear trend. We could have drawn almost any line, and some of them would have looked like pretty good summaries of the trend. What makes the regression line special? How do we find the intercept and slope of the regression line?

The regression line is chosen because it is the line that comes closest to most of the points. More precisely, the square of the vertical distances between the points and the line, on average, is bigger for any other line we might draw than for the regression line. Figure 4.17 shows these vertical distances with black lines. The regression line is sometimes called the "best fit" line because, in this sense, it provides the best fit to the data.

Details

Least Squares
The regression line is also called the least squares line because it is chosen so that the sum of the squares of the differences between the observed y-value, y, and the value predicted by the line, $\hat{y}$, is as small as possible. Mathematically, this means that the slope and intercept are chosen so that $\Sigma(y - \hat{y})^2$ is as small as possible.



◄ **FIGURE 4.17** The "best fit" regression line, showing the vertical distance between each observation and the line. For any other line, the average of the squared distances is larger.

To find this best fit line, we need to find the slope and the intercept. The slope, b, of the regression line is the ratio of the standard deviations of the two variables, multiplied by the correlation coefficient:

$$\textbf{Formula 4.2a:}\ \text{The slope, } b = r\frac{s_y}{s_x}$$

Once we have found the slope, we can find the intercept. Finding the intercept, a, requires that we first find the means of the variables, $\bar{x}$ and $\bar{y}$:

$$\textbf{Formula 4.2b:}\ \text{The intercept, } a = \bar{y} - b\bar{x}$$

Now put these quantities into the equation of a line, and the regression line is given by

$$\textbf{Formula 4.2c:}\ \text{The regression line, Predicted } y = a + bx$$

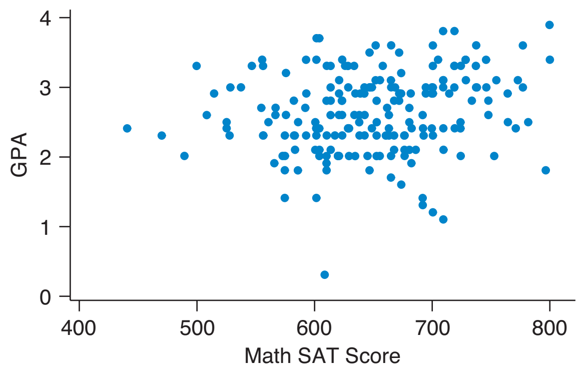
EXAMPLE 4 SAT Scores and GPAs

A college in the northeastern United States used SAT scores to help decide which applicants to admit. To determine whether the SAT was useful in predicting success, the college examined the relationship between the SAT scores and first-year GPAs of admitted students. Figure 4.18a shows a scatterplot of the math SAT scores and first-year GPAs for a random sample of 200 students. The scatterplot suggests a weak, positive linear association: Students with higher math SAT scores tend to get higher first-year GPAs. The average math SAT score of this sample was 649.5 with standard deviation 66.3. The average GPA was 2.63 with standard deviation 0.58. The correlation between GPA and math SAT score was 0.194.

QUESTIONS

a. Find the equation of the regression line that best summarizes this association. Note that the *x*-variable is the math SAT score and the *y*-variable is the GPA.

b. Using the equation, find the predicted GPA for a person in this group with a math SAT score of 650.



▶ **FIGURE 4.18a** Students with higher math SAT scores tend to have higher first-year GPAs, but this linear association is fairly weak. (Source: ETS validation study of an unnamed college and Fathom™ Sample Documents, 2007, Key Curriculum Press)

SOLUTIONS

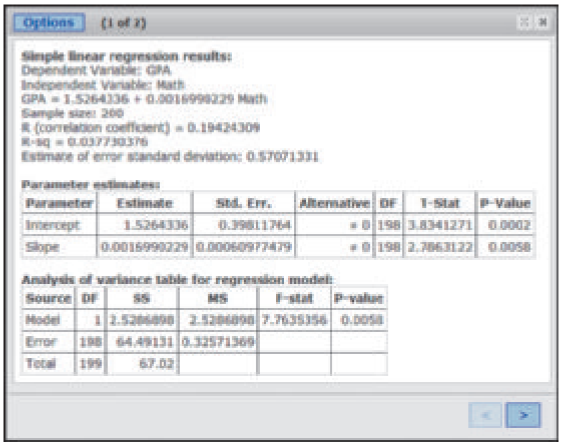
a. With access to the full data set at the text's website, we can use technology to find (and plot) the regression line. Still, when summary statistics are provided (means and standard deviations of the two variables as well as their correlation), it is not time-consuming to use Formula 4.2 to find the regression line.

Tech

Figure 4.18b shows StatCrunch output for the regression line. (StatCrunch provides quite a bit more information than we will use in this chapter.)

According to technology (and after some rounding), the equation of the regression line is

$$\text{Predicted GPA} = 1.53 + 0.0017 \text{ Math}$$



Options (1 of 2)

Simple linear regression results:
Dependent Variable: GPA
Independent Variable: Math
GPA = 1.5264336 + 0.0016990229 Math
Sample size: 200
R (correlation coefficient) = 0.19424309
R-sq = 0.037730376
Estimate of error standard deviation: 0.57071331

Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	P-Value
Intercept	1.5264336	0.39811764	≠ 0	198	3.8341271	0.0002
Slope	0.0016990229	0.00060977479	≠ 0	198	2.7863122	0.0058

Analysis of variance table for regression model:

Source	DF	SS	MS	F-stat	P-value
Model	1	2.5286898	2.5286898	7.7635356	0.0058
Error	198	64.49131	0.32571369		
Total	199	67.02			

▶ **FIGURE 4.18b** StatCrunch output for the regression line to predict GPA from math SAT score. The structure of the output is typical for many statistical software packages.

We now check this calculation by hand, using Formula 4.2.

We are given that

For math SAT scores: $\bar{x} = 649.5,\ s_x = 66.3$

For GPAs: $\bar{y} = 2.63,\ s_y = 0.580$

and $r = 0.194$

First we must find the slope:

$$b = r\frac{s_y}{s_x} = 0.194 \times \frac{0.580}{66.3} = 0.001697$$

Now we can use the slope to find the intercept:

$$a = \bar{y} - b\bar{x} = 2.63 - 0.001697 \times 649.5 = 2.63 - 1.102202 = 1.53$$

Rounding off yields

Predicted GPA = 1.53 + 0.0017 math

b. Predicted GPA = 1.53 + 0.0017 math
= 1.53 + 0.0017 × 650
= 1.53 + 1.105
= 2.635

CONCLUSION We would expect someone with a math SAT score of 650 to have a GPA of about 2.64.

TRY THIS! Exercise 4.33

Different software packages present the intercept and slope differently. Therefore, you need to learn how to read the output of the software you are using. Example 5 shows output from several packages.

EXAMPLE 5 Technology Output for Regression

Figure 4.19 shows outputs from Minitab, StatCrunch, the TI-84, and Excel for finding the regression equation in Example 4 for GPA and math SAT scores.

QUESTION For each software package, explain how to find the equation of the regression line from the given output.

CONCLUSION Figure 4.19a: Minitab gives us a simple equation directly: GPA = 1.53 + 0.00170 Math (after rounding).

However, the more statistically correct format includes the word "predicted":

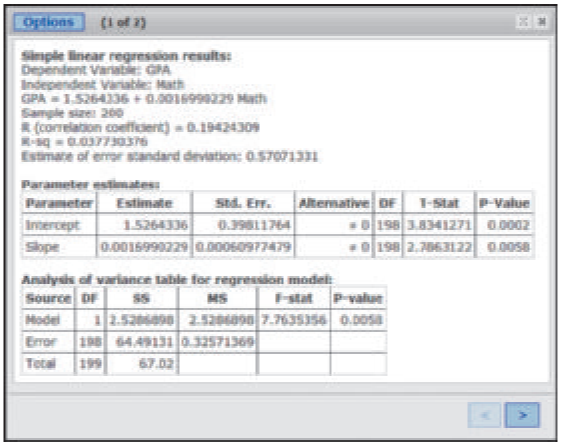
Predicted GPA = 1.53 + 0.00170 Math

Regression Analysis: GPA versus Math

```
The regression equation is
GPA = 1.526 + 0.001699 Math
```

◀ **FIGURE 4.19a** Minitab output.

Figure 4.19b: StatCrunch gives the equation directly near the top, but it also lists the intercept and slope separately in the table near the bottom.



Options (1 of 2)

Simple linear regression results:
Dependent Variable: GPA
Independent Variable: Math
GPA = 1.5264336 + 0.0016990229 Math
Sample size: 200
R (correlation coefficient) = 0.19424309
R-sq = 0.037730376
Estimate of error standard deviation: 0.57071331

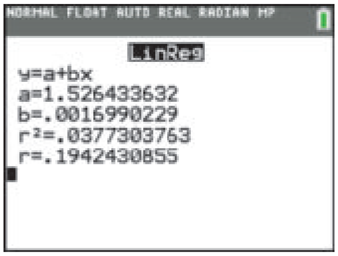
Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	P-Value
Intercept	1.5264336	0.39811764	≠ 0	198	3.8341271	0.0002
Slope	0.0016990229	0.00060977479	≠ 0	198	2.7863122	0.0058

Analysis of variance table for regression model:

Source	DF	SS	MS	F-stat	P-value
Model	1	2.5286898	2.5286898	7.7635356	0.0058
Error	198	64.49131	0.32571369		
Total	199	67.02			

▲ **FIGURE 4.19b** StatCrunch output.



▲ **FIGURE 4.19c** TI-84 output.

	Coefficients	Standard Error	t Stat	P-valve	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	1.526434	0.398118	3.834127	0.000169	0.741339	2.311528	0.741339	2.311528
Variable 1	0.001699	0.00061	2.786312	0.00585	0.000497	0.002902	0.000497	0.002902

▲ **FIGURE 4.19d** Excel output.

Figure 4.19c: The TI-84 gives us the coefficients to put together. The "a" value is the intercept, and the "b" value is the slope. If the diagnostics are on (use the CATALOG button), the TI-84 also gives the correlation.

Figure 4.19d: Excel shows the coefficients in the column labeled "Coefficients." The intercept is in the first row, labeled "Intercept," and the slope is in the row labeled "Variable 1."

TRY THIS! Exercise 4.35

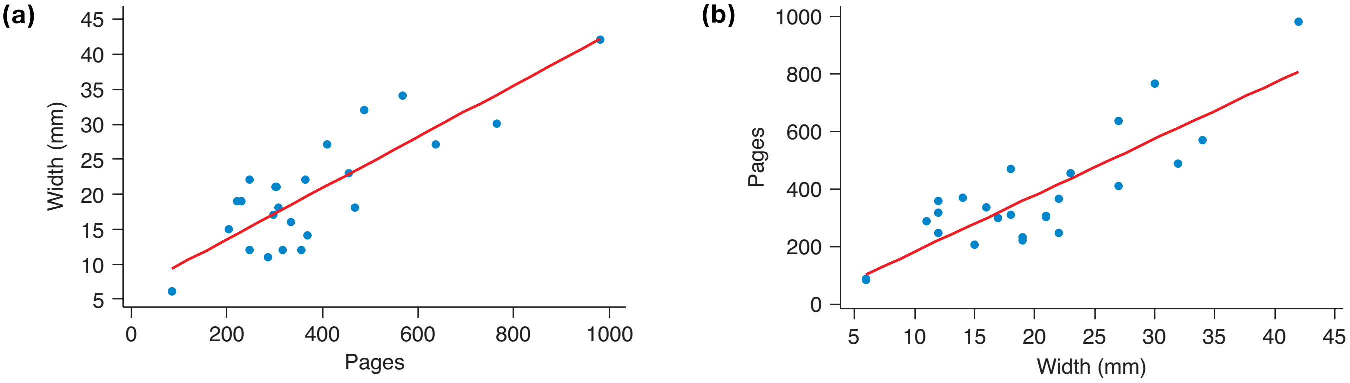
Interpreting the Regression Line

An important use of the regression line is to make predictions about the value of y that we will see for a given value of x. However, the regression line provides more information than just a predicted y-value. The regression line can also tell us about the rate of change of the mean value of y with respect to x, and it can help us understand the underlying theory behind cause-and-effect relationships.

Choosing x and y: Order Matters In Section 4.2, you saw that the correlation coefficient is the same no matter which variable you choose for x and which you choose for y. With regression, however, order matters.

Consider the collection of data about book widths. We used it earlier to predict the width of a book, given the number of pages. The equation of the regression line for this prediction problem (shown in Figure 4.20a) is

$$\text{Predicted Width} = 6.22 + 0.0366 \text{ Pages}$$



▲ **FIGURE 4.20** **(a)** Predicting width from number of pages. The slope is 0.037. **(b)** Predicting number of pages from width. The slope is 19.6.

But what if we instead wanted to predict how many pages there are in a book on the basis of the width of the book?

To do this, we would switch the order of the variables and use *Pages* as our *y*-variable and *Width* as our *x*-variable. Then the slope is calculated to be 19.6 (Figure 4.20b).

It is tempting to think that because we are flipping the graph over when we switch *x* and *y*, we can just flip the slope over to get the new slope. If this were true, then we could find the new slope simply by calculating 1/(old slope). However, that approach doesn't work. That would give us a slope of

$$\frac{1}{0.0366} = 27.3, \text{ which is not the same as the correct value of 19.6.}$$

How, then, do we know which variable goes where?

We use the variable plotted on the horizontal axis to make predictions about the variable plotted on the vertical axis. For this reason, the *x*-variable is called the **explanatory variable**, the **predictor variable**, or the **independent variable**. The *y*-variable is called the **response variable**, the **predicted variable**, or the **dependent variable**. These names reflect the different roles played by the *x*- and *y*-variables in regression. Which variable is which depends on what the regression line will be used to predict.

You'll see many pairs of terms used for the *x*- and *y*-variables in regression. Some are shown in Table 4.1.

x-Variable	*y*-Variable
Predictor variable	Predicted variable
Explanatory variable	Response variable
Independent variable	Dependent variable

◀ **TABLE 4.1** Terms used for the *x*- and *y*-variables.

EXAMPLE 6 Bedridden

It is hard to measure the height of people who are bedridden, and for many medical reasons it is often important to know a bedridden patient's height. However, it is not so difficult to measure the length of the ulna (the bone that runs from the elbow to the wrist). Data collected on non-bedridden people shows a strong linear association between ulnar length and height.

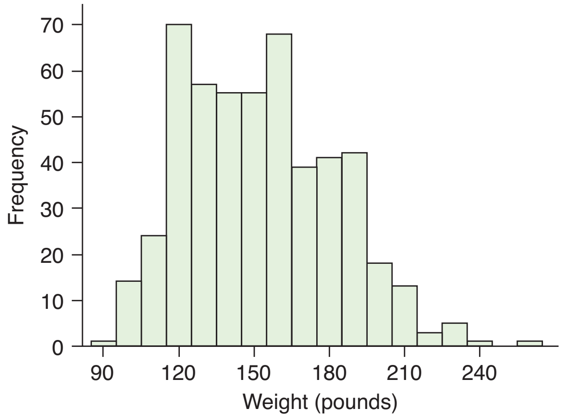
QUESTION When making a scatterplot to predict height from ulnar length, which variable should be plotted on the x-axis and which on the y-axis?

CONCLUSION We are measuring ulnar length to predict a person's height. Therefore, ulnar length is the predictor (independent variable) and is plotted on the x-axis, while height is the response (dependent variable) and is plotted on the y-axis.

TRY THIS! Exercise 4.41

The Regression Line Is a Line of Averages

Figure 4.21 shows a histogram of the weights of a sample of 507 active people. What is the typical weight?



▶ **FIGURE 4.21** A histogram of the weights of 507 active people.

One way of answering this question is to calculate the mean of the sample. The distribution of weights is a little right-skewed but not terribly so, and so the mean will probably give us a good idea of what is typical. The average weight of this group is 152.5 pounds.

Now, we know that an association exists between height and weight and that shorter people tend to weigh less. If we know someone's height, then, what weight would we guess for that person? Surely not the average weight of the whole group! We can do better than that. For instance, what's the typical weight of someone 66 inches tall? To answer this, it makes sense to look only at the weights of those people in our sample who are about 66 inches tall. To make sure we have enough 66-inch-tall people in our sample, let's include everyone who is *approximately* 66 inches tall. So let's look at a slice of people whose height is between 65.5 inches and 66.5 inches. We come up with 47 people. Some of their weights are

130, 150, 142, 149, 118, 189 . . .

Figure 4.22a shows this slice, which is centered at 66 inches.

The mean of these numbers is 141.5 pounds. We put a special symbol on the plot to record this point—a triangle at the point (66, 141.5), shown in Figure 4.22b.

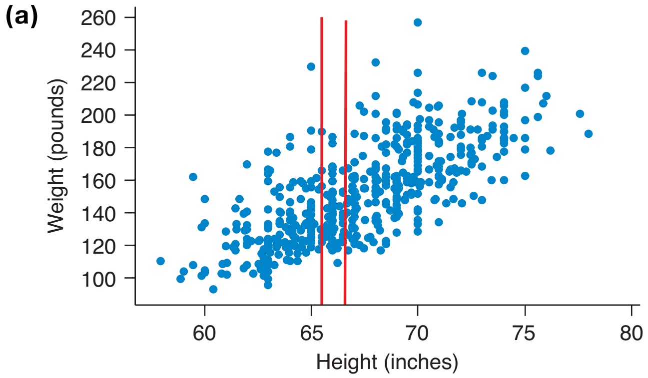
The reason for marking this point is that if we wanted to predict the weights of those who were 66 inches tall, one good answer would be 141.5 pounds.

What if we wanted to predict the weight of someone who is 70 inches tall? We could take a slice of the sample and look at those people who are between 69.5 inches and 70.5 inches tall. Typically, they're heavier than the people who are about 66 inches tall. Here are some of their weights:

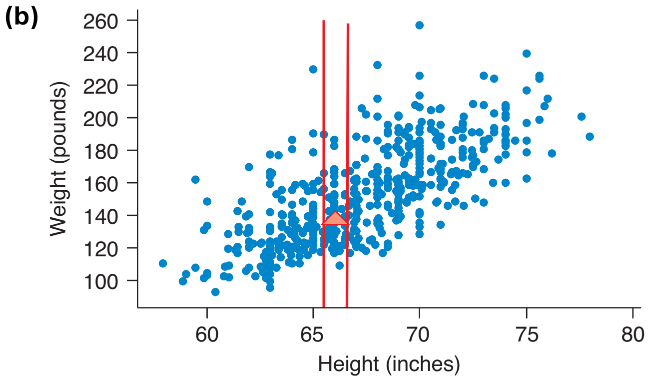
189, 197, 151, 207, 157 . . .

Their mean weight is 173.0 pounds. Let's put another special triangle symbol at (70, 173.0) to record this. We can continue in this fashion, and Figure 4.22c shows where the mean weights are for a few more heights.

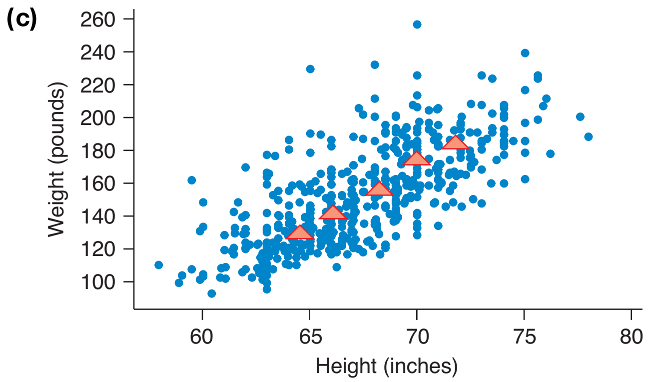
Note that the means fall (nearly) on a straight line. What could be the equation of this line? Figure 4.22d shows the regression line superimposed on the scatterplot with the means. They're nearly identical.



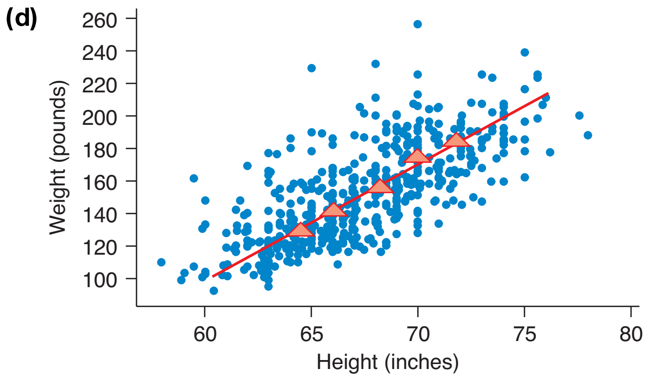
▲ **FIGURE 4.22a** Heights and weights with a slice at 66 inches.



▲ **FIGURE 4.22b** Heights and weights with the average at 66 inches, which is 141.5 pounds.



▲ **FIGURE 4.22c** Heights and weights with more means marked.



▲ **FIGURE 4.22d** Heights and weights with means, and a straight line superimposed on the scatterplot.

In theory, these means should lie exactly on the regression line. However, when working with real data, we often find that the theory doesn't always fit perfectly.

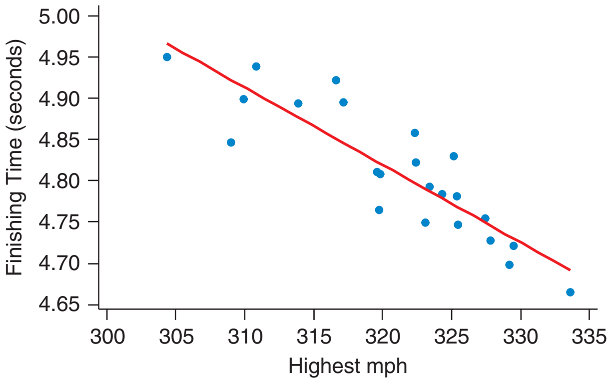
The series of graphs in Figure 4.22 illustrates a fundamental feature of the regression line: It is a line of means. You plug in a value for x, and the regression line "predicts" the mean y-value for everyone in the population with that x-value.

KEY POINT When the trend is linear, the regression line connects the points that represent the mean value of y for each value of x.

EXAMPLE 7 Funny-Car Race Finishing Times

A statistics student collected data on National Hot Rod Association driver John Force. The data consist of a sample of funny-car races. The drivers race in several different trials and change speed frequently. One question of interest is whether the fastest speed (in miles per hour, mph) driven during a trial is associated with the time to finish that trial (in seconds). If a driver's fastest speed cannot be maintained and the driver is forced to slow down, then it might be best to avoid going that fast. An examination of Figure 4.23, the scatterplot of finishing time (response variable) versus fastest speed (predictor) for a sample of trials for John Force, shows a reasonably linear association. The regression line is

$$\text{Predicted Time} = 7.84 - 0.0094\ (\text{Highest mph})$$



▶ **FIGURE 4.23** A scatterplot for the funny-car race. (Source: Hettinga, J. 2005, UCLA, Department of Statistics)

QUESTION What is the predicted mean finishing time when the driver's fastest speed is 318 mph?

SOLUTION We predict by estimating the mean finishing time when the speed is 318 mph. To find this, plug 318 into the regression line:

$$\text{Predicted Time} = 7.84 - 0.0094 \times 318 = 4.85 \text{ seconds}$$

CONCLUSION We predict that the finishing time will be 4.85 seconds.

TRY THIS! Exercise 4.43

Interpreting the Slope The slope tells us how to compare the mean *y*-values for objects that are 1 unit apart on the *x*-variable. For example, how different is the mean finishing time when the driver's fastest speed is 1 mph faster? The slope in Example 7 tells us that the means differ by a slim 0.0094 second. (Of course, races can be won or lost by such small differences.) What if the fastest time is 10 mph faster? Then the mean finishing time is $10 \times 0.0094 = 0.094$ second faster. We can see that in a typical race, if the driver wants to lose 1 full second from racing time, he or she must go over 100 mph faster, which is clearly not possible.

We should pay attention to whether the slope is 0 or very close to 0. When the slope is 0, we say that no (linear) relationship exists between *x* and *y*. The reason is that a 0 slope means that no matter what value of *x* you consider, the predicted value of *y* is always the same. The slope is 0 whenever the correlation coefficient is 0.

For example, the slope for runners' ages and their marathon times (see Figure 4.3a) is very close to 0 (0.000014 second per year of age).

KEY POINT For linear trends, a slope of 0 means there is no association between *x* and *y*.

SNAPSHOT THE SLOPE

WHAT IS IT? ▶ The slope of a regression line.

WHAT DOES IT DO? ▶ Tells us how different the mean *y*-value is for observations that are 1 unit apart on the *x*-variable.

HOW DOES IT DO IT? ▶ The regression line tells us the average *y*-values for any value of *x*. The slope tells us how these average *y*-values differ for different values of *x*.

HOW IS IT USED? ▶ To measure the linear association between two variables.

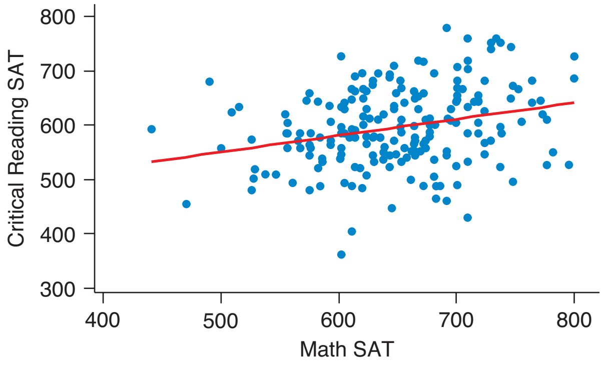
EXAMPLE 8 Math and Critical Reading SAT Scores

A regression analysis to study the relationship between SAT math and SAT critical reading scores resulted in this regression model:

$$\text{Predicted Critical Reading} = 398.81 + 0.3030\ \text{Math}$$

This model was based on a sample of 200 students.

The scatterplot in Figure 4.24 shows a linear relationship with a weak positive trend.

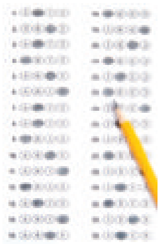


◀ **FIGURE 4.24** The regression line shows a weak positive trend between critical reading SAT score and math SAT score.

QUESTION Interpret the slope of this regression line.

SOLUTION The slope tells us that students who score 10 points higher on the math SAT had an average critical reading SAT score that was

$$10 \times 0.3030 = 3.03 \text{ points higher}$$



CONCLUSION Students who score 10 points higher on the math SAT score, on average, 3.03 points higher on critical reading.

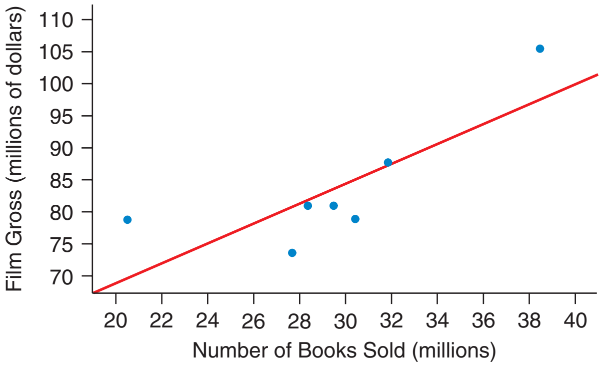
TRY THIS! Exercise 4.61a

Interpreting the Intercept The intercept tells us the predicted mean *y*-value when the *x*-value is 0. Quite often, this is not terribly helpful. Sometimes it is ridiculous. For example, the regression line to predict weight, given someone's height, tells us that if a person is 0 inches tall, then his or her predicted weight is negative 231.5 pounds!

Before interpreting the intercept, ask yourself whether it makes sense to talk about the *x*-variable taking on a 0 value. For the SAT data, you might think it makes sense to talk about getting a 0 on the math SAT. However, the lowest possible score on the SAT math portion is 200, so it is not possible to get a score of 0. (One lesson statisticians learn early is that you must know something about the data you analyze—knowing only the numbers is not enough!)

EXAMPLE 9 Magic and Movie Making

Many factors determine a movie's popularity. For this reason, movie studios like to bet on a sure thing, and so often make movies based on books or comics that are already popular. Figure 4.25 shows the amount of money made on each Harry Potter film (in millions of dollars) against the number of copies of each of the seven books (in



▶ **FIGURE 4.25** Film gross and number of books sold for the Harry Potter franchise.

millions of books sold). (Source: The Guardian DataBlog. Dollar amounts converted from pounds sterling using late 2013 exchange rate.)
The regression line is

$$\text{Predicted Film Gross} = 37.50 + 1.57 \text{ Sales}$$

QUESTION Interpret the intercept and slope.

CONCLUSION The intercept is the predicted mean gross for a book that sold no copies. The regression equation tells us that if the linear trend continues for 0 books sold, then we could expect the film to make about $37.50 million. (While it is unlikely that a new Harry Potter book would sell 0 copies, we might imagine that a movie studio would make a film based on the Harry Potter franchise that was not directly based on a book.)

The slope tells us that each additional million books sold translated into an average increase of 1.57 million dollars for the film. We could conclude that, at least as far as the Harry Potter books are concerned, more popular books meant more profitable movies.

TRY THIS! Exercise 4.65

SNAPSHOT THE INTERCEPT

WHAT IS IT? ▶ The intercept of a regression line.

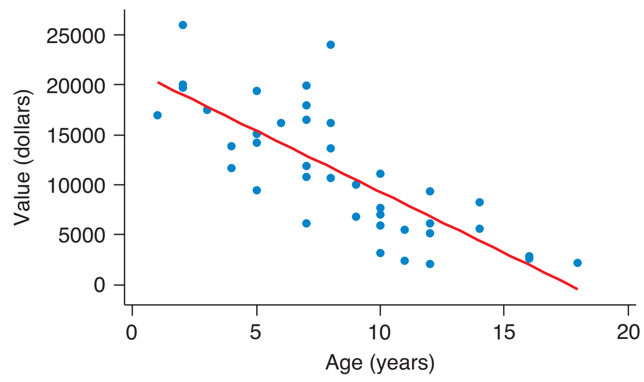
WHAT DOES IT DO? ▶ Tells us the average y-value for all observations that have a zero x-value.

HOW DOES IT DO IT? ▶ The regression line is the best fit to a linear association, and the intercept is the best prediction of the y-value when the x-value is 0.

HOW IS IT USED? ▶ It is not always useful. Often, it doesn't make sense for the x-variable to assume the value 0.

EXAMPLE 10 Age and Value of Cars

Figure 4.26 shows the relationship between age and Kelley Blue Book value for a sample of cars. These cars were owned by students in one of the author's statistics classes.



◀ **FIGURE 4.26** The age of a car and its value for a chosen sample.

The regression line is

$$\text{Predicted Value} = 21375 - 1215 \text{ Age}$$

QUESTION Interpret the slope and intercept.

CONCLUSION The intercept estimates that the average value of a new car (0 years old) owned by students in this class is \$21,375. The slope tells us that, on average, cars lost \$1215 in value each year.

TRY THIS! Exercise 4.67

SECTION 4.4

Evaluating the Linear Model

Regression is a very powerful tool for understanding linear relationships between numerical variables. However, we need to be aware of several potential interpretation pitfalls so that we can avoid them. We will also discuss methods for determining just how well the regression model fits the data.

Pitfalls to Avoid

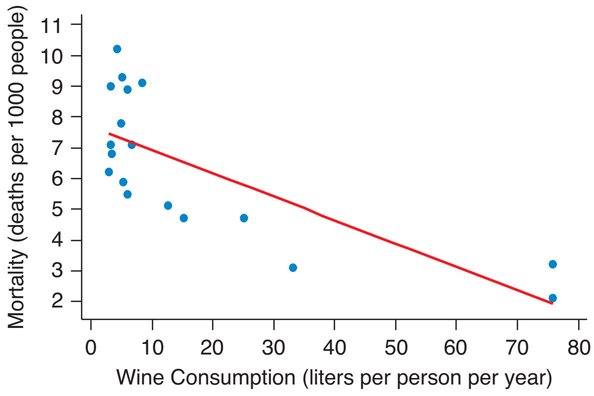
You can avoid most pitfalls by simply making a graph of your data and examining it closely before you begin interpreting your linear model. This section will offer some advice for sidestepping a few subtle complications that might arise.

Don't Fit Linear Models to Nonlinear Associations Regression models are useful only for linear associations. (Have you heard that somewhere before?) If the association is not linear, a regression model can be misleading and deceiving. For this reason, before you fit a regression model, you should always make a scatterplot to verify that the association seems linear.

Figure 4.27 shows an example of a bad regression model. The association between mortality rates for several industrialized countries (deaths per 1000 people) and wine consumption (liters per person per year) is nonlinear. The regression model is

$$\text{Predicted Mortality} = 7.69 - 0.0761 \text{ (Wine Consumption)}$$

but it provides a poor fit. The regression model suggests that countries with middle values of wine consumption should have higher mortality rates than they actually do.



▶ **FIGURE 4.27** The straight-line regression model is a poor fit to this nonlinear relationship. (Source: Leger et al. 1979)

Correlation Is Not Causation One important goal of science and business is to discover relationships between variables such that if you make a change to one variable, you know the other variable will change in a reliably predictable way. This is what is meant by "*x* causes *y*": Make a change in *x,* and a change in *y* will usually follow. For example, the distance it takes to stop your car depends on how fast you were traveling when you first applied the brakes (among other things); the amount of memory an mp3 file takes on your hard drive depends on the length of the song; and the size of your phone bill depends on how many minutes you talked. In these cases, a strong causal relationship exists between two variables, and if you were to collect data and make a scatterplot, you would see an association between those variables.

In statistics, however, we are often faced with the reverse situation, in which we see an association between two variables and wonder whether there is a cause-and-effect relationship. The correlation coefficient for the association could be quite strong, but as we saw earlier, correlation does not mean cause and effect. A strong correlation or a good-fitting regression line is not evidence of a cause-and-effect relationship.

KEY POINT An association between two variables is not sufficient evidence to conclude that a cause-and-effect relationship exists between the variables, no matter how strong the correlation or how well the regression line fits the data.

Be particularly careful about drawing cause-and-effect conclusions when interpreting the slope of a regression line. For example, for the SAT data,

$$\text{Predicted Critical Reading} = 398.81 + 0.3030\ \text{Math}$$

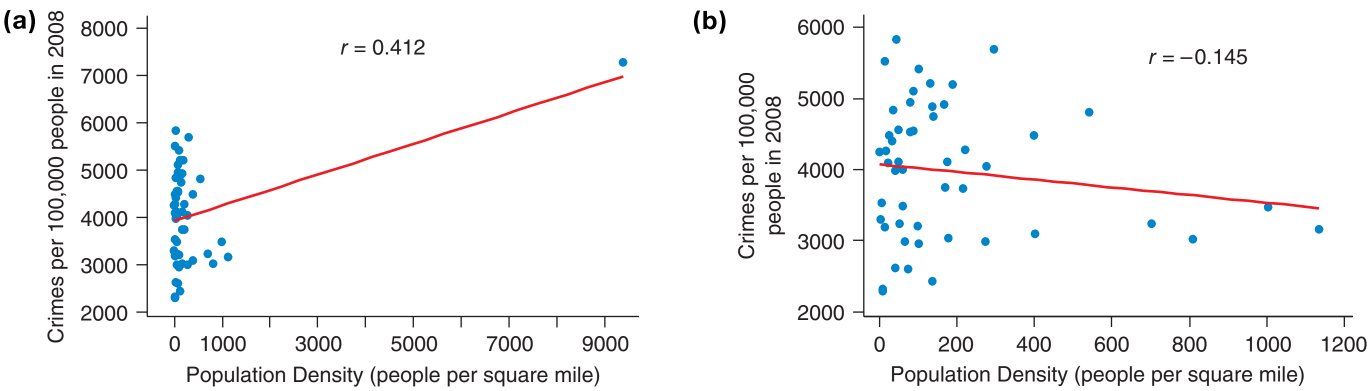
Even if this regression line fits the association very well, it does not give us sufficient evidence to conclude that if you improve your math score by 10 points, your critical reading score will go up by 3.03 points. As you learned in Chapter 1, because these data were not collected from a controlled experiment, the presence of confounding factors could prevent you from making a causal interpretation.

Beware of the algebra trap. In algebra, you were taught to interpret the slope to mean that "as *x* increases by 1 unit, *y* goes up by *b* units." However, quite often with data, the phrase "as *x* increases" doesn't make sense. When looking at the height and weight data, where *x* is height and *y* is weight, to say "*x* increases" means that people are growing taller! This is not accurate. It is much more accurate to interpret the slope as making comparisons between groups. For example, when comparing people of a certain height with those who are 1 inch taller, you can see that the taller individuals tend to weigh, on average, *b* pounds more.

When can we conclude that an association between two variables means a cause-and-effect relationship is present? Strictly speaking, never from an observational study and only when the data were collected from a controlled experiment. (Even in a controlled experiment, care must be taken that the experiment was designed correctly.) However, for many important questions, controlled experiments are not possible. In these cases, we can sometimes make conclusions about causality after a number of observational studies have been collected and examined, and if there is a strong theoretical explanation for why the variables should be related in a cause-and-effect fashion. For instance, it took many years of observational studies to conclude that smoking causes lung cancer, including studies that compared twins—one twin who smoked and one who did not—and numerous controlled experiments on lab animals.

Beware of Outliers Recall that when calculating sample means, we must remember that outliers can have a big effect. Because the regression line is a line of means, you might think that outliers would have a big effect on the regression line. And you'd be right. You should always check a scatterplot before performing a regression analysis to be sure there are no outliers.

The graphs in Figure 4.28 illustrate this effect. Both graphs in Figure 4.28 show crime rates (number of reported crimes per 100,000 people) versus population density (people per square mile) in 2008. Figure 4.28a includes all 50 states and the District of Columbia. The District is an outlier and has a strong influence on the regression line. Figure 4.28b excludes the District of Columbia.



▲ **FIGURE 4.28** Crime rates and population densities in the states. Part **(a)** includes Washington, D.C. and part **(b)** does not. (Source: Joyce 2009)

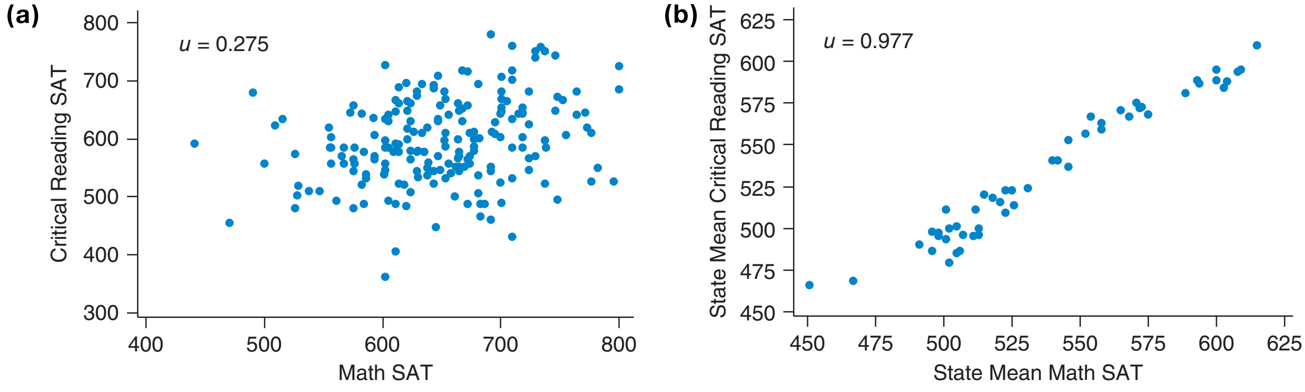
Without the District of Columbia, the slope of the regression line actually changes sign! Thus, whether you conclude that a positive or a negative association occurs between crime and population density depends on whether this one city is included. These types of observations are called **influential points** because their presence or absence has a big effect on conclusions. When you have influential points in your data, it is good practice to try the regression and correlation with and without these points (as we did) and to comment on the difference.

Looking Back

Outliers
You learned about outliers for one variable in Chapter 3. Outliers are values that lie a great distance from the bulk of the data. In two-variable associations, outliers are points that do not fit the trend or are far from the bulk of the data.

Regressions of Aggregate Data Researchers sometimes do regression analysis based on what we call **aggregate data**. Aggregate data are those for which each plotted point is from a summary of a group of individuals. For example, in a study to examine the relationship between SAT math and critical reading scores, we might use the *mean* of each of the 50 states rather than the scores of individual students. The regression line provides a summary of linear associations between two variables. The mean provides a summary of the center of a distribution of a variable. What happens if we have a collection of means of two variables and we do a regression with these?

This is a legitimate activity, but you need to proceed with caution. For example, Figure 4.29 shows scatterplots of critical reading and math SAT scores. However, in



▲ **FIGURE 4.29** Critical reading and math SAT scores. **(a)** Scores for individuals. **(b)** Means for states.

Figure 4.29a, each point represents an individual: an SAT score for each first-year student enrolled in one northeastern university for a fall semester. The scatterplot in Figure 4.29b seems to show a much stronger relationship, because in this case each point is an aggregate. Specifically, each point represents a state in the United States: the mean SAT score for all students in the state taking SAT tests.

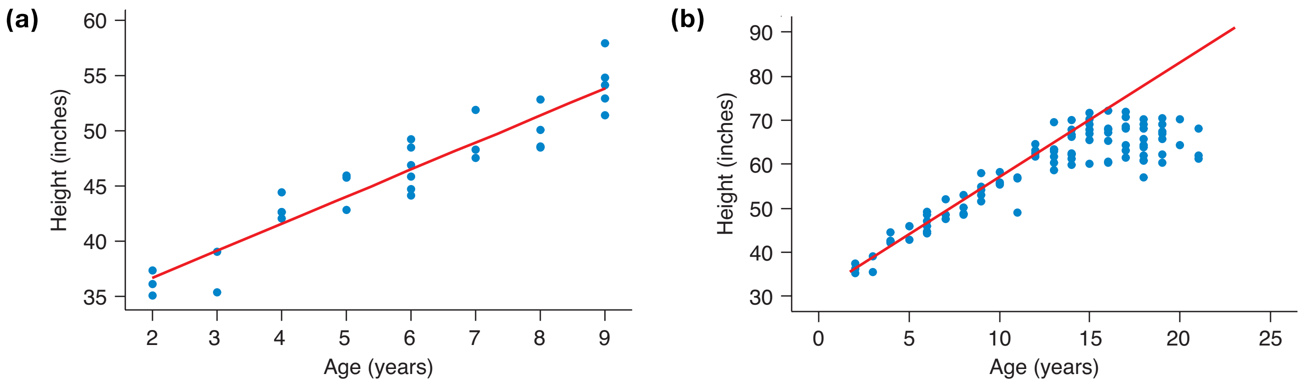
We can still interpret Figure 4.29b, as long as we're careful to remember that we are talking about states, not people. We *can* say a strong correlation exists between a state's mean math SAT score and a state's mean critical reading SAT score. We *cannot* say that there is a strong correlation between individual students' math and critical reading scores.

Don't Extrapolate! **Extrapolation** means that we use the regression line to make predictions beyond the range of our data. This practice can be dangerous, because although the association may have a linear shape for the range we're observing, that might not be true over a larger range. This means that our predictions might be wrong outside the range of observed x-values.

Figure 4.30a shows a graph of height versus age for children between 2 and 9 years old from a large national study. We've superimposed the regression line on the graph (Figure 4.30a), and it looks like a pretty good fit. However, although the regression model provides a good fit for children ages 2 through 9, it fails when we use the same model to predict heights for older children.

The regression line for the data shown in Figure 4.30a is

$$\text{Predicted Height} = 31.78 + 2.45 \text{ Age}$$



▲ **FIGURE 4.30** **(a)** Ages and heights for children between 2 and 9 years old. **(b)** Heights included for people up to age 21 years. The straight line is not valid in the upper range of ages. (Source: National Health and Nutrition Examination Survey, Centers for Disease Control)

However, we observed only children between the ages of 2 and 9. Can we use this line to predict the height of a 20-year-old?

The regression model predicts that the mean height of 20-year-old is 80.78 inches:

$$\text{Predicted Height} = 31.78 + 2.45\ \text{Age} = 31.78 + 2.45 \times 20 = 80.78\text{: nearly 7 feet!}$$

We can see from Figure 4.30b that the regression model provides a poor fit if we include people over the age of 9. Beyond that age, the trend is no longer linear, so we get bad predictions from the model.

It is often tempting to use the regression model beyond the range of data used to build the model. Don't. Unless you are very confident that the linear trend continues without change beyond the range of observed data, you must collect new data to cover the new range of values.

KEY POINT Don't extrapolate!

The Origin of the Word Regression *(Regression toward the Mean)*

The first definition in the *Oxford Dictionary* says that regression is a "backward movement or return to a previous state." So why are we are using it to describe predictions of one numerical variable (such as value of a car) from another numerical variable (such as age of the car)?

The term *regression* was coined by Francis Galton, who used the regression model to study genetic relationships. He noticed that even though taller-than-average fathers tended to have taller-than-average sons, the sons were somewhat closer to average than the fathers were. Also, shorter-than-average fathers tended to have sons who were closer to the average than their fathers. He called this phenomenon regression toward mediocrity, but later it came to be known as **regression toward the mean**.

You can see how regression toward the mean works by examining the formula for the slope of the regression line:

$$b = r\frac{s_y}{s_x}$$

This formula tells us that fathers who are one standard deviation taller than average (s_x inches above average) have sons who are not one standard deviation taller than average (s_y) but are instead r times s_y inches taller than average. Because r is a number between -1 and 1, r times s_y is usually smaller than s_y. Thus the "rise" will be less than the "run" in terms of standard deviations.

The *Sports Illustrated* jinx is an example of regression toward the mean. According to the jinx, athletes who make the cover of *Sports Illustrated* end up having a really bad year after appearing. Some professional athletes have refused to appear on the cover of *Sports Illustrated*. (Once, the editors published a picture of a black cat in that place of honor, because no athlete would agree to grace the cover.) However, if an athlete's performance in the first year is several standard deviations above average, the second year is likely to be closer to average. This is an example of regression toward the mean. For a star athlete, closer to average can seem disastrous.

The Coefficient of Determination, r^2, Measures Goodness of Fit

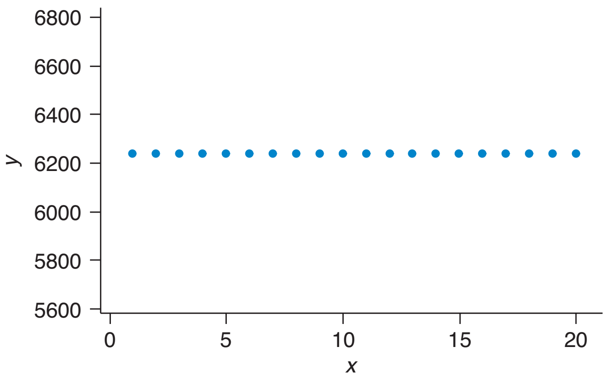
If we are convinced that the association we are examining is linear, then the regression line provides the best numerical summary of the relationship. But how good is "best"? The correlation coefficient, which measures the strength of linear relationships, can also be used to measure how well the regression line summarizes the data.

The **coefficient of determination** is simply the correlation coefficient squared: r^2. In fact, this statistic is often called ***r*-squared**. Usually, when reporting *r*-squared, we multiply by 100% to convert it to a percentage. Because *r* is always between −1 and 1, *r*-squared is always between 0% and 100%. A value of 100% means the relationship is perfectly linear and the regression line perfectly predicts the observations. A value of 0% means there is no linear relationship and the regression line does a very poor job.

For example, when we predicted the width of a book from the number of pages in the book, we found the correlation between these variables to be $r = 0.9202$. So the coefficient of determination is $0.9202^2 = 0.8468$, which we report as 84.7%.

What does this value of 84.7% mean? A useful interpretation of *r*-squared is that it measures how much of the variation in the response variable is explained by the explanatory variable. For example, 84.7% of the variation in book widths was explained by the number of pages. What does this mean?

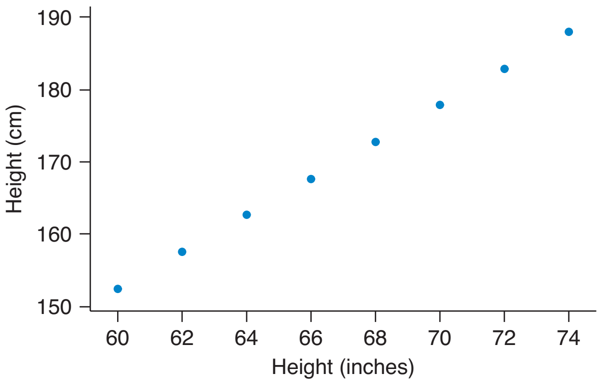
Figure 4.31 shows a scatterplot (simulated data) with a constant value for y ($y = 6240$) no matter what the x-value is. You can see that there is no variation in y, so there is also nothing to explain.



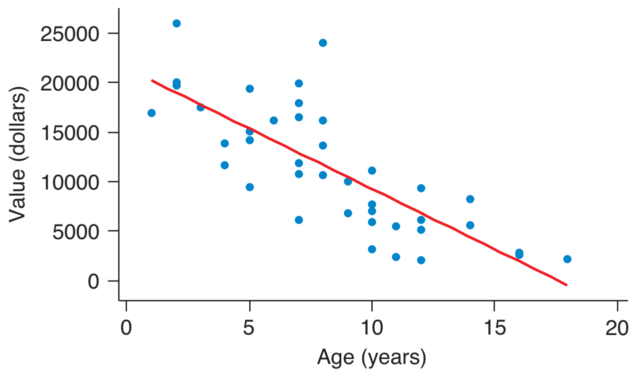
▶ **FIGURE 4.31** Because *y* has no variation, there is nothing to explain.

Figure 4.32 shows the height in both inches and in centimeters for several people. Here, we see variation in the *y*-variable because people naturally vary in height. Note that the points fall perfectly on a line because we're simply recording the same variable–height–for each person. The only difference is that the *x*-variable shows height in inches and the *y*-variable in centimeters. In other words, if you are given an *x*-value (a person's height in inches), then you know the *y*-value (the person's height in centimeters) precisely. Thus all of the variation in *y* is explained by the regression model. In this case, the coefficient of determination is 100%; all variation in *y* is perfectly explained by the best-fit line.

Real data are messier. Figure 4.33 shows a plot of the age and value of some cars. The regression line has been superimposed to remind us that there is, in fact, a linear



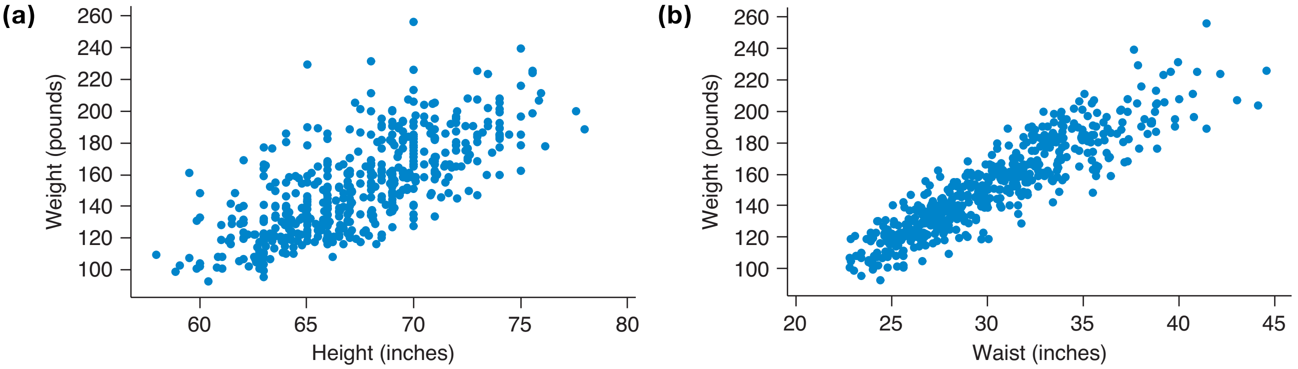
▲ **FIGURE 4.32** Heights of people in inches and in centimeters (cm) with a correlation of 1. The coefficient of determination is 100%.



▲ **FIGURE 4.33** Age and value of cars; the correlation is −0.778, and the coefficient of determination is 60.5%.

trend and that the regression line does capture it. The regression model explains some of the variation in y, but as we can see, it's not perfect; plugging the value of x into the regression line gives us an imperfect prediction of what y will be. In fact, for these data, $r = -0.778$, so with this regression line we've explained $(-0.778)^2 = 0.605$, or about 60.5%, of the variation in y.

The practical implication of r-squared is that it helps determine which explanatory variable would be best for making predictions about the response variable. For example, is waist size or height a better predictor of weight? We can see the answer to this question from the scatterplots in Figure 4.34, which show that the linear relationship is stronger (has less scatter) for waist size.



▲ **FIGURE 4.34** Scatterplots of height vs. weight **(a)** and waist size vs. weight **(b)**. Waist size has the larger coefficient of determination.

The r-squared for predicting weight from height is 51.4% (Figure 4.34a), and the r-squared for predicting weight from waist size is 81.7% (Figure 4.34b). We can explain more of the variation in these people's weights by using their waist sizes than by using their heights, and therefore, we can make better (more precise) predictions using waist size as the predictor.

KEY POINT If the association is linear, the larger the coefficient of determination (r-squared), the smaller the variation or scatter about the regression line, and the more accurate the predictions tend to be.

SNAPSHOT R-SQUARED

WHAT IS IT? ▶ r-squared or coefficient of determination.

WHAT DOES IT DO? ▶ Measures how well the data fit the linear model.

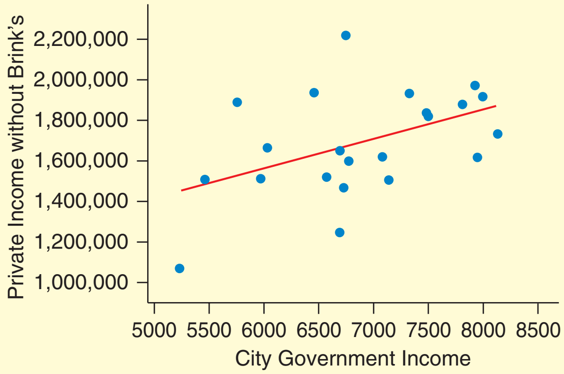
HOW DOES IT DO IT? ▶ If the points are tightly clustered around the regression line, r^2 has a value near 1, or 100%. If the points are more spread out, it has a value closer to 0. If there is no linear trend, and the points are a formless blob, it has a value near 0.

HOW IS IT USED? ▶ Works only with linear associations. Large values of r-squared tell us that predicted y-values are likely to be close to actual values. The coefficient of determination shows us the percentage of variation that can be explained by the regression line.

CASE STUDY REVISITED

Brink's was contracted by New York City to collect money from parking meters in some months, but some Brink's employees had been stealing some of the money. The city knew how much money its own employees had collected from meters each month and how much money honest private contractors had collected. It used this relation to predict the amount that Brink's should have collected.

The first step in this analysis was to determine the relation between the amount of money collected by the city itself and that collected by the honest contractors. Figure 4.35 shows a positive (though somewhat weak) linear association. In months in which the city collected more money, the (honest) private contractors also tended to collect more money.



▶ **FIGURE 4.35** Regression plot without Brink's. The association between amounts of meter income collected from honest private contracts (vertical axis) and amounts collected by the city government. The trend is positive, linear, and somewhat weak.

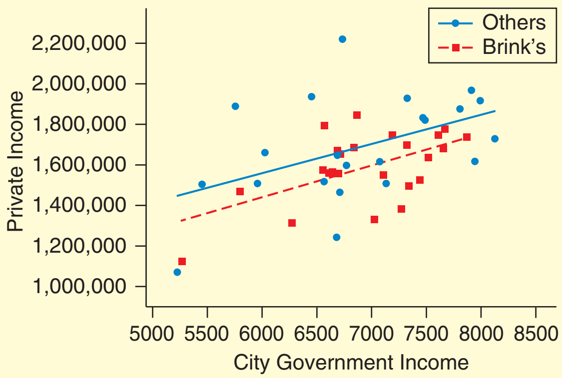
The regression line is

$$\text{Predicted Contractor Collection} = 688497 + 145.5\ (\text{City Income})$$

This model is useful for predicting how much Brink's should have collected in any given month. For example, in one particular month during which Brink's was the contractor, the city collected \$7016. The regression line predicts that if Brink's were like the other private contractors, it should have collected about

$$688497 + 145.5 \times 7016 = \$1{,}709{,}325$$

However, Brink's collected only \$1,330,143—about \$400,000 too little. This is just one month, but as Figure 4.36 shows, Brink's was consistently low in its collections. Figure 4.36 shows data both from Brink's and from the other private contractors. The Brink's data points are shown with red squares, and the data points from other contractors are shown with blue circles. You can see that the regression line for the Brink's collection (the dotted line) is below the regression line for the other contractors, indicating that the mean amount collected by Brink's was consistently lower than the mean amount collected by the other contractors.

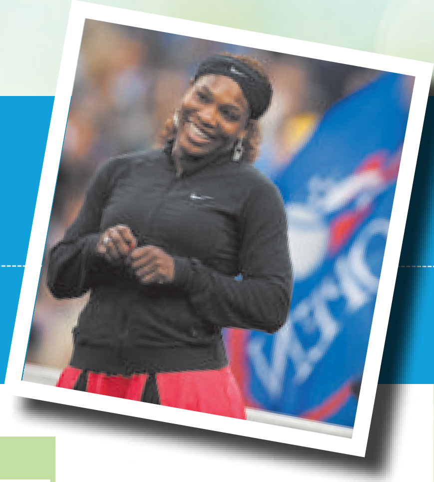


▶ **FIGURE 4.36** Scatterplot of income from parking meters. The dotted red line represents the regression line for the money that Brink's collected. The solid blue line represents the money they should have collected, judging on the basis of the amount collected by the city. Notice that the predicted means for Brink's tend to be lower than those of the other contractors.

For each month that Brink's had worked for the city, the city used the regression line to compute the amount that Brink's owed, by finding the difference between the amount that Brink's would presumably have collected if there had been no stealing and the amount they actually collected.

EXPLORING STATISTICS
CLASS ACTIVITY

Guessing the Age of Famous People



GOALS	MATERIALS
In this activity, you will learn how to interpret the slope and intercept of a regression line, using data you collect in class.	Graph paper and a calculator or computer.

ACTIVITY

Your instructor will give you a list of names of famous people. Beside each name, write your guess of the person's age, in years. Even if you don't know the age or don't know who the person is, give your best guess. If you work in a group, your group should discuss the guessed ages and record the best guess of the group. After you've finished, your instructor will give you a list of the actual ages of these people.

To examine the relationship between the actual ages and the ages you guessed, make a scatterplot with actual age on the x-axis and guessed age on the y-axis. Use technology to find the equation for the regression line and insert the line in the graph. Calculate the correlation.

BEFORE THE ACTIVITY

1. Suppose you guessed every age correctly. What would be the equation of the regression line? What would be the correlation?
2. What correlation do you think you will actually get? Why?
3. Suppose you consistently guess that people are older than they actually are. How will the intercept of the regression line compare with your intercept in Question 1? How about the slope?

AFTER THE ACTIVITY

1. How would you describe the association between the ages you guessed and the actual ages?
2. Is a regression line appropriate for your data? Explain why or why not.
3. What does it mean if a point falls above your regression line? Below your line?
4. What is the intercept of your line? What is the slope? Interpret the slope and intercept. Explain what these tell you about your ability to guess the ages of these people.

CHAPTER REVIEW

KEY TERMS

scatterplot, *168*
trend, *168*
strength, *168*
shape, *168*
positive association (positive trend), *168*
negative association (negative trend), *168*
linear, *170*
correlation coefficient, *172*
regression line, *180*
intercept, *180*
slope, *180*
explanatory variable, predictor variable, independent variable, *187*
response variable, predicted variable, dependent variable, *187*
influential point, *195*
aggregate data, *195*
extrapolation, *196*
regression toward the mean, *197*
coefficient of determination, r^2, r-squared, *198*

LEARNING OBJECTIVES

After reading this chapter and doing the assigned homework problems, you should

- Be able to write a concise and accurate description of an association between two numerical variables based on a scatterplot.
- Understand how to use a regression line to summarize a linear association between two numerical variables.
- Interpret the intercept and slope of a regression line in context, and know how to use the regression line to predict mean values of the response variable.
- Critically evaluate a regression model.

SUMMARY

The first step in looking at the relationship between two numerical variables is to make a scatterplot and learn as much about the association as you can. Examine trend, strength, and shape. Look for outliers.

Regression lines and correlation coefficients can be interpreted meaningfully only if the association is linear. A correlation coefficient close to 1 or –1 does *not* mean the association is linear. When the association is linear, the correlation coefficient can be calculated by Formula 4.1:

$$\textbf{Formula 4.1: } r = \frac{\sum z_x z_y}{n - 1}$$

A regression line can summarize the relationship in much the same way as a mean and standard deviation can summarize a distribution. Interpret the slope and intercept, and use the regression line to make predictions about the mean y-value for any given value of x. The coefficient of determination, r^2, indicates the strength of the association by measuring the proportion of variation in y that is explained by the regression line.

The regression line is given by

$$\textbf{Formula 4.2c: } \text{Predicted } y = a + bx$$

where the slope, b, is

$$\textbf{Formula 4.2a: } b = r\frac{s_y}{s_x}$$

and the intercept, a, is

$$\textbf{Formula 4.2b: } a = \bar{y} - b\bar{x}$$

When interpreting a regression analysis, be careful:

Don't extrapolate.
Don't make cause-and-effect conclusions if the data are observational.
Beware of outliers, which may (or may not) strongly affect the regression line.
Proceed with caution when dealing with aggregated data.

SOURCES

Anscombe, F.J., (1973) "Graphs in Statistical Analysis," *American Statistician* 27(1): 17–21.

Bentow, S., Afshartous, D. "Statistical Analysis of Brink's Data," www.stat.ucla.edu/cases (accessed April 4, 2014).

Faraway, J. J. 2005. *Extending the linear model with R*. Boca Raton, Florida: CRC Press.

Heinz, G., L. Peterson, R. Johnson, and C. Kerk. 2003. Exploring relationships in body dimensions. *Journal of Statistics Education* 11(2). http://www.amstat.org/publications/jse/

Joyce, C. A., ed. *The 2009 world almanac and book of facts*. New York: World Almanac.

Leger, A., A. Cochrane, and F. Moore. 1979. Factors associated with cardiac mortality in developed countries with particular reference to the consumption of wine. *The Lancet* 313(8124): 1017–1020.

SECTION EXERCISES

SECTION 4.1

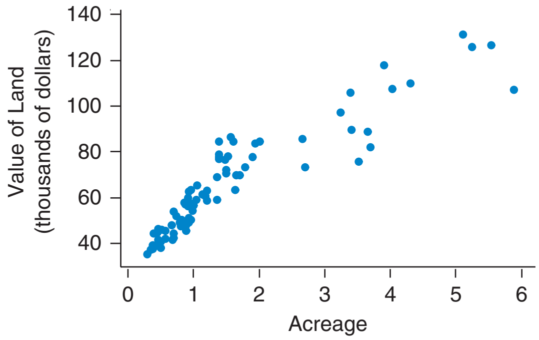
4.1 Predicting Land Value Both scatterplots concern the assessed value of land (with homes on the land), and both depict the same observations.

a. Which do you think has a stronger relationship with value of the land—the number of acres of land or the number of rooms in the homes? Why?

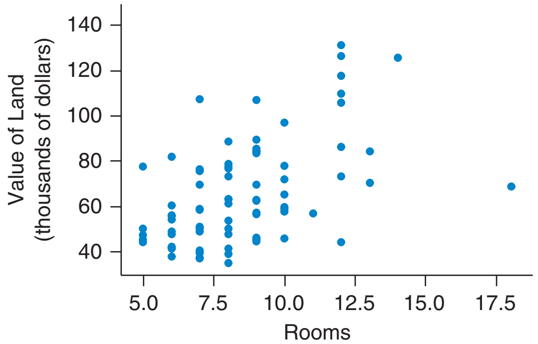
b. If you were trying to predict the value of a parcel of land (on which there is a home) in this area, would you be able to make a better prediction by knowing the acreage or the number of rooms in the house? Explain.

(Source: Minitab File, Student 12, “Assess.”)

(A)



(B)

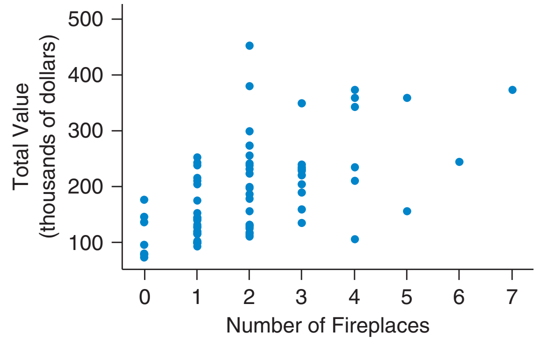


4.2 Predicting Total Value of Property Both scatterplots concern the total assessed value of properties that include homes, and both depict the same observations.

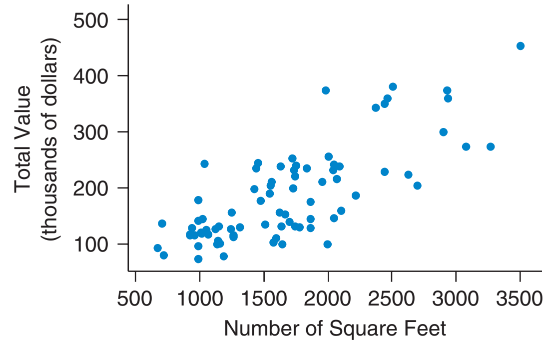
a. Which do you think has a stronger relationship with value of the property—the number of square feet in the home (shown in part B of the figure) or the number of fireplaces in the home (shown in part A of the figure)? Why?

b. If you were trying to predict the value of a property (where there is a home) in this area, would you be able to make a better prediction by knowing the number of square feet or the number of fireplaces? Explain.

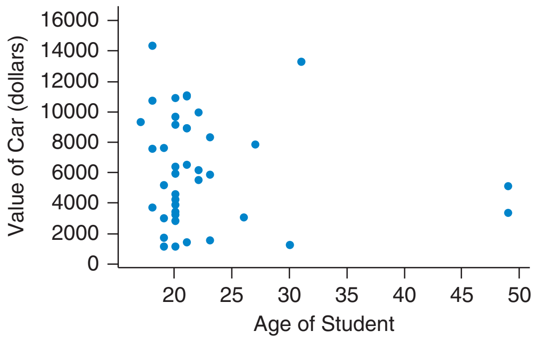
(A)



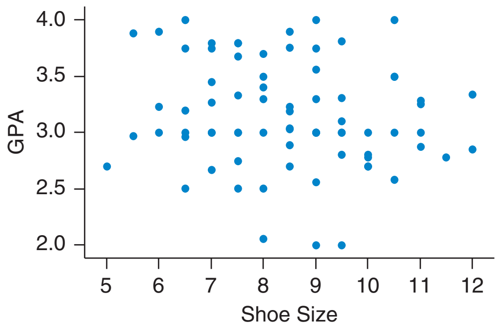
(B)



4.3 Car Value and Age of Student The scatterplot shows the age of students and the value of their cars according to the Kelley Blue Book. Does it show an increasing trend, a decreasing trend, or very little trend? Explain.

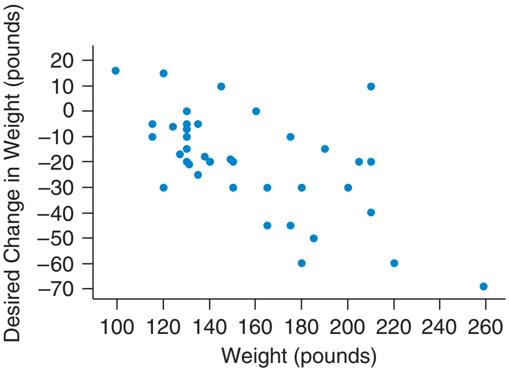


4.4 Shoe Size and GPA The figure shows a scatterplot of shoe size and GPA for some college students. Does it show an increasing trend, a decreasing trend, or no trend? Is there a strong relationship?

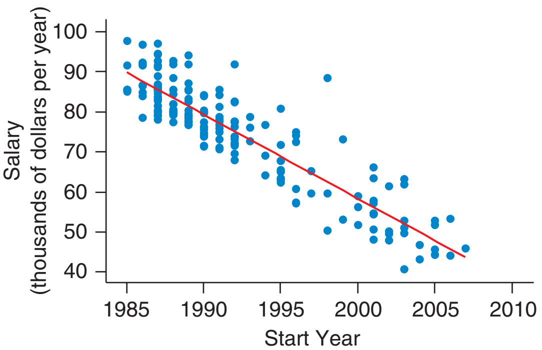


TRY **4.5 Weight Loss (Example 1)** The scatterplot shows the actual weight and desired weight change of some students. Thus, if a student weighed 220 and wanted to weigh 190, the desired weight change would be negative 30.

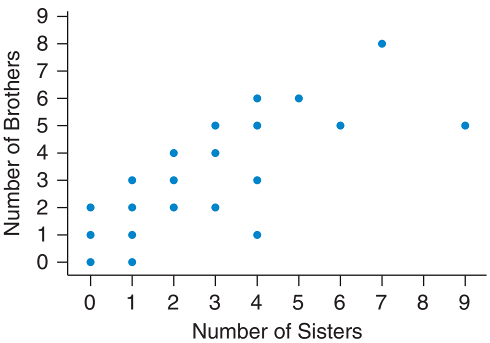
Explain what you see. In particular, what does it mean that the trend is negative?



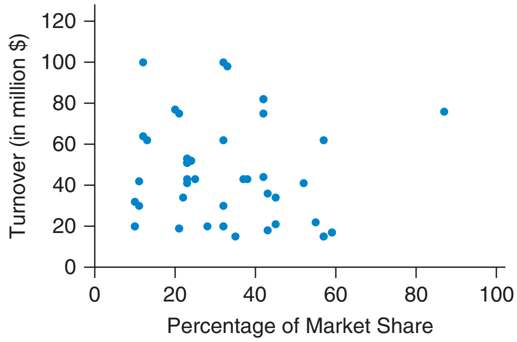
4.6 Comparing Salaries The scatterplot shows the salary and year of first employment for some professors at a college. Explain, in context, what the negative trend shows. Who makes the most and who makes the least?
(Source: Minitab, Student 12, "Salary.")



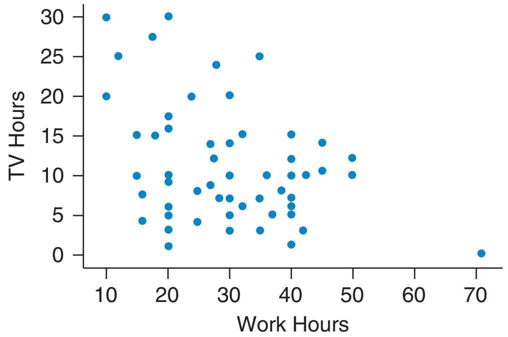
4.7 Sisters and Brothers The scatterplot shows the numbers of brothers and sisters for a large number of students. Do you think the trend is somewhat positive or somewhat negative? What does the direction (positive or negative) of the trend mean? Does the direction make sense in this context?



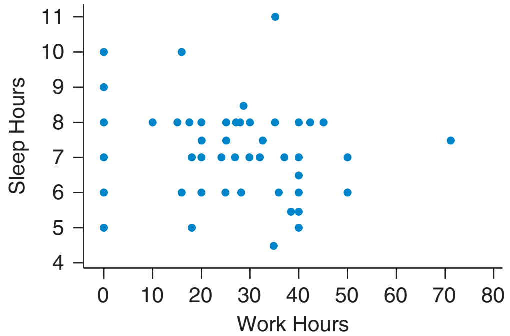
4.8 Market Share and Turnover The scatterplot shows data about the products of 40 corporate houses—the percentage of the market share for three or more products and the median annual turnover (in million \$). Describe and interpret the trend. (The outlier is a pharmaceutical company).



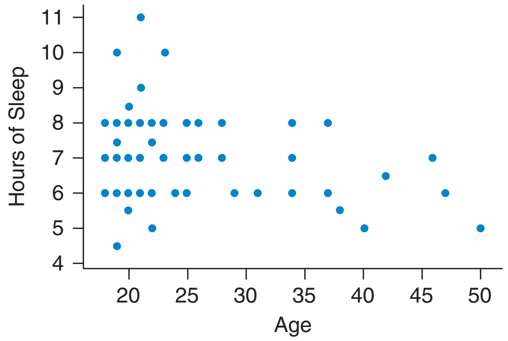
4.9 Work and TV The scatterplot shows the number of work hours and the number of TV hours per week for some college students who work. There is a very slight trend. Is the trend positive or negative? What does the direction of the trend mean in this context? Identify any unusual points.



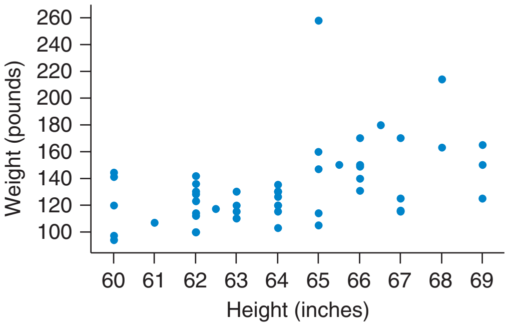
4.10 Work and Sleep The scatterplot shows the number of hours of work per week and the number of hours of sleep per night for some college students. Does the graph show a strong increasing trend, a strong decreasing trend, or very little trend? Explain.



4.11 Age and Sleep The scatterplot shows the age and number of hours of sleep "last night" for some students. Do you think the trend is slightly positive or slightly negative? What does that mean?



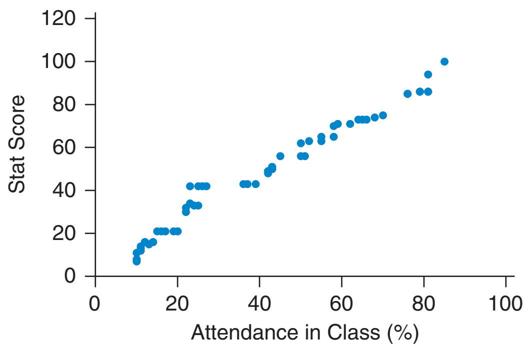
4.12 Height and Weight for Women The figure shows a scatterplot of the heights and weights of some women taking statistics. Describe what you see. Is the trend positive, negative, or near zero? Explain.



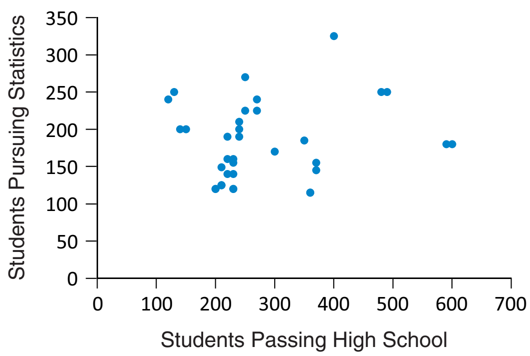
SECTION 4.2

4.13 Statistics Score and B.Stat

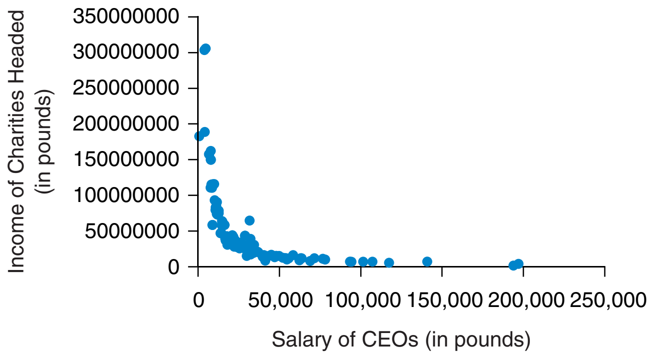
a. The first scatterplot shows scores in statistics and attendance of students majoring in statistics in a college. Would it make sense to find the correlation using this data set? Why or why not?



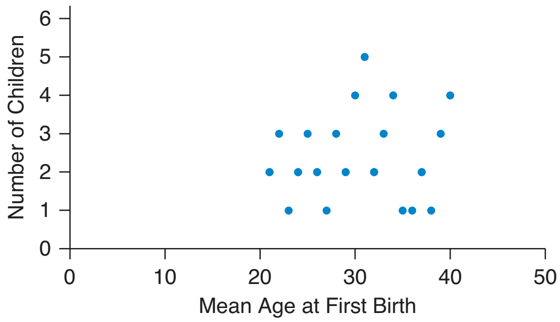
b. The second scatterplot shows the number of students passing high school examinations and the number of students pursuing higher studies in statistics. Would it make sense to find the correlation using this data set? Why or why not?



4.14 Salary of CEOs and Charities' Income The figure shows a scatterplot of the salaries of CEOs (in pounds) and income of the charities headed by them. Would it make sense to find the correlation for this data set? Explain. According to this graph, what is the mean range of the income of charities?



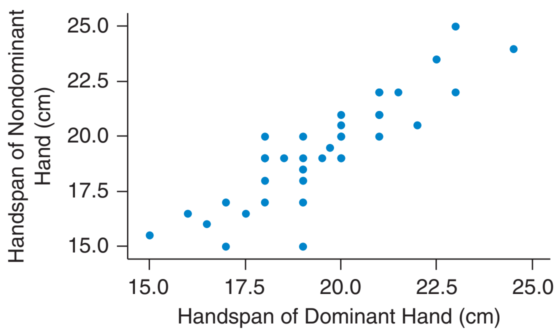
4.15 Do Working Mothers Have Children at a Later Age? On the basis of the scatterplot, do you think that the correlation coefficient between the mean age of working mothers when they give birth for the first time and the number of children for this figure is positive, negative, or no correlation?



4.16 Handspans Refer to the figure.

a. Would it make sense to find the correlation with this data set? Why or why not?

b. Would the correlation be positive, negative, or near 0?



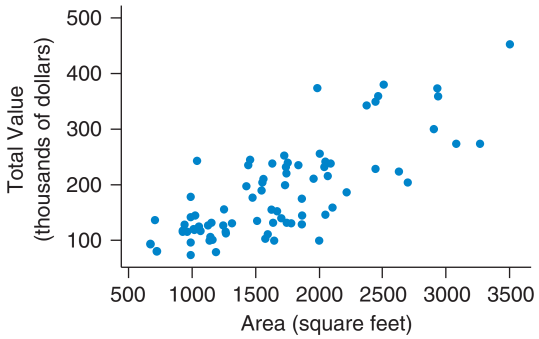
4.17 Matching Pick the letter of the graph that goes with each numerical value listed below for the correlation. Correlations:

0.767 __________

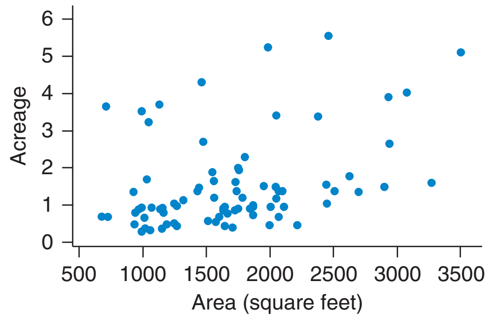
0.299 __________

−0.980 __________

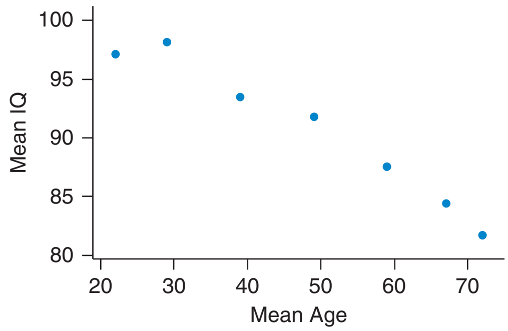
(A)



(B)



(C)



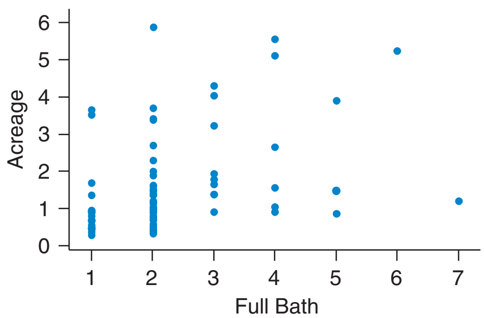
4.18 Matching Pick the letter of the graph that goes with each numerical value listed below for the correlation. Correlations:

−0.903 __________

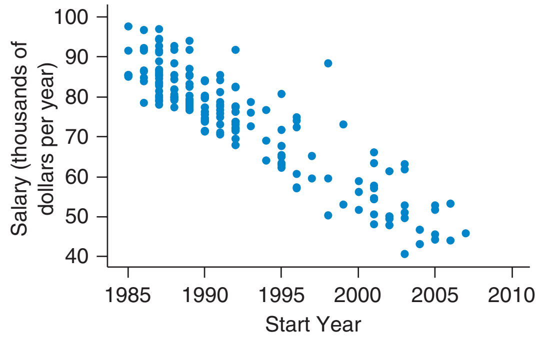
0.374 __________

0.777 __________

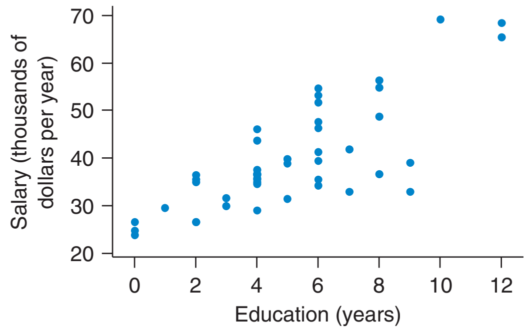
(A)



(B)



(C)



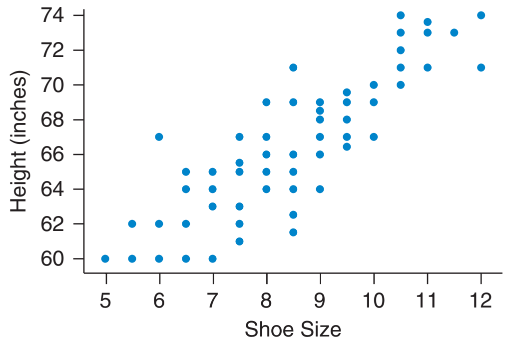
4.19 Matching Match each of the following correlations with the corresponding graph.

0.87 __________

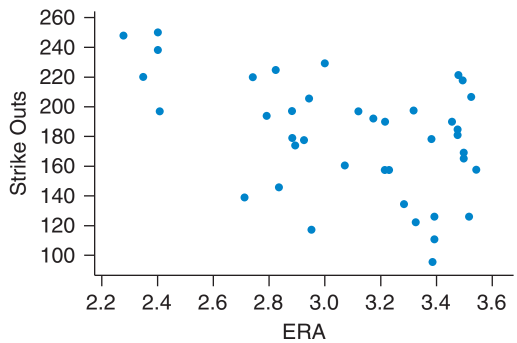
−0.47 __________

0.67 __________

(A)

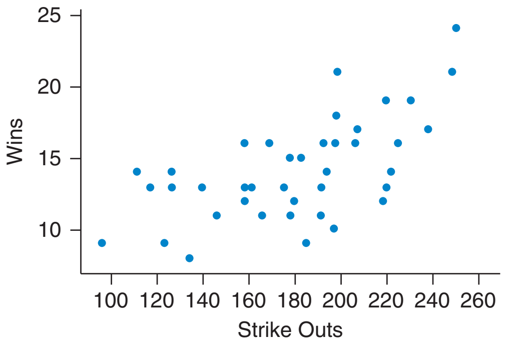


(B)



(Source: StatCrunch: 2011 MLB Pitching Stats according to owner: IrishBlazeFighter.)

(C)



(Source: StatCrunch: 2011 MLB Pitching Stats according to owner: IrishBlazeFighter.)

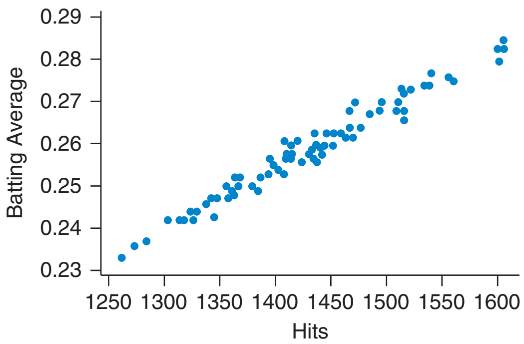
4.20 Matching Match each of the following correlations with the corresponding graph.

−0.51 __________

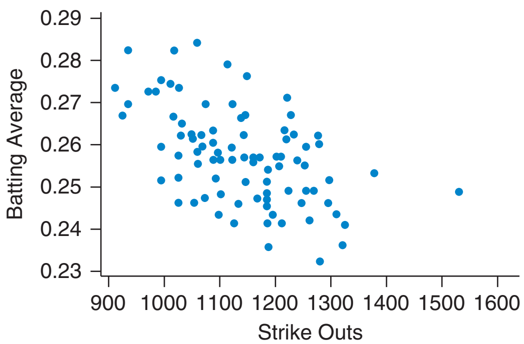
0.98 __________

0.18 __________

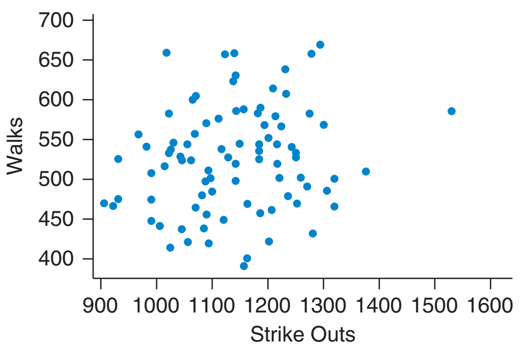
(A)



(B)



(C)



TRY **4.21 Trash (Example 2)** The table shows the number of people living in a house and the weight of trash (in pounds) at the curb just before trash pickup.

People	Trash (pounds)
2	18
3	33
6	93
1	23
7	83

a. Find the correlation between these numbers by using a computer or a statistical calculator.

b. Suppose some of the weight was from the container (each container weighs 3 pounds). Subtract 3 pounds from each weight, and find the new correlation with the number of people. What happens to the correlation when a constant is added (we added negative 3) to each number?

c. Suppose each house contained exactly twice the number of people, but the weight of the trash was the same. What happens to the correlation when numbers are multiplied by a constant?

4.22 Distance and Time A car is being driven at an average speed range of 50–70 kmph. The table shows distances between selected cities and the time taken by the car to cover these kilometers.

a. Calculate the correlation of the numbers shown in the part a table by using a computer or statistical calculator.

Distance (km)	Time (hrs)
120	2
294	4
160	3
340	6
310	5

b. The table for part b shows the same information, except that the distance was converted to meters by multiplying the number of kilometers by 1000. What happens to the correlation when numbers are multiplied by a constant?

Distance (m)	Time (hrs)
120000	2
294000	4
160000	3
340000	6
310000	5

c. Suppose the 0.5 hour that is lost at toll booths is added to the hours during each travel, no matter how long the distance is. The table for part c shows the new data. What happens to the correlation when a constant is added to each number?

Distance (km)	Time (hrs)
120	2.5
294	4.5
160	3.5
340	6.5
310	5.5

4.23 Work Hours and TV Hours In Exercise 4.9 there was a graph of the relationship between hours of TV and hours of work. Work hours was the predictor and TV hours was the response. If you reversed the variables so that TV hours was the predictor and work hours the response, what effect would that have on the numerical value of the correlation?

4.24 Price and Engine Capacity of Cars The correlation between the car price (in dollars) and the cubic capacity of the engine (in liters) for some cars is 0.89. If you found the correlation between the car price (by adding flat $1000 toward taxes and duties) and the cubic capacity of the engine (in liters) for the same cars, what would the correlation be?

4.25 Rate My Hotels Arnold, a destination traveler, went to the website TopTenReviews.com and looked up the reservation process rating and booking help rating of six hotels in a city. The ratings are 1 (worst reservation process) to 10 (best reservation process) and 1 (non-cooperative) to 10 (helpful). The numbers given are averages for each hotel. Assume the trend is linear, find the correlation, and comment on what it means.

Reservation Process	Booking Help
8.75	7.50
9.50	5.00
8.75	7.50
8.75	10.00
6.88	7.50

4.26 Cousins Five people were asked how many female first cousins they had and how many male first cousins. The data are shown in the table. Assume the trend is linear, find the correlation, and comment on what it means.

Female	Male
2	4
1	0
3	2
5	8
2	2

4.27 Video Games and BMI The table gives some hypothetical data for number of hours of video games played in a day and BMI (body mass index) for some young teenagers. Assume that the trend is linear, calculate the correlation, and explain what the sign shows. (Although these are hypothetical data, there have been studies that found the same direction for the trend.)

Games	BMI
0	27
2	25
4.5	36
3	40
4	35
1	19
2	34

4.28 See-Saw The table gives data on the heights (above ground) of the left and right seats of a see-saw (in feet). Assume the trend is linear, calculate the correlation, and explain what it shows.

Left	Right
4	0
3	1
2	2
1	3
0	4

SECTION 4.3

TRY **4.29 Used Car Market (Example 3)** Assume that the scatterplot shows the median volume and median value of used cars over the years in the UK used car market.

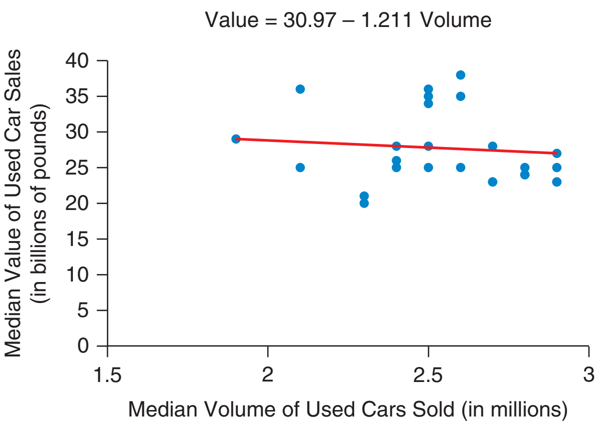
a. As the data are graphed, which is the independent and which is the dependent variable?

b. Why do you suppose median value and volume have been used instead of the mean?

c. Using the graph, estimate the median value of sales if the median volume of used cars sold is 2.3 million.

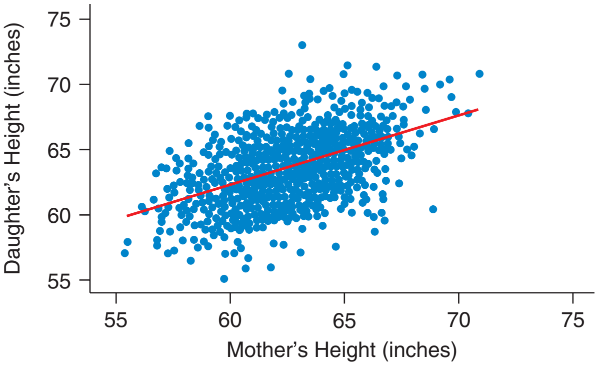
d. Use the equation to predict the median value of sales if the median volume of used cars sold is 2.3 million.

e. What other factors besides the price of used cars might influence the volume of used cars sold?



4.30 Mother and Daughter Heights The graph shows the heights of mothers and daughters. (Source: StatCrunch: Mother and Daughter Heights.xls. Owner: craig_slinkman)

Daughter = 29.92 + 0.5417 Mother



a. As the data are graphed, which is the independent variable and which the dependent variable?

b. From the graph, approximate the predicted height of the daughter of a mother who is 60 inches (5 feet) tall.

c. From the equation, determine the predicted height of the daughter of a mother who is 60 inches tall.

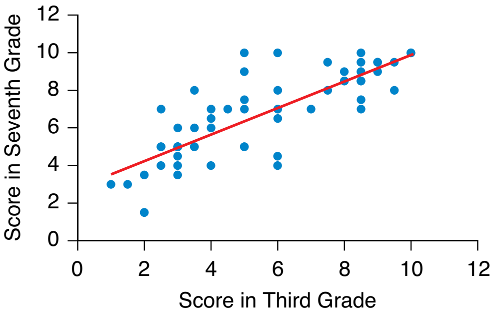
d. Interpret the slope.

e. What other factors besides mother's height might influence the daughter's height?

4.31 Median Annual Score The scatterplot shows the median annual score obtained by students of different schools in the third grade and obtained by the same students when promoted in the seventh grade. The correlation is 0.794. The regression equation is above the graph.

a. Find a rough estimate (by using the scatterplot) of median score for a seventh-grade student who had a median score of 8 in third grade.

Score in Seventh Grade = 2.952 + 0.6968
Score in Third Grade



b. Use the regression equation above the graph to get a more precise estimate of the median score in the seventh grade for a student who had a median score of about 8 in the third grade.

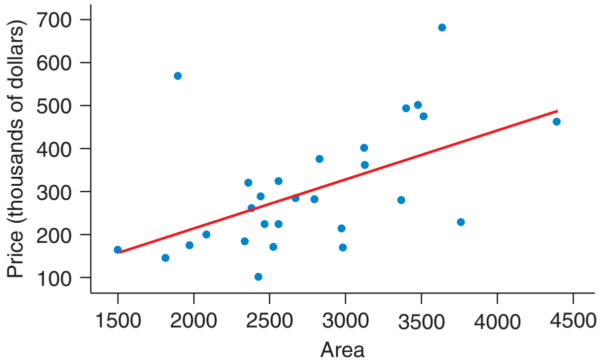
4.32 Home Prices and Areas of Four Bedroom Homes

a. Using the graph, estimate the predicted price for a home with 3000 square feet.

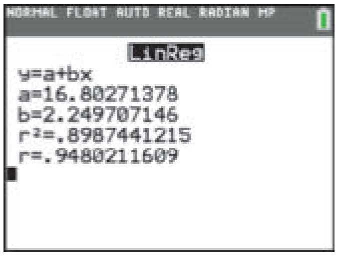
b. Use the equation to predict the price for a home with 3000 square feet.

(Source: Asking prices for a sample of 4 bedroom homes in Bryan-College Station, Texas, Yahoo Real Estate accessed via StatCrunch. Owner: Webster West)

Predicted Price = –11.77 + 0.1146 Sqft



TRY * **4.33 Height and Armspan for Women (Example 4)** TI-84 output from a linear model for predicting armspan (in centimeters) from height (in inches) is given in the figure. Summary statistics are also provided.



	Mean	Standard Deviation
Height, x	63.59	3.41
Armspan, y	159.86	8.10

To do parts a–c, assume that the association between armspan and height is linear.

a. Report the regression equation, using the words "Height" and "Armspan," not x and y, employing the output given.

b. Verify the slope by using the formula $b = r\dfrac{s_y}{s_x}$.

c. Verify the y-intercept using $a = \bar{y} - b\bar{x}$.

d. Using the regression equation, predict the armspan (in centimeters) for someone 64 inches tall.

4.34 Hand and Foot Length for Women The computer output shown below is for predicting foot length from hand length (in cm) for a group of women. Assume the trend is linear. Summary statistics for the data are shown in the table below.

	Mean	Standard Deviation
Hand, x	17.682	1.168
Foot, y	23.318	1.230

```
The regression equation is
Y = 5.67 + 0.998 X
Pearson correlation of HandL and FootL = 0.948
```

a. Report the regression equation, using the words "Hand" and "Foot," not x and y.

b. Verify the slope by using the formula $b = r\frac{s_y}{s_x}$.

c. Verify the y-intercept by using the formula $a = \bar{y} - b\bar{x}$.

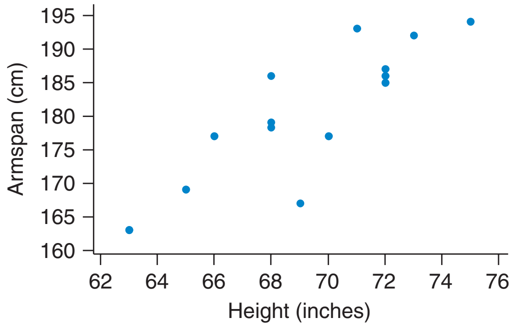
d. Using the regression equation, predict the foot length (in cm) for someone who has a hand length of 18 cm.

TRY **4.35 Height and Armspan for Men (Example 5)** Measurements were made for a sample of adult men. A regression line was fit to predict the men's armspan from their height. The output from several different statistical technologies is provided. The scatterplot confirms that the association between armspan and height is linear.

a. Report the equation for predicting armspan from height. Use words such as "Armspan," not just x and y.

b. Report the slope and intercept from each technology, using all the digits given.

(A)



(B)

```
The regression equation is
Armspan = 6.2 + 2.51 Height
```

From Minitab

(C)

Simple linear regression results:
Dependent Variable: Armspan
Independent Variable: Height
Armspan = 6.2408333 + 2.514674 Height
Sample size: 15
R (correlation coefficient) = 0.8681
R-sq = 0.7535989
Estimate of error standard deviation: 5.409662

From StatCrunch

(D)

	Coefficients
Intercept	6.240833
X Variable	2.514674

From Excel

(E)

```
NORMAL FLOAT AUTO REAL RADIAN MP
LinReg
y=a+bx
a=6.240833333
b=2.514673913
r²=.7535988997
r=.8681007428
```

From TI-84

4.36 Hand Length and Foot Length for Men Measurements were made for a sample of adult men. Assume that the association between their hand length and foot length is linear. Output for predicting foot length from hand length is provided from several different statistical technologies.

a. Report the equation for predicting foot length from hand length. Use the variable names *FootL* and *HandL* in the equation, rather than x and y.

b. Report the slope and intercept from each technology, using all the digits given.

(A)

```
The regression equation is
FootL = 15.8 + 0.563 HandL
```

From Minitab

(B)

Simple linear regression results:
Dependent Variable: FootL
Independent Variable: HandL
FootL = 15.807631 + 0.5626551 HandL
Sample size: 17
R (correlation coefficient) = 0.404
R-sq = 0.1632489
Estimate of error standard deviation: 1.6642156

From StatCrunch

(C)

	Coefficients
Intercept	15.80763
X Variable	0.562655

From Excel

(D)

```
NORMAL FLOAT AUTO REAL RADIAN MP
LinReg
y=a+bx
a=15.80763027
b=.5626550868
r²=.1632489032
r=.4040407197
```

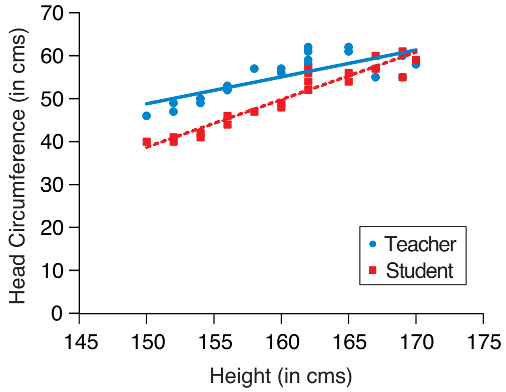
From TI-84

4.37 Comparing Correlation for Height and Weight The correlation between height and weight in a sample of students was found to be $r = 0.865$. The correlation between height and weight in a sample of teachers was found to be $r = 0.912$. Assuming both associations are linear, which association—the association between height and weight of students, or the association between height and weight of teachers—is stronger? Explain.

* **4.38 Height and Head Circumference for Students and Teachers** The scatterplot shows a solid blue line for predicting head circumference from height of teachers; the dotted red line is for predicting head circumference from height of students. The data were collected from a college.

a. Which line is higher, and what does that mean?

b. Which line has a steeper slope, and what does that mean?

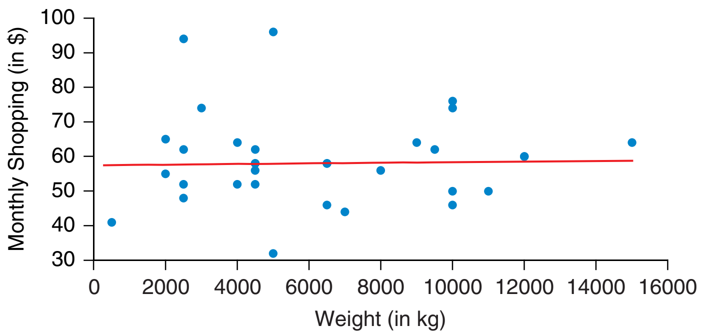


4.39 Monthly Shopping and Weight The figure shows a scatterplot of the monthly shopping (in \$) and weights (in kg) of some residents in a locality.

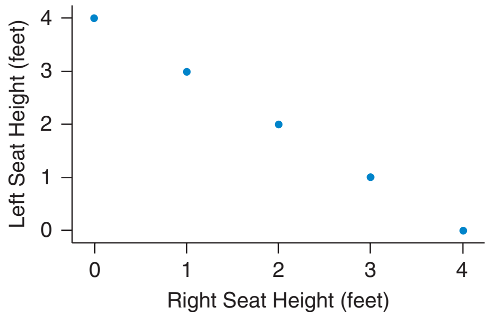
a. If a regression line were drawn on this graph, would it have a positive slope, a negative slope, or a slope near 0?

b. Give an estimate of the numerical value of correlation between monthly shopping and weight.

c. Explain what this graph tells us about the relation between monthly shopping and weight.



4.40 Seesaw The figure shows a scatterplot of the height of the left seat of a seesaw and the height of the right seat of the same seesaw. Estimate the numerical value of the correlation, and explain the reason for your estimate.



TRY **4.41 Choosing the Predictor and Response (Example 6)** Indicate which variable you think should be the predictor (x) and which variable should be the response (y). Explain your choices.

a. You collect data on the number of gallons of gas it takes to fill up the tank after driving a certain number of miles. You wish to know how many miles you've driven based on the number of gallons it took to fill up the tank.

b. Data on salaries and years of experience at a two-year college are used in a lawsuit to determine whether a faculty member is being paid the correct amount for her years of experience.

c. You wish to buy a belt for a friend and know only his weight. You have data on the weight and waist sizes for a large sample of adult men.

4.42 Choosing the Predictor and Response Indicate which variable you think should be the predictor (x) and which variable should be the response (y). Explain your choices.

a. Weights of nuggets of gold (in ounces) and their market value over the last few days are provided, and you wish to use this to estimate the value of a gold bracelet that weighs 4 ounces.

b. You have data collected on the amount of time since chlorine was added to the public swimming pool and the concentration of chlorine still in the pool. (Chlorine evaporates over time.) Chlorine was added to the pool at 8 A.M., and you wish to know what the concentration is now, at 3 P.M.

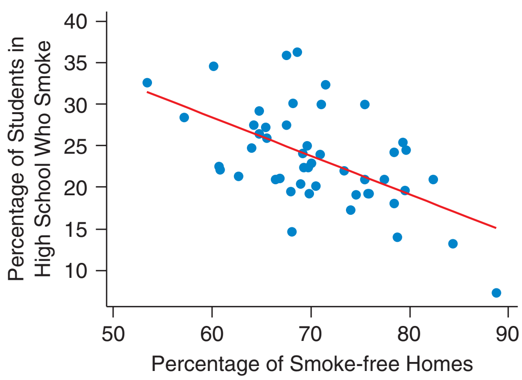
c. You have data on the circumference of oak trees (measured 12 inches from the ground) and their age (in years). An oak tree in the park has a circumference of 36 inches, and you wish to know approximately how old it is.

TRY **4.43 Percentage of Smoke-Free Homes and Percentage of High School Students Who Smoke (Example 7)** The figure shows a scatterplot with the regression line. The data are for the 50 states. The predictor is the percentage of smoke-free homes. The response is the percentage of high school students who smoke. The data came from the Centers for Disease Control and Prevention (CDC).

a. Explain what the trend shows.

b. Use the regression equation to predict the percentage of students in high school who smoke, assuming that there are 70% smoke-free homes in the state. Use 70 not 0.70.

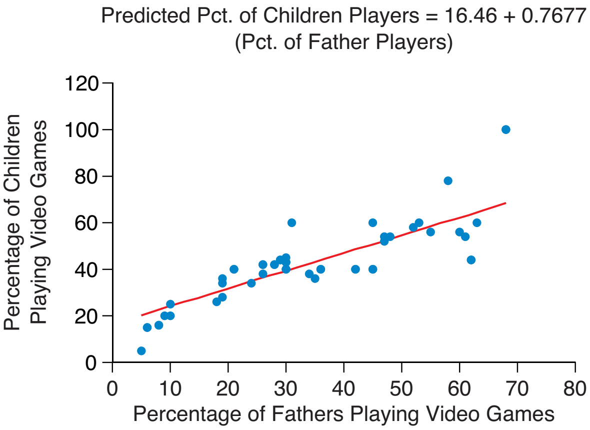
Predicted Pct. Smokers = 56.32 – 0.464 (Pct. Smoke-free)



4.44 Effect of Fathers Playing Video Games on Children The figure shows a scatterplot with a regression line. The data are for 40 schools. The predictor is the percentage of fathers who play video games. The response is the percentage of those fathers' offspring who are addicted to video games.

a. Explain what the trend shows.

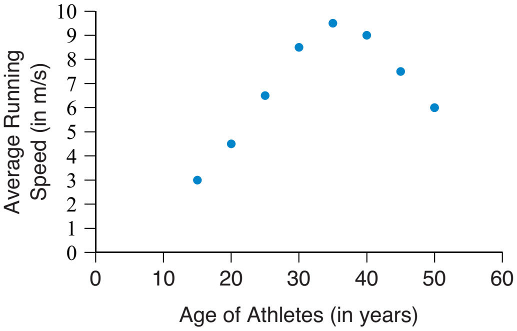
b. Use the regression equation to predict the percentage of children who play video games, assuming that 30% of fathers play video games. Use 30, not 0.30.



4.45 Athletes' Age and Speed The figure shows a graph of the running speed and age of a few athletes. The numbers came from various athletic records.

a. Explain what the graph tells us about athletes at different ages, and state which ages show the fastest athletes and which show the slowest athletes.

b. Explain why it would not be appropriate to use these data for linear regression.



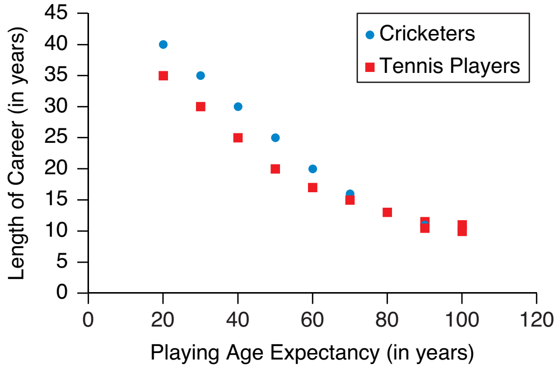
4.46 Do Cricketers Tend to Have Longer Careers Than Tennis Players? The figure shows mean length of career versus playing age expectancy (in years) for cricketers and tennis players, until the age of 50. Cricketers are represented by blue diamonds, and tennis players are represented by red squares.

a. Find your own age on the graph and estimate your playing age expectancy as a cricketer and tennis player.

b. Would it make sense to find the best straight line for this graph? Why or why not?

c. Is it reasonable to predict the length of career for a person who is 55 from the regression line for these data? Why or why not?

d. Explain what it means that nearly all of the blue diamonds (for cricketers) are above the red squares (for tennis players). (Above the age of 35, the red squares cover the blue diamonds because both are in the same place).



4.47 How is the time of a flight related to the distance of the flight? The table gives the distance from Boston to each city (in thousands of miles) and gives the time for one randomly chosen, commercial airplane to make that flight. Do a complete regression analysis that includes a scatterplot with the line, interprets the slope and intercept, and predicts how much time a nonstop flight from Boston to Seattle would take. The distance from Boston to Seattle is 3000 miles. *See page 222 for guidance.*

City	Distance (1000s of miles)	Time (hours)
St. Louis	1.141	2.83
Los Angeles	2.979	6.00
Paris	3.346	7.25
Denver	1.748	4.25
Salt Lake City	2.343	5.00
Houston	1.804	4.25
New York	0.218	1.25

4.48 Gender Gap in Universities This problem concerns the increasing gender gap in the UK universities. *The Guardian* surveyed and collected data on various students pursuing higher studies in different subject areas. Reported in the table are the number of students in various subject areas and the number of female students in those subject areas.

Subject Area	Number of Students	Number of Female Students
Medicine	10140	5845
Biological science	42040	25570
Physical science	17975	7650
Mathematical science	8895	3755
Computer science	20060	3500
Law	20440	12620
Business studies	77280	39715
Languages	28705	19775
Education	38465	30930

Assume that the association between the number of students in different subject areas and the number of female students in those subject areas is linear enough to proceed. Find the regression equation for predicting female students from the number of students in a subject area and report it. Interpret the sign of the slope clearly.

4.49 Do Countries with Higher Populations Have More Billionaires? The table gives the number of billionaires and the population (in hundreds of thousands) for various countries in the world in 2015. The number of billionaires comes from *Forbes Magazine* in July 2015.

a. Without doing any calculations, predict whether the correlation and the slope will be positive or negative. Explain your prediction.

b. Make a scatterplot with the population (in hundreds of thousands) on the x-axis and the number of billionaires on the y-axis. Was your prediction correct?

c. Find the numerical value for correlation.

d. Find the value of the slope and explain what it means in context. Be careful with the units.

e. Explain why interpreting the value for the intercept does not make sense in this situation.

Country	Billionaires	Population
Israel	17	84
Sweden	23	98
Iceland	1	3
Singapore	19	55
Switzerland	29	83
Cyprus	5	9
Hong Kong	55	73
Guernsey	1	1
St. Kitts and Nevis	1	1
Monaco	3	1

* **4.50 Semesters and Units** The table shows the self-reported number of semesters completed and the number of units completed for 15 students at a community college. All units were counted, but attending summer school was not included as a semester.

a. Make a scatterplot with the number of semesters on the x-axis and the number of units on the y-axis. Does one point stand out as unusual? Explain why it is unusual. (At most colleges, full-time students take between 12 and 18 units per semester.)

Finish each part *two ways,* with and without the unusual point, and comment on the differences.

b. Find the numerical values for the correlation between semesters and units.

c. Find the two equations for the two regression lines.

d. Insert the lines. Use technology if possible.

e. Report the slopes and intercepts of the regression lines and explain what they show. If the intercepts are not appropriate to report, explain why.

Sems	Units
2	21.0
4	130.0
5	50.0
7	112.0
3	45.5
3	32.0
8	140.0
0	0.0

Sems	Units
3	30.0
4	60.0
3	45.0
5	70.0
3	32.0
8	70.0
6	60.0

4.51 Pitchers The table shows the number of wins and the number of strike-outs (SO) for 40 baseball pitchers in the major leagues in 2011. (Source: 2011 MLB PITCHING STATS, http://www.baseball-reference.com/leagues/MLB/2011-pitching-leaders.shtml, accessed via StatCrunch. Owner: IrishBlazeFighter)

a. Make a scatterplot of the data, and state the sign of the slope from the scatterplot. Use strike-outs to predict wins.

b. Use linear regression to find the equation of the best-fit line. Insert the line on the scatterplot using technology or by hand.

c. Interpret the slope.

d. Interpret the intercept and comment on it.

Wins	SOs
21	248
19	220
17	238
24	250
18	198
13	139
13	220
14	194
16	225
11	146
21	198
12	179
13	175
15	178
16	206
13	117
19	230
13	161
16	197
16	192

Wins	SOs
12	158
13	191
16	158
8	134
10	197
9	123
9	96
11	178
14	111
14	126
11	191
14	222
9	185
15	182
16	169
11	166
12	218
13	126
17	207
13	158

4.52 Text Messages The table shows the number of text messages sent and received by some people in one day. (Source: StatCrunch: Responses to survey How often do you text? Owner: Webster West. A subset was used.)

a. Make a scatterplot of the data, and state the sign of the slope from the scatterplot. Use the number sent as the independent variable.

b. Use linear regression to find the equation of the best-fit line. Graph the line with technology or by hand.

c. Interpret the slope.

d. Interpret the intercept.

Sent	Received
1	2
1	1
0	0
5	5
5	1
50	75
6	8
5	7
300	300
30	40

Sent	Received
10	10
3	5
2	2
5	5
0	0
2	2
200	200
1	1
100	100
50	50

SECTION 4.4

4.53 Answer the questions using complete sentences.

a. What is an influential point?

b. It has been noted that people who go to church frequently tend to have lower blood pressure than people who don't go to church. Does this mean you can lower your blood pressure by going to church? Why or why not? Explain.

4.54 Answer the questions, using complete sentences.

a. What is extrapolation and why is it a bad idea in regression analysis?

b. How is the coefficient of determination related to the correlation, and what does the coefficient of determination show?

* c. When testing the IQ of a group of adults (aged 25 to 50), an investigator noticed that the correlation between IQ and age was negative. Does this show that IQ goes down as we get older? Why or why not? Explain.

4.55 If there is a positive correlation between number of years studying math and shoe size (for children), does that prove that larger shoes cause more studying of math, or vice versa? Can you think of a confounding variable that might be influencing both of the other variables?

4.56 Suppose that the growth rate of children looks like a straight line if the height of a child is observed at the ages of 24 months, 28 months, 32 months, and 36 months. If you use the regression obtained from these ages and predict the height of the child at 21 years, you might find that the predicted height is 20 feet. What is wrong with the prediction and the process used?

4.57 Coefficient of Determination If the correlation between height and weight of a large group of people is 0.67, find the coefficient of determination (as a percent) and explain what it means. Assume that height is the predictor and weight is the response, and assume that the association between height and weight is linear.

4.58 Coefficient of Determination Does a correlation of −0.70 or +0.50 give a larger coefficient of determination? We say that the linear relationship that has the larger coefficient of determination is more strongly correlated. Which of the values shows a stronger correlation?

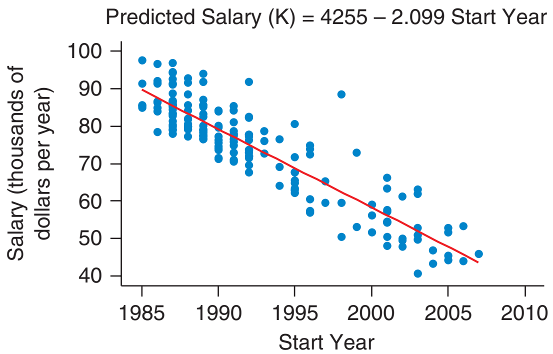
4.59 Up the Down Market Dinner Up the Down Market Dinner is a stock market game where, each year, in four cities across Canada, finance heavyweights meet and invest in made-up companies. They pick several individuals with Down syndrome and collectively raise a million dollars in their support. Explain why these collections rise and will probably improve in the future.

4.60 Blood Pressure Suppose a doctor telephones those patients who are in the highest 10% with regard to their recently recorded blood pressure and asks them to return for a clinical review. When she retakes their blood pressures, will those new blood pressures, as a group (that is, on average), tend to be higher than, lower than, or the same as the earlier blood pressures, and why?

TRY **4.61 Salary and Year of Employment (Example 8)** The equation for the regression line relating the salary and the year first employed is given above the figure.

a. Report the slope and explain what it means.

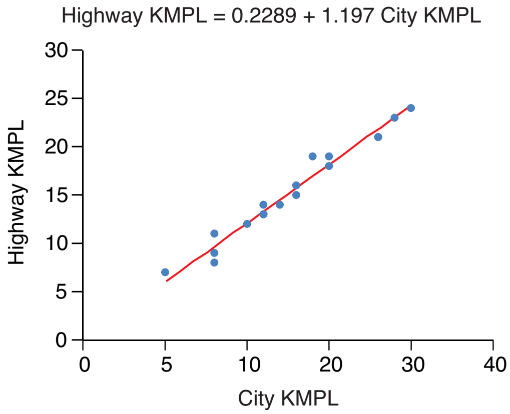
b. Either interpret the intercept (4,255,000) or explain why it is not appropriate to interpret the intercept.



4.62 KMPL: Highway and City The figure shows the relationship between the number of kilometers per liter on the highway and that in the city for some cars.

a. Report the slope and explain what it means.

b. Either interpret the intercept (0.2289) or explain why it is not appropriate to interpret the intercept.



4.63 Cost of Turkeys The table shows the weights and prices of some turkeys at different supermarkets.

a. Make a scatterplot with weight on the *x*-axis and cost on the *y*-axis. Include the regression line on your scatterplot.

b. Find the numerical value for the correlation between weight and price. Explain what the sign of the correlation shows.

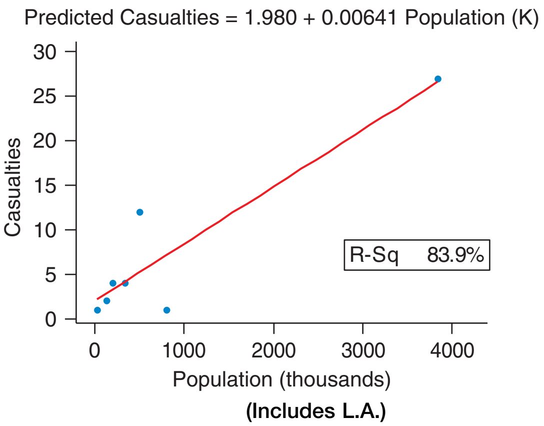
c. Report the equation of the best-fit straight line, using weight as the predictor (x) and cost as the response (y).

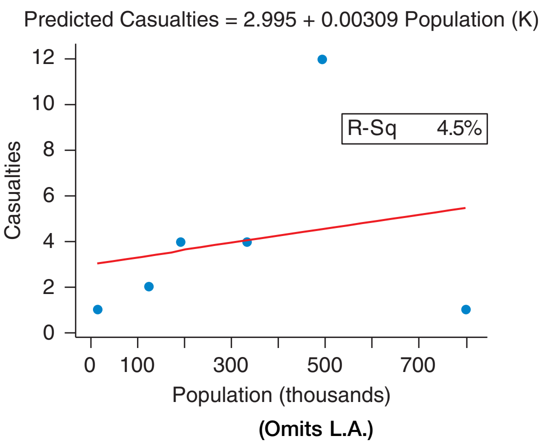
d. Report the slope and intercept of the regression line, and explain what they show. If the intercept is not appropriate to report, explain why.

e. Add a new point to your data: a 30-pound turkey that is free. Give the new value for r and the new regression equation. Explain what the negative correlation implies. What happened?

Weight (pounds)	Price
12.3	\$17.10
18.5	\$23.87
20.1	\$26.73
16.7	\$19.87
15.6	\$23.24
10.2	\$ 9.08

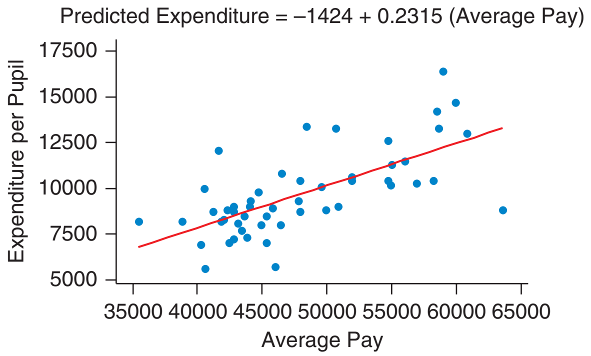
4.64 Iraq Casualties and Population of Hometowns The figures show the number of Iraq casualties through October 2009 and the population of some hometowns from which the servicemen or servicewomen came, according to the *Los Angeles Times*. Comment on the difference in graphs and in the coefficient of determination between the top scatterplot that included L.A. and the bottom scatterplot that did not include L.A. L.A. is the point with a population of nearly 4 million.





TRY **4.65 Teachers' Pay and Costs of Education (Example 9)** The figure shows a scatterplot with a regression line for teachers' average pay and the expenditure per pupil for each state for public schooling in 2007, according to *The 2009 World Almanac and Book of Facts*.

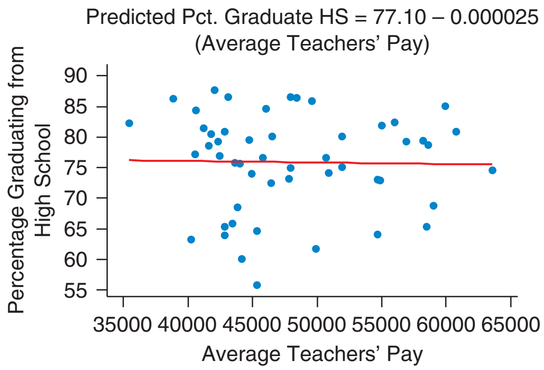
a. From the graph, is the correlation between teachers' average pay and the expenditure per pupil positive or negative?



b. Interpret the slope.

c. Interpret the intercept or explain why it should not be interpreted.

4.66 Teachers' Pay The figure shows a scatterplot with a regression line for the average teacher's pay and the percentage of students graduating from high school for each state in 2007, according to *The 2009 World Almanac and Book of Facts*. On the basis of the graph, do you think the correlation is positive, negative, or near 0? Explain what this means.

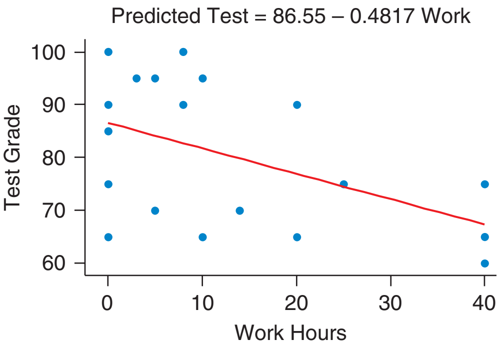


TRY **4.67 Does Having a Job Affect Students' Grades? (Example 10)** Grades on a political science test and the number of hours of paid work in the week before the test were recorded. The instructor was trying to predict the grade on a test from the hours of work. The figure shows a scatterplot and the regression line for these data.

a. Referring to the figure, state whether you think the correlation is positive or negative, and explain your prediction.

b. Interpret the slope.

c. Interpret the intercept.



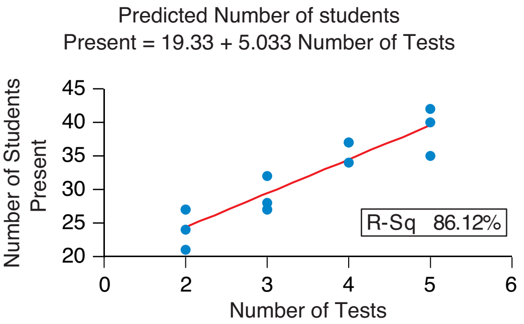
4.68 Attendance and Tests Data were collected that included information on the number of first-grade students in different sections who are present in one week and the number of tests in that week. The figure shows a scatterplot with the regression line.

a. Is the trend positive or negative? What does that mean?

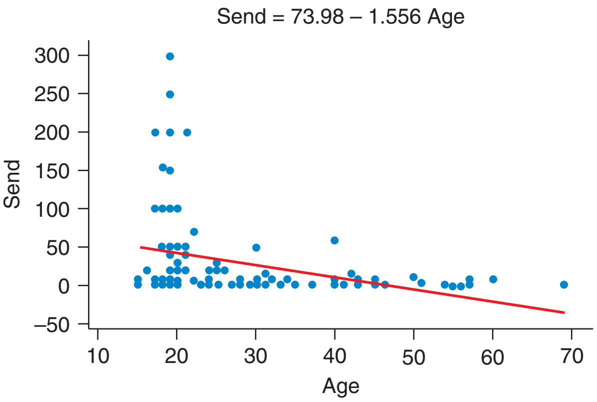
b. Now calculate the correlation between the number of present students and the number of tests in that week. (Use R-Sq from the figure and take the square root of it.)

c. Report the slope. For each additional test on a day, there are, on average, how many students present?

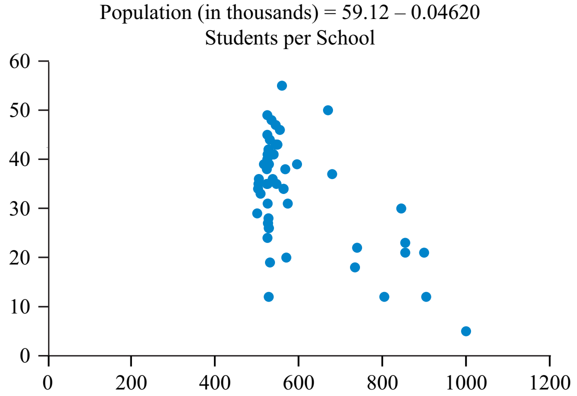
d. Either report the intercept or explain if it is appropriate to interpret it or not.



4.69 Age and Text Messages The scatterplot shows the relationship between age and number of text messages sent in a day. Comment on the appropriateness of linear regression. (Source: StatCrunch: Responses to survey How often do you text? Owner: Webster West)



4.70 Population and Schools The scatterplot shows the average population of some cities and the number of students per school in those cities. Comment on the appropriateness of the regression. What do you think accounts for the relationship? Do you think you could reduce the population of a city by having more students per school? Explain.



* **4.71 Education of Fathers and Mothers** The data shown in the table are the numbers of years of formal education of the fathers and mothers of a sample of 29 statistics students at a small community college in an area with many recent immigrants. (The means are both about 8, and the standard deviations are both about 4.6.) The scatterplot (not shown) suggests a linear trend.

Father	Mother	Father	Mother
3	3	4	7
11	13	16	13
3	4	13	3
12	12	11	10
8	16	6	6
1	2	3	2
8	8	12	14
6	4	14	12
12	8	12	12
0	10	12	13
8	12	12	18
5	5	13	8
8	8	0	0
5	5	3	5
12	6		

a. Find and report the regression equation for predicting the mother's years of education from the father's. Then find the predicted number of years for the mother if the father has 12 years of education, and find the predicted number of years for a mother if the father has 4 years of education.

b. Find and report the regression equation for predicting the father's years of education from the mother's. Then find the predicted number of years for the father if the mother has 12 years of education, and find the predicted number of years for the father if the mother has 4 years of education.

c. What phenomenon from the chapter does this demonstrate? Explain.

* **4.72 Heights of Fathers and Sons** The table shows some data from a sample of heights of fathers and their sons.

The scatterplot (not shown) suggests a linear trend.

a. Find and report the regression equation for predicting the son's height from the father's height. Then predict the height of a son with a father 74 inches tall. Also predict the height of a son of a father who is 65 inches tall.

b. Find and report the regression equation for predicting the father's height from the son's height. Then predict a father's height from that of a son who is 74 inches tall and also predict a father's height from that of a son who is 65 inches tall.

c. What phenomenon does this show?

Father's Height	Son's Height
75	74
72.5	71
72	71
71	73
71	68.5
70	70
69	69
69	66.5
69	72
68.5	66.5
67.5	65.5
67.5	70
67	67
65.5	64.5
64	67

g * **4.73 Test Scores** Assume that in a political science class, the teacher gives a midterm exam and a final exam. Assume that the association between midterm and final scores is linear. The summary statistics have been simplified for clarity.

Midterm: Mean = 75, Standard deviation = 10

Final: Mean = 75, Standard deviation = 10

Also, $r = 0.7$ and $n = 20$.

According to the regression equation, for a student who gets a 95 on the midterm, what is the predicted final exam grade? What phenomenon from the chapter does this demonstrate? Explain. *See page 223 for guidance.*

* **4.74 Salaries** Assume that data are collected on salaries in two cities (City A and City B). Assume that the association between these salaries is linear. Here are the summary statistics:

City A: Mean = $20,000, Standard deviation = $1,500

City B: Mean = 20,000, Standard deviation = $1,500

Also, $r = 0.84$ and $n = 50$.

a. Find and report the equation of the regression line to predict the salary in City B from the salary in City A.

b. For a person who has a salary of $18,000 in City A, predict the salary in City B.

c. Your answer to part b should be higher than $18,000. Why?

d. Consider a person who gets $45,000 in City A. Without doing any calculation, state whether the predicted salary in City B would be higher, lower, or the same as $45,000.

CHAPTER REVIEW EXERCISES

* **4.75 Heights and Weights of People** The table shows the heights and weights of some people. The scatterplot shows that the association is linear enough to proceed.

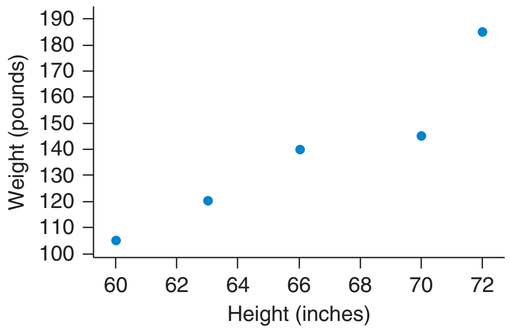
Height (inches)	Weight (pounds)
60	105
66	140
72	185
70	145
63	120

a. Calculate the correlation, and find and report the equation of the regression line, using height as the predictor and weight as the response.

b. Change the height to centimeters by multiplying each height in inches by 2.54. Find the weight in kilograms by dividing the weight in pounds by 2.205. Retain at least six digits in each number so there will be no errors due to rounding.

c. Report the correlation between height in centimeters and weight in kilograms, and compare it with the correlation between the height in inches and weight in pounds.

d. Find the equation of the regression line for predicting weight from height, using height in cm and weight in kg. Is the equation for weight (in pounds) and height (in inches) the same as or different from the equation for weight (in kg) and height (in cm)?



* **4.76 Heights and Weights of Men** The table shows the heights (in inches) and weights (in pounds) of 14 college men. The scatterplot shows that the association is linear enough to proceed.

Height (inches)	Weight (pounds)	Height (inches)	Weight (pounds)
68	205	70	200
68	168	69	175
74	230	72	210
68	190	72	205
67	185	72	185
69	190	71	200
68	165	73	195

a. Find the equation for the regression line with weight (in pounds) as the response and height (in inches) as the predictor. Report the slope and intercept of the regression line, and explain what they show. If the intercept is not appropriate to report, explain why.

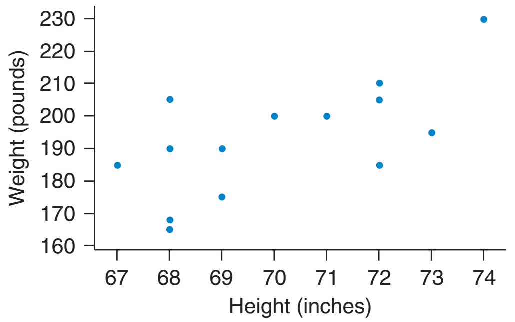
b. Find the correlation between weight (in pounds) and height (in inches).

c. Find the coefficient of determination and interpret it.

d. If you changed each height to centimeters by multiplying heights in inches by 2.54, what would the new correlation be? Explain.

e. Find the equation with weight (in pounds) as the response and height (in cm) as the predictor, and interpret the slope.

f. Summarize what you found: Does changing units change the correlation? Does changing units change the regression equation?



* **4.77 Homes** The table shows the asking price (in thousands of dollars) and area (square feet) of 30 homes in a town in Texas. (Source: Yahoo Real Estate, accessed via StatCrunch. Owner: Webster West)

Price	Sqft	Price	Sqft
2400	4918	499	3486
680	3645	492	3400
570	1900	475	3517
400	3123	460	4398
320	2365	375	2835
280	3361	360	3131
260	2383	323	2561
229	3770	288	2450
215	2979	285	2667
200	2088	281	2797
183	2343	225	2474
170	2526	225	2565
145	1812	176	1978
100	2432	165	1504
168	2988	160	1496

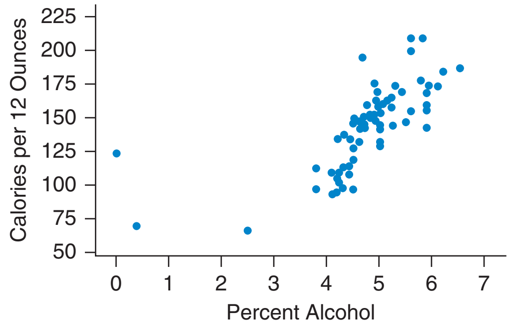
a. Do a complete analysis of the data (with square feet as the independent variable), including the graph, equation, interpretation of slope and intercept, and coefficient of determination.

b. Remove the high-end outlier and do another complete analysis.

c. Explain the changes from part a to part b.

* **4.78 Alcohol and Calories in Beer** At the text's website there is a data set that provides the number of calories per 12 ounces of beer and the percentage alcohol for several different brands of beer.

a. Refer to the scatterplot that follows. Is a linear analysis appropriate? Why or why not?



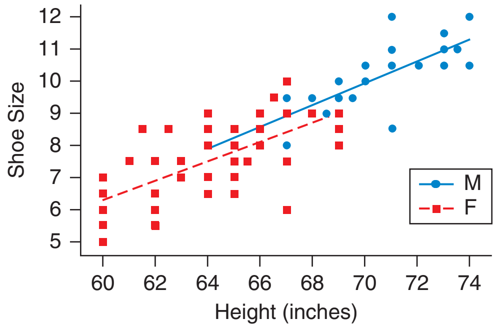
b. Remove the two low-end outliers (the beer with 0.0% alcohol and the beer with 0.4% alcohol), because most of the data come from alcoholic beers. Do a complete analysis of the data to help predict the calorie content based on the percent alcohol. Include the graph, equation, interpretation of slope and intercept, and coefficient of determination. (Source: beer100.com, accessed via StatCrunch. Owner: Webster West)

4.79 Shoe Size and Height The scatterplot shows the shoe size and height for some men (M) and women (F).

a. Why did we not extend the red line (for the women) all the way to 74 inches, instead stopping at 69 inches?

b. How do we interpret the fact that the blue line is above the red line?

c. How do we interpret the fact that the two lines are (nearly) parallel?



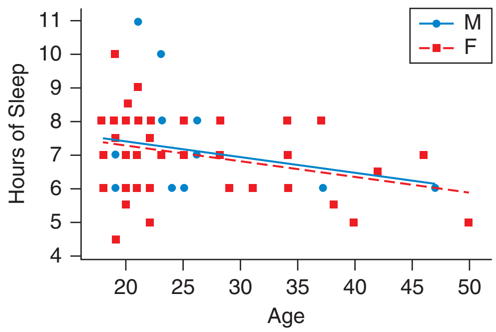
4.80 Age and Sleep The scatterplot shows the age in years and the number of hours of sleep for some college men (M) and women (F).

a. How do we interpret the fact that both lines have a negative slope?

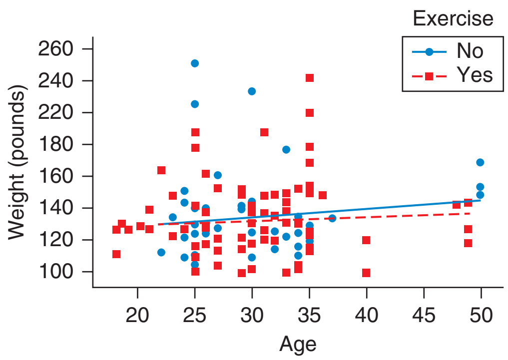
b. How do we interpret the fact that the slopes are the same for both lines?

c. How do we interpret the fact that the lines are nearly the same?

d. Why is the line for the men shorter than the line for the women?



4.81 Age and Weight The scatterplot shows the age and weight for some women. Some of them exercised regularly, and some did not. Explain what it means that the blue line (for those who did not exercise) is a bit steeper than the red line (for those who did exercise). (Source: StatCrunch: 2012 Women's final. Owner: molly7son@yahoo.com)

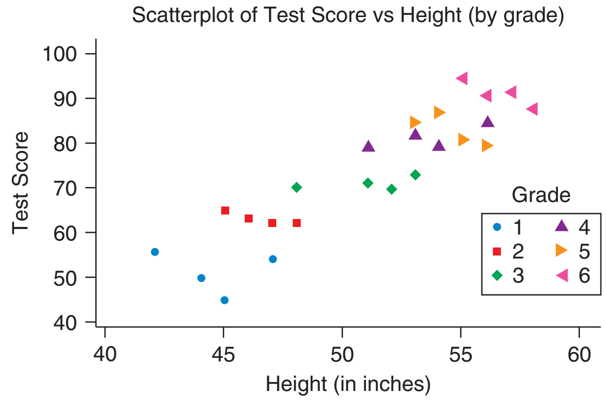


4.82 Heights and Test Scores

a. The figure shows hypothetical data for a group of children. By looking at the figure, state whether the correlation between height and test score is positive, negative, or near zero.

b. The shape and color of the each marker show what grade these children were in at the time they took the test. Look at the six different groupings (for grades 1, 2, 3, 4, 5, and 6) and decide whether the correlation (the answer to part a) would stay the same if you controlled for grade (that is, if you looked only within specific grades).

c. Suppose a school principal looked at this scatterplot and said, "This means that taller students get better test scores, so we should give more assistance to shorter students." Do the data support this conclusion? Explain. If yes, say why. If no, give another cause for the association.



4.83 Prices at Target and Whole Foods The price (in dollars) is given for some foods at Target and at Whole Foods. Assume the price at Target is the predictor. (Source: StatCrunch: organic food price comparison fall 2011. Owner: kerrypaulson)

Food	Target	Whole
bananas/1 lb	0.79	0.99
grape tomato	4.49	3.99
russet potato	4.49	4.99
red potato	1.56	2.00
yellow onion	4.94	1.99
red tomato	4.99	3.99
kiwi each	0.69	0.79
green grapes	2.50	3.69
red grapes	2.50	2.99
raspberries/.5 lb	3.09	3.99
strawberries/1 lb	3.04	6.99
gala apples	1.50	2.49
romaine lettuce	3.29	3.50
cauliflower	3.99	2.99
broccoli	3.04	2.99
celery/1 lb	2.19	1.99
celery hearts	2.29	2.99
green onion	1.99	0.99
baby carrots/2 lb	3.09	3.99
baby carrots/1 lb	1.64	1.99
stick carrots	1.90	0.99
romaine/5 oz	3.99	3.99
baby spring mix	3.99	3.99
herb salad	4.19	3.99
spinach	4.19	3.99
green pepper	1.89	3.49
cucumber	2.04	2.99
can black bean	1.79	2.19
Cascadian cereal	3.29	4.99
nature crackers	2.99	3.39

a. Make a graph and report whether the trend is linear. If the trend is not linear, comment on what it shows, and do not go on to subsequent parts.

b. If the graph is linear, find the equation of the best-fit line, and put the line into the graph.

c. Interpret the slope in context.

d. Either interpret the intercept or explain why it is not appropriate to do so.

4.84 Age and Happiness Happiness ratings were from 1 (least happy) to 100 (most happy). Data are at the text's website. Use age as the independent variable. (Source: StatCrunch: Responses to Happiness Survey. Owner: Webster West)

a. Make a graph and report whether the trend is linear. If the trend is not linear, comment on what it shows, and do not go on to subsequent parts.

b. If the graph is linear, do a complete analysis.

4.85 Tree Heights Loggers gathered information about some trees. The diameter is in inches, the height is in feet, and the volume of the wood is in cubic feet. Loggers are interested in whether they can estimate the volume of the tree given any single dimension. Which is the better predictor of volume: the diameter or the height? Data are at the text's website.

4.86 Salary and Education Does education pay? The salary per year in dollars, the number of years employed (YrsEm), and the number of years of education after high school (Educ) for the employees of a company were recorded. Determine whether number of years employed or number of years of education after high school is a better predictor of salary. Explain your thinking. Data are at the text's website. (Source: Minitab File)

4.87 Film Budgets and Grosses Movie studios exert much effort trying to predict how much money their movies will make. One possible predictor is the amount of money spent on the production of the movie. The table shows the budget and the amount of money made worldwide for the ten movies with the highest profits. The budget (amount spent on production) and gross are in millions of dollars. Make a scatterplot and comment on what you see. If appropriate, find, report, and interpret the regression line. If it is not appropriate to do so, explain why. (Source: www.the-numbers.com)

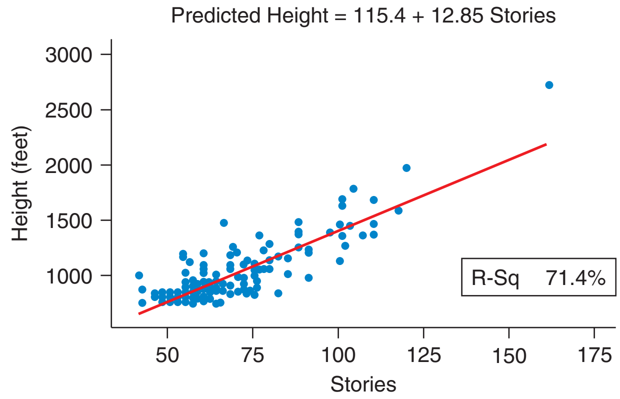
Film	Budget	Gross
Avatar	237	2784
Titanic	200	1843
Harry Potter and the Deathly Hallows: Part II	125	1328
Lord of the Rings: The Return of the King	94	1141
Jurassic Park	63	924
The Lion King	79	953
Shrek 2	70	919
Star Wars Episode 1: Phantom Menace	115	1007
Star Wars Episode 4: A New Hope	11	798
ET: The Extra-Terrestrial	10	793

4.88 Gas Mileage of Cars The table gives the number of miles per gallon in the city and on the highway for the coupes and compact cars reported to have the best gasoline mileage, according to autobytel.com. Make a scatterplot, using the city mileage as the predictor. Find the equation of the regression line for predicting the number of miles per gallon (mpg) on the highway from the number of miles per gallon in the city. Use the equation to predict the highway mileage from a city mileage of 100 mpg. Also find the coefficient of determination and explain what it means. Finally, state which car is an outlier; find the coefficient of determination without the outlier, and comment on it. (Source: http://www.autobytel.com/top-10-cars/best-gas-mileage-cars/coupes)

Car	City	Highway
Chevy Spark EV	128	109
Fiat 500e	122	108
Honda Fit EV	132	105
Nissan Leaf	129	102
Ford Focus Electric	110	99
Mitsubishi i-MiEV	126	99
Chevy Volt	101	93
Smart fortco ED	122	93
Ford C-Max Energi	108	92
VW Jetta Hybrid	42	48

4.89 Tall Buildings The scatterplot shows information about the world's tallest 169 buildings. "Stories" means "Floors."

a. What does the trend tell us about the relationship between stories and height (feet)?



b. The regression line for predicting the height (in feet) from the number of stories is shown above the graph. What height would you predict for a building with 100 stories?

c. Interpret the slope.

d. What, if anything, do we learn from the intercept?

e. Interpret the coefficient of determination.

(This data set is available at the text's website, and it contains several other variables. You might want to check to see whether the year the building was constructed is related to its height, for example.)

4.90 Bar-Passing Rate To become a lawyer, you must pass the bar exam in your state, and law schools often attract students by advertising their bar-passing rate: the percentage of their graduates who pass the bar exam. What qualities make for a good law school? You might think that a low student/faculty ratio was good; this would mean that the school typically has small class sizes.

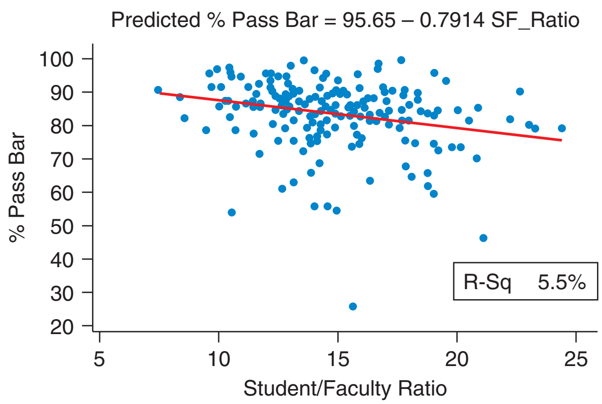
a. The scatterplot shows the bar-passing rate against the student/faculty ratio for a large number of law schools in the United States. What does the trend tell us about the role of the student/faculty ratio?

b. The regression line for predicting the bar-passing rate is shown above the graph. What bar-passing rate would you predict for a school with a student/faculty ratio of 12?

c. Interpret the slope.

d. What, if anything, do we learn from the intercept?

e. Interpret the coefficient of determination.



(This data set is available at the text's website, and other variables are also shown, such as the minimum score on the LSAT (the Law School Admission Test), and the minimum GPA for the students accepted at the law schools. Several other variables give higher coefficients of determination. You could also discover what named law schools have the low outliers, such as the bar-passing rate of 26%.)

For 4.91–4.94 show your points in a rough scatterplot and give the coordinates of the points.

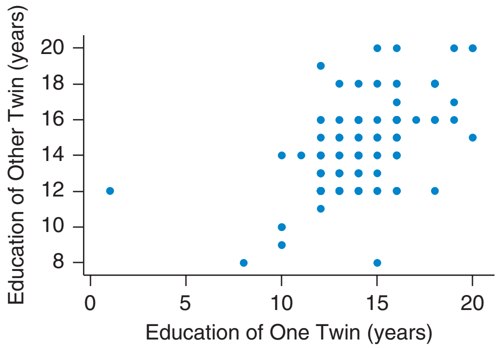
* **4.91** Construct a small set of numbers with at least three points with a perfect positive correlation of 1.00.

* **4.92** Construct a small set of numbers with at least three points with a perfect negative correlation of −1.00.

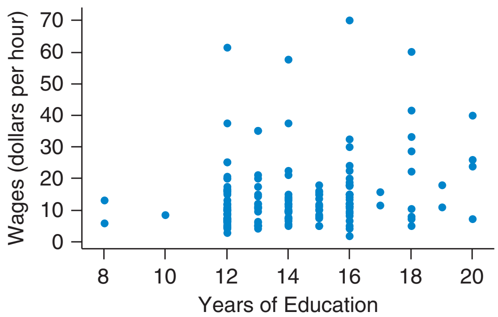
* **4.93** Construct a set of numbers (with at least three points) with a strong negative correlation. Then add one point (an influential point) that changes the correlation to positive. Report the data and give the correlation of each set.

* **4.94** Construct a set of numbers (with at least three points) with a strong positive correlation. Then add one point (an influential point) that changes the correlation to negative. Report the data and give the correlation of each set.

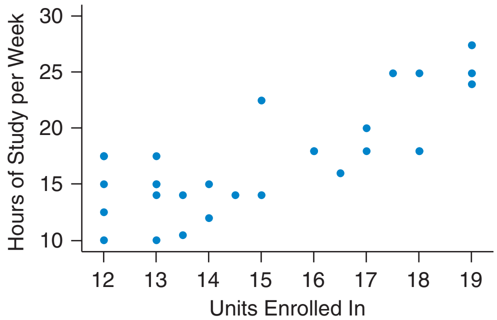
4.95 The figure shows a scatterplot of the educational level of twins. Describe the scatterplot. Explain the trend and mention any unusual points. (Source: www.stat.ucla.edu)



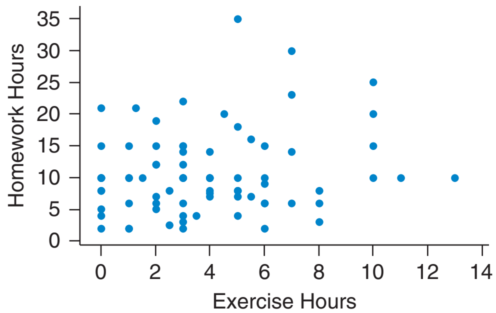
4.96 Wages and Education The figure shows a scatterplot of the wages and educational level of some people. Describe what you see. Explain the trend and mention any unusual points. (Source: www.stat.ucla.edu)



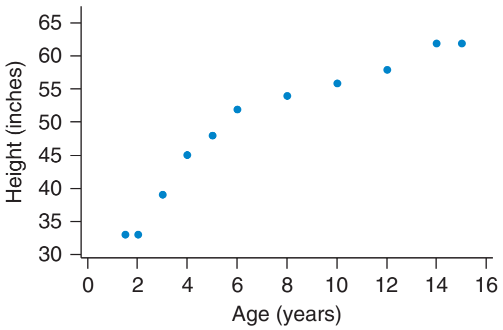
4.97 Do Students Taking More Units Study More Hours? The figure shows the number of units that students were enrolled in and the number of hours (per week) that they reported studying. Do you think there is a positive trend, a negative trend, or no noticeable trend? Explain what this means about the students.



4.98 Hours of Exercise and Hours of Homework The scatterplot shows the number of hours of exercise per week and the number of hours of homework per week for some students. Explain what it shows.

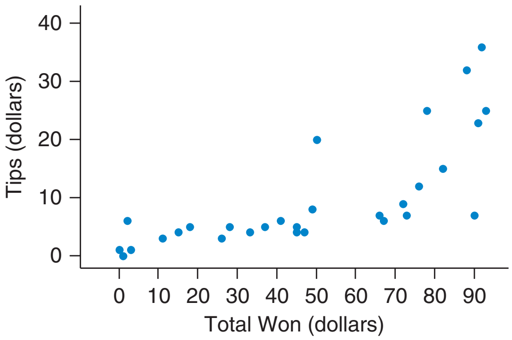


4.99 Children's Ages and Heights The figure shows information about the ages and heights of several children. Why would it not make sense to find the correlation or to perform linear regression with this data set? Explain.



4.100 Blackjack Tips The figure shows the amount of money won by people playing blackjack and the amount of tips they gave to the dealer (who was a statistics student), in dollars.

Would it make sense to find a correlation for this data set? Explain.



* **4.101 Decrease in Cholesterol** A doctor is studying cholesterol readings in his patients. After reviewing the cholesterol readings, he calls the patients with the highest cholesterol readings (the top 5% of readings in his office) and asks them to come back to discuss cholesterol-lowering methods. When he tests these patients a second time, the average cholesterol readings tended to have gone down somewhat. Explain what statistical phenomenon might have been partly responsible for this lowering of the readings.

* **4.102 Test Scores** Suppose that students who scored much lower than the mean on their first statistics test were given special tutoring in the subject. Suppose that they tended to show some improvement on the next test. Explain what might cause the rise in grades other than the tutoring program itself.

GUIDED EXERCISES

g **4.47 How is the time of a flight related to the distance of the flight?** The table gives the distance from Boston to each city (in thousands of miles) and gives the time for one randomly chosen, commercial airplane to make that flight. Do a complete regression analysis that includes a scatterplot with the line, interprets the slope and intercept, and predicts how much time a nonstop flight from Boston to Seattle would take. The distance from Boston to Seattle is 3000 miles.

City	Distance (1000s of miles)	Time (hours)
St. Louis	1.141	2.83
Los Angeles	2.979	6.00
Paris	3.346	7.25
Denver	1.748	4.25
Salt Lake City	2.343	5.00
Houston	1.804	4.25
New York	0.218	1.25

Step 1 ▶ Make a scatterplot
Be sure that *distance* is the *x*-variable and *time* is the *y*-variable, because we are trying to predict time from distance. Graph the line using technology. See the TechTips starting on page 225.

Step 2 ▶ Is the linear model appropriate?
Does it seem that the trend is linear, or is there a noticeable curve?

Step 3 ▶ Find the equation
Find the equation for predicting time (in hours) from miles (in thousands).

Step 4 ▶ Slope
Interpret the slope in context.

Step 5 ▶ Intercept
Interpret the intercept in context. Although there are no flights with a distance of zero, try to explain what might cause the added time that the intercept represents.

Step 6 ▶ Time to Seattle
Answer the question. About how long should it take to fly nonstop from Boston to Seattle?

g **4.73 Test Scores** Assume that in a political science class, the teacher gives a midterm exam and a final exam. Assume that the association between midterm and final scores is linear. The summary statistics have been simplified for clarity.

Midterm:	Mean = 75,	Standard deviation = 10
Final:	Mean = 75,	Standard deviation = 10

Also, $r = 0.7$ and $n = 20$.

For a student who gets 95 on the midterm, what is the predicted final exam grade? Assume the graph is linear.

Step 1 ▶ Find the equation of the line to predict the final exam score from the midterm score.
Standard form: $y = a + bx$

a. First find the slope: $b = r\left(\dfrac{s_{\text{final}}}{s_{\text{midterm}}}\right)$

b. Then find the y-intercept, a, from the equation

$$a = \bar{y} - b\bar{x}$$

c. Write out the following equation:

$$\text{Predicted } y = a + bx$$

However, use "Predicted Final" instead of "Predicted y" and "Midterm" in place of x.

Step 2 ▶ Use the equation to predict the final exam score for a student who gets 95 on the midterm.

Step 3 ▶ Your predicted final exam grade should be less than 95. Why?

CHECK YOUR TECH

Verifying Minitab Output for Correlation and Regression

Husband	Wife
20	20
30	30
40	25

At the left, data are given for the ages at which husbands and wives married. The data were simulated to make the calculations easier.

We used Minitab to find the correlation and regression equation given in Figure A, using the husband's age (x) to predict his wife's age (y).

▶ FIGURE A

```
Pearson correlation of Husband and Wife = 0.500
The regression equation is
Wife = 17.5 + 0.250 Husband
```

From Figure B: for the husband, $\bar{x} = 30$ and $s_x = 10$; for the wife, $\bar{y} = 25$ and $s_y = 5$.

$$r = \frac{\sum z_x z_y}{n - 1} \text{ and } z_x = \frac{x - \bar{x}}{s_x} \text{ and } z_y = \frac{y - \bar{y}}{s_y}$$

▶ FIGURE B

```
Variable   N   Mean   StDev
Husband    3  30.00   10.00
Wife       3  25.00    5.00
```

QUESTION Using the summary statistics provided in Figure B, verify the output given in Figure A by following the steps below.

SOLUTION

Step 1 ▶ Fill in the missing numbers (indicated by the blanks) in the table.

x	$x - \bar{x}$	z_x	y	$y - \bar{y}$	z_y	$z_x z_y$
20	$20 - 30 = -10$	$-10/10 = -1$	20	$20 - 25 = -5$	$-5/5 = -1$	$(-1) \times (-1) = 1$
30	$30 - 30 = 0$	$0/10 = \_$	30	$30 - \_ = \_$	$5/5 = \_$	$0 \times (\_) = \_$
40	$40 - 30 = 10$	$\_/\_ = \_$	25	$\_ - \_ = \_$	$0/\_ = \_$	$\_ \times (\_1) = \_$

Step 2 ▶ Add the last column to get $\sum z_x z_y = \_$.

Step 3 ▶ Find the correlation, and check it with the output.

$$r = \frac{\sum z_x z_y}{n - 1} = \frac{\_}{3 - 1} = \frac{\_}{2} = \_$$

Step 4 ▶ Find the slope: $b = r\frac{s_y}{s_x} = r\frac{5}{10} = \_\_$.

Step 5 ▶ Find the *y*-intercept: $a = \bar{y} - b\bar{x} = 25 - 30b = \_\_\_$.

Step 6 ▶ Finally, put together the equation:

$$y = a + bx$$
$$\text{Predicted Wife} = a + b \text{ Husband}$$

and check the equation with the Minitab output. The equations should match.

TechTips

General Instructions for All Technology

Upload data from the text's website, or enter data manually using two columns of equal length. Refer to TechTips in Chapter 2 for a review of entering data. Each row represents a single observation, and each column represents a variable. All technologies will use the example that follows.

EXAMPLE ▶ Analyze the six points in the data table with a scatterplot, correlation, and regression. Use heights (in inches) as the x-variable and weight (in pounds) as the y-variable.

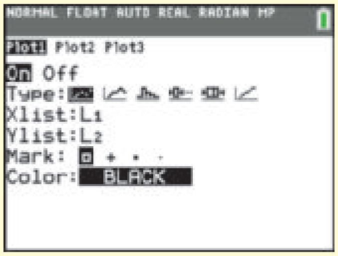
Height	Weight
61	104
62	110
63	141
64	125
66	170
68	160

TI-84

These steps assume you have entered the heights into **L1** and the weights into **L2**.

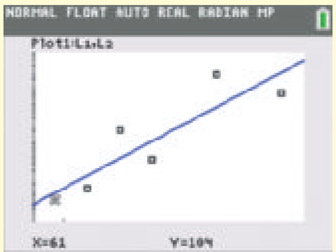
Making a Scatterplot

1. Press **2ND**, **STATPLOT** (which is the button above **2ND**), **4**, and **ENTER**, to turn off plots made previously.
2. Press **2ND**, **STATPLOT**, and **1** (for Plot1).
3. Refer to Figure 4A: Turn on **Plot1** by pressing **ENTER** when **On** is highlighted.



▲ **FIGURE 4A** TI-84 Plot1 Dialogue Screen

4. Use the arrows on the keypad to get to the scatterplot (first of the six plots) and press **ENTER** when the scatterplot is highlighted. Be careful with the **Xlist** and **Ylist**. To get **L1**, press **2ND** and **1**. To get **L2**, press **2ND** and **2**.
5. Press **GRAPH, ZOOM** and **9** (**Zoomstat**) to create the graph.
6. Press **TRACE** to see the coordinates of the points, and use the arrows on the keypad to go to other points. Your output will look like Figure 4B, but without the line.



▲ **FIGURE 4B** TI-84 Plot with Line

7. To get the output with the line in it, shown in Figure 4B: **STAT**, **CALC**, 8:LinReg(a + bx), **L1**, **ENTER**, **L2**, **ENTER**, **ENTER**, **Y1** (You get the **Y1** by pressing **VARS**, **Y-VARS**, 1: **Function**, 1:**Y1**), **ENTER**, **ENTER**.
8. Press **GRAPH**, **ZOOM** and **9**.
9. Press **TRACE** to see the numbers, and use the arrows on the keypad to get to other numbers.

Finding the Correlation and Regression Equation Coefficients

Before finding the correlation, you must turn the diagnostics on, as shown here.

Press **2ND**, **CATALOG**, and scroll down to **DiagnosticOn** and press **ENTER** twice. The diagnostics will stay on unless you **Reset** your calculator or change the batteries.

1. Press **STAT**, choose **CALC**, and **8** (for LinReg (a + bx)).
2. Press **2ND L1** (or whichever list is X, the predictor), press **ENTER**, press **2ND L2** (or whichever list is Y, the response), and press **ENTER, ENTER, ENTER.**

Figure 4C shows the output.

▲ **FIGURE 4C** TI-84 Output

MINITAB

Making a Scatterplot

1. **Graph > Scatterplot**
2. Leave the default **Simple** and click **OK.**
3. Double click the column containing the weights so that it goes under the **Y Variables**. Then double click the column containing the heights so that it goes under the **X Variables**.
4. Click **OK**. After the graph is made, you can edit the labels by clicking on them.

Finding the Correlation

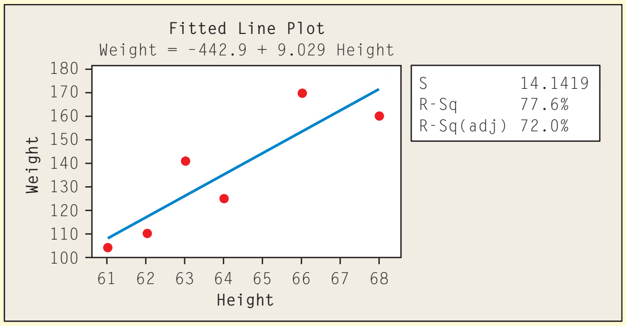
1. **Stat > Basic Statistics > Correlation**
2. Double click both the predictor column and the response column (in either order).
3. Click **OK**. You will get 0.881.

Finding the Regression Equation Coefficients

1. **Stat > Regression > Regression > Fit Regression Model**
2. Put in the **Responses:** (y) and **Continuous Predictors:** (x) columns.
3. Click **OK**. You may need to scroll up to see the regression equation. It will be easier to understand if you have put in labels for the columns, such as "Height" and "Weight." You will get: Weight = −443 + 9.03 Height.

To Display the Regression Line on a Scatterplot

1. **Stat > Regression > Fitted Line Plot**
2. Double click the **Response (Y)** column and then double click the **Predictor (X)** column.
3. Click **OK**. Figure 4D shows the fitted line plot.



▲ **FIGURE 4D** Minitab Fitted Line Plot

EXCEL

Making a Scatterplot

1. Select (highlight) the two columns containing the data, with the predictor column to the *left* of the response column. You may include the labels at the top or not include them.
2. Click **INSERT**, in **Charts** click the picture of a scatterplot, and click the upper left option shown here:

3. Note that the lower left corner of the chart is not at the origin, (0, 0). If you want to zoom in or out on the data by changing the minimum value for an axis, right-click on the axis numbers, click **Format axis**, in **AXIS OPTIONS** change the **Minimum** to the desired value. You may want to do this twice: once for the *x*-axis and once for the *y*-axis. Then close the **Format Axis** menu.
4. When the chart is active (click on it), then click on **DESIGN**, then **Add Chart Elements**, then **Axis Titles**, and **Chart Title** to add appropriate labels. After the labels are added, you can click on them to change the spelling or add words. Delete the labels on the right-hand side, such as **Series 1**, if you see any.

Finding the Correlation

1. Click on **DATA**, click on **Data Analysis**, select **Correlation**, and click **OK**.
2. For the **Input Range**, select (highlight) both columns of data (if you have highlighted the labels as well as the numbers, you must also click on the **Labels in first row**).
3. Click **OK**. You will get 0.880934.

(Alternatively, just click the f_x button, for **category** choose **statistical**, select **CORREL**, click **OK**, and highlight the two columns containing the numbers, one at a time. The correlation will show up on the dialogue screen, and you do *not* have to click **OK**.)

Finding the Coefficients of the Regression Equation

1. Click on **DATA**, **Data Analysis, Regression**, and **OK**.
2. For the **Input Y Range**, select the column of numbers (not words) that represents the response or dependent variable. For the **Input X Range**, select the column of numbers that represents the predictor or independent variable.
3. Click **OK**.

A large summary of the model will be displayed. Look under **Coefficients** at the bottom. For the **Intercept** and the slope (next to **XVariable1**), see Figure 4E, which means the regression line is

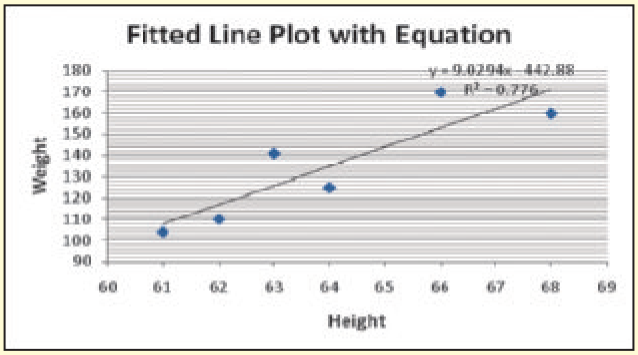
$$y = -442.9 + 9.03x$$

	Coefficients
Intercept	-442.882
X Variable 1	9.029412

▲ **FIGURE 4E** Excel Regression Output

To Display the Regression Line on a Scatterplot

4. After making the scatterplot, click on the scatterplot and click **Design**. In the **Chart Layouts** group, click the triangle to the right of **Quick Layout**. Choose Layout 9 (the option in the lower right portion, which shows a line in it and also *fx*).
   Refer to Figure 4F.



▲ **FIGURE 4F** Excel Fitted Line Plot with Equation

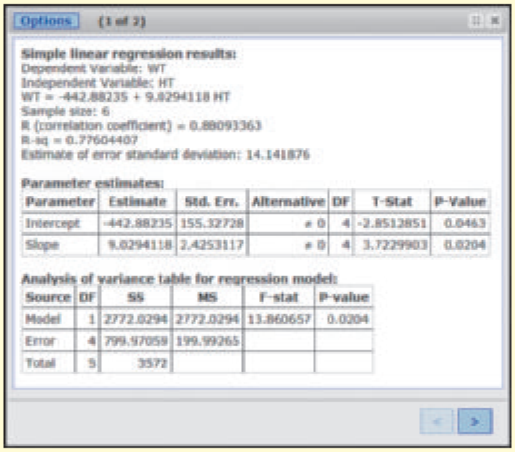
STATCRUNCH

Making a Scatterplot

1. **Graph > Scatter plot**
2. Select an **X variable** and a **Y variable** for the plot.
3. Click **Compute!** to construct the plot.
4. To copy the graph, click **Options** and **Copy**.

Finding the Correlation and Coefficients for the Equation

1. **Stat > Regression > Simple Linear**
2. Select the **X variable** and **Y variable** for the regression.
3. Click **Compute!** to view the equation and numbers, which are shown in Figure 4G.



Options (1 of 2)

Simple linear regression results:
Dependent Variable: WT
Independent Variable: HT
WT = -442.88235 + 9.0294118 HT
Sample size: 6
R (correlation coefficient) = 0.88093363
R-sq = 0.77604407
Estimate of error standard deviation: 14.141876

Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	P-Value
Intercept	-442.88235	155.32728	≠ 0	4	-2.8512851	0.0463
Slope	9.0294118	2.4253117	≠ 0	4	3.7229903	0.0204

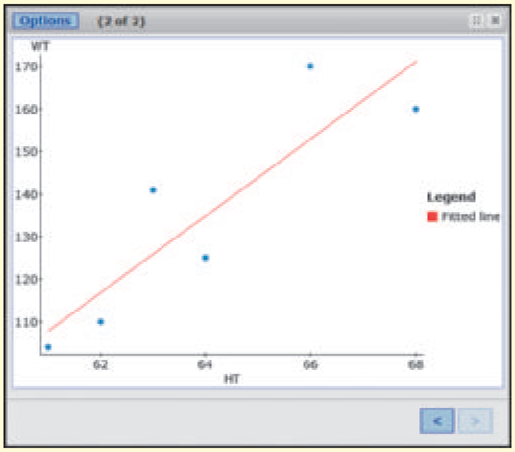
Analysis of variance table for regression model:

Source	DF	SS	MS	F-stat	P-value
Model	1	2772.0294	2772.0294	13.860657	0.0204
Error	4	799.97059	199.99265		
Total	5	3572			

▲ **FIGURE 4G** StatCrunch Regression Output

Plotting the Regression Line on a Scatterplot

1. **Stat > Regression > Simple Linear**
2. Select your columns for X and Y.
3. Click **Compute!**
4. Click the **>** in the lower right corner (see Figure 4G).
5. To copy the graph, click **Options** and **Copy**.



▲ **FIGURE 4H** StatCrunch Fitted Line Plot

5 Modeling Variation with Probability

THEME

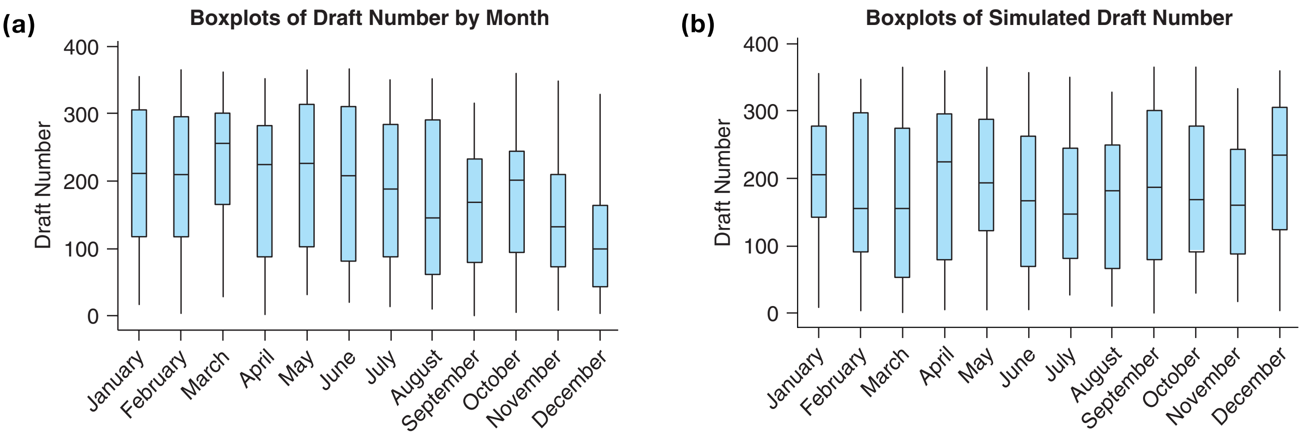
Probabilities are long-run relative frequencies used to describe processes where the outcome depends on chance, such as flipping a coin or rolling a die. Theoretical probabilities are based on specific assumptions (usually based on a theory) about the chance process. Empirical probabilities are based on observation of the actual process. The two types of probabilities are closely related, and we need both of them to analyze samples of data in the real world.

In 1969, the United States was fighting the Vietnam War and drafting men to serve in the military. To determine who was chosen, government officials wrote the days of the year (January 1, January 2, and so on) on capsules. The capsules were placed in a large container and mixed up. They were then drawn out one at a time. The first date chosen was assigned the rank 1, the second date was assigned the rank 2, and so on. Men were drafted on the basis of their birthday. Those whose birthday had rank 1 were drafted first, then those whose birthday had rank 2, and so on until the officials had enough men.

Although the officials thought that this method was random, some fairly convincing evidence indicates that it was not (Starr 1997). Figure 5.1a shows boxplots with the actual ranks for each month. Figure 5.1b shows what the boxplots might have looked like if the lottery had been truly random. In Figure 5.1b, each month has roughly the same rank. However, in Figure 5.1a, a few months had notably lower ranks than the other months. Bad news if you were born in December—you were more likely to be called up first.

What went wrong? The capsules, after having dates written on them, were clustered together by month as they were put into the tumbler. But the capsules weren't mixed up enough to break up these clusters. The mixing wasn't adequate to create a truly random mix.

It's not easy to generate true randomness, and humans have a hard time recognizing random events when they see them. Probability gives us a tool for understanding randomness. Probability helps us answer the question "How often does something like this happen by chance?" By answering that question, we create an important link between our data and the real world. In previous chapters, you learned how to organize, display, and summarize data to see patterns and trends. Probability is a vital tool because it gives us the ability to generalize our understanding of data to the larger world. In this chapter, we'll explore issues of randomness and probability: What is randomness? How do we measure it? And how do we use it?



▲ **FIGURE 5.1** Boxplots of **(a)** actual Vietnam draft numbers by month, and **(b)** what might have happened if the draft had really been random.

CASE STUDY

SIDS or Murder?

In November 2000, Sally Clark was convicted in England of killing her two children. Her children had died several years apart, and the initial diagnosis was sudden infant death syndrome (SIDS). SIDS is the abrupt, unexplained death of an infant under

1 year of age. It is estimated that in the United States, about 5500 infants die of SIDS every year. Although some risk factors have been identified (most are related to the mother's health during pregnancy), the cause or causes are as yet unknown.

Clark was convicted of murder on the basis of the expert testimony of Sir Roy Meadow, a prominent British physician who was an expert on SIDS. Dr. Meadow quoted a published statistic: The probability of an infant dying of SIDS is $1/8543$. That is, in the U.K., one child in every 8543 dies of SIDS. Dr. Meadow concluded that the probability of two children in the same family dying of SIDS was therefore $(1/8543) \times (1/8543)$, or about 1 in 73 million. The jury agreed that this event—two children in the same family dying of SIDS—was so unlikely that they convicted Ms. Clark of murder, and she was sent to prison.

But then in 2003, Sally Clark's conviction was overturned and she was released from prison. Dr. Meadow was accused of professional misconduct. Why did Dr. Meadow multiply the two probabilities together? Why was the verdict overturned? We will answer these questions at the end of the chapter.

SECTION 5.1

What Is Randomness?

What exactly is randomness? You might see a definition similar to: "Haphazard; for no apparent reason or purpose." You probably use the word to describe events that seem to happen with no purpose at all. Sometimes the word is used to describe things that have no apparent pattern. However, in statistics the word *random* has a more precise meaning, as you will see.

We can do a small experiment to compare our natural understanding of randomness with real-life randomness. We asked a student to imagine flipping a coin and to write down the resulting heads (H) and tails (T). To compare this to a real random sequence, we flipped an actual coin 20 times and recorded the results. Which of the sequences below do you think is real, and which is invented by the student?

H	H	T	H	T	H	H	T	T	T	H	T	H	H	H	T	H	T	T	H
T	T	T	H	H	H	T	T	T	T	T	H	H	H	H	T	T	H	T	H

The first row of results is the one made up by the student, and the second row records the results of actually tossing a coin 20 times. Is it possible to tell by comparing the two sequences? Not always, but this time the student did something that most people do in this situation. Most of the time, he wrote very short "streaks." (A streak is any sequence of consecutive heads or tails. TT is a streak of two tails.) Only once did he write as many as three heads or three tails in a row. It's as if, after writing three heads in a row, he thought, "Wait! If I put a fourth, no one will believe it is random."

However, the real, truly random sequence had one streak of five tails (beginning at the seventh flip) and another streak with four heads. These long streaks are examples of the way chance creates things that look like a pattern or structure but really are not.

KEY POINT People are not good at identifying truly random samples or random experiments, so we need to rely on outside mechanisms such as coin flips or random number tables.

Real randomness is hard to achieve without help from a computer or some other randomizing device. If a computer is not available to generate random numbers, another useful approach is to use a random number table. (An example is provided in Appendix A of this text.) A random number table provides a sequence of digits from 0 to 9 in a random order. Here, **random** means that no predictable pattern occurs and that no digit is more likely to appear than any other. (Of course, if you use the same table often enough, it might seem predictable to you, but to an outsider, it will not be predictable.)

For example, if we are doing a controlled experiment, we might assign each subject in our study a number from this table as the subjects come into our office. The odd-numbered subjects might then go into the Control group, and the even-numbered subjects might go into the Treatment group.

To use a random number table to simulate coin flipping, we assign a number to each outcome. For example, suppose we let even numbers (0, 2, 4, 6, 8) represent tails and let odd numbers (1, 3, 5, 7, 9) represent heads. Now choose any row of the table you wish and any column. For example, we arbitrarily chose column 11 and row 30 because that's the date of the day we wrote this paragraph (November 30, or 11/30). Read off the next 20 digits, but translate them into "heads" and "tails." What's the longest streak you get?

Line						
28	3 1 4 9 8	8 5 3 0 4	2 2 3 9 3	2 1 6 3 4	3 4 5 6 0	7 7 4 0 4
29	9 3 0 7 4	2 7 0 8 6	6 2 5 5 9	8 6 5 9 0	1 8 4 2 0	3 3 2 9 0
30	9 0 5 4 9	5 3 0 9 4	7 6 2 8 2	5 3 1 0 5	4 5 5 3 1	9 0 0 6 1
31	1 1 3 7 3	9 6 8 7 1	3 8 1 5 7	9 8 3 6 8	3 9 5 3 6	0 8 0 7 9
32	5 2 0 2 2	5 9 0 9 3	3 0 6 4 7	3 3 2 4 1	1 6 0 2 7	7 0 3 3 6

◀ **TABLE 5.1** Lines 28–32 (indicated at the left side of the table) from the random number table in Appendix A. The red 7 in the 11th column and the 30th line, or row, is our starting point.

EXAMPLE 1 Simulating Randomness

Let's play a game. You roll a six-sided die until the first 6 appears. You win one point for every roll. You pass the die to your opponent, who also rolls until the first 6 appears. Whoever has the most points, wins.

QUESTION Simulate rolling the die until a 6 appears. Use Table 5.1, and start at the very first entry of line 28 in Table 5.1. How many rolls did it take?

SOLUTION We'll let the digits 1 through 6 represent the outcome shown on the die. We'll ignore the digits 7, 8, 9, and 0. Starting at line 28, we get these random digits:

3, 1, 4, (ignore, ignore, ignore) 5, 3, (ignore), 4, 2, 2, 3, (ignore), 3, 2, 1, 6

CONCLUSION We rolled the die 12 times before the first 6 appeared.

TRY THIS! Exercise 5.1

Random number tables are a bit old-fashioned, though they still have their uses. Computers and calculators have built-in random number generators that come close to true randomness. Computer-generated random numbers are sometimes called pseudo-random numbers, because they are generated on the basis of a seed value—a number that starts the random sequence. If you always input the same seed number, you will always see the same sequence of pseudo-random numbers. However, it is not possible to predict what number will come next in the sequence, and in this sense the generated numbers are considered random. For most practical work (and certainly for everything we cover in this text), these pseudo-random numbers are as random as we need. (You should be aware, however, that not all statistics packages produce equally convincing sequences of random numbers.)

Achieving a convincing "randomness" is difficult. This is the lesson the government learned from its flawed Vietnam draft lottery, which was described in the chapter introduction. Quite often, statisticians are employed to check whether random processes, such as the way winners are selected for state lotteries, are truly random.

Empirical and Theoretical Probabilities

Probability is used to measure how often random events occur. When we say, "The probability of getting heads when you flip a coin is 50%," we mean that flipping a coin results in heads half the time (assuming that the outcome of the flips is random). This is sometimes called the "frequentist" approach to probability, because in it, probabilities are defined as relative frequencies.

We will examine two kinds of probabilities. These two kinds of probabilities are connected, as we will see later in this chapter. **Theoretical probabilities** are long-run relative frequencies. The theoretical probability is the relative frequency at which an event happens after *infinitely* many repetitions. When we say that a coin has a 50% probability of coming up heads, we mean that if it were possible to flip the coin *infinitely* many times, then exactly 50% of the flips would be heads.

> **Looking Back**
> **Relative Frequencies**
> Relative frequencies, introduced in Chapter 2, are the same as proportions.

Because we can't do *anything* infinitely many times, finding theoretical probabilities requires relying on theory. (That's why they're called theoretical!) As an example, a coin has two sides. My theory is that either side is equally likely to come up when I flip a coin. I don't know for a fact that this is true; I *assume* it is true. And so I conclude that the probability of seeing heads is 50%. However, I haven't actually done an experiment. I have simply reasoned on the basis of a theory about how coins behave.

Empirical probabilities, on the other hand, are relative frequencies based on an experiment or on observations of a real-life process. I toss a coin 10 times and get 6 heads. My empirical probability of getting heads is therefore $6/10 = 0.6$, or 60%.

Empirical and theoretical probabilities have some striking differences. Chief among them is that theoretical probabilities are always the same value; if we all agree on the theory, then we all agree on the values for the theoretical probabilities. Empirical probabilities, however, change with every experiment. Suppose I flip the exact same coin 10 times again, and this time I get 3 heads. Now my empirical probability of heads is 0.3. Empirical probabilities are themselves random and vary from experiment to experiment.

> **KEY POINT** Empirical probabilities tell us how often an event occurred in an actual set of experiments or observations. Theoretical probabilities are based on theory and tell us how many times an event would occur if an experiment were repeated *infinitely* many times.

Why Do We Need Both?

Theoretical probabilities are very abstract things. They measure what proportion of the time events would occur if an experiment were repeated *infinitely many times*. This means it is impossible to carry out an experiment that will provide the exact value of a theoretical probability; to do so would literally take forever. However, we can use empirical probabilities to *estimate* and to *test* theoretical probabilities.

Why do we need to estimate a theoretical probability? Sometimes, it is just too difficult to compute a theoretical probability. Instead, it may be adequate to get a good approximate value using an experiment that allows us to estimate how often a certain event might happen if we could repeat the experiment infinitely many times. On other occasions, the event for which we need a probability may be too complex for theory, and if we really need to know the probability, then running an experiment is the only way of finding a useful value.

Why do we need to test? We might not trust the theoretical probability value. For example, it might be based on assumptions that we are not sure are true. If done well, an experiment that produces an empirical probability can be used to verify or refute a theoretically derived value. In fact, much of the rest of this text will develop this theme.

If you do the "Let's Make a Deal" activity described at the end of this chapter, you will find empirical probabilities of winning for each of the two strategies offered to contestants: stay or switch. These empirical probabilities allow you to estimate the probability of winning, and they also allow you to test whether your own ideas about which strategy is best are correct.

Simulations are experiments used to produce empirical probabilities, because the investigators hope that these experiments simulate the situation they are examining. As you will see in Section 5.4, a mathematical discovery called the Law of Large Numbers tells us that if the relationship between the situation and our experiment is strong, then our empirical probability will be a good estimate of the true theoretical probability.

SECTION 5.2

Finding Theoretical Probabilities

We can use empirical probabilities as an estimate of theoretical probabilities, and if we do our experiment enough times, the estimate can be pretty good. But how good is pretty good? To understand how to use empirical probabilities to estimate and to verify theoretical probabilities, we will first learn how to calculate theoretical probabilities.

Facts about Theoretical Probabilities

Probabilities are always numbers between 0 and 1 (including 0 and 1). Probabilities can be expressed as fractions, decimals, or percents: 1/2, 0.50, and 50% are all used to represent the probability that a coin comes up heads.

Some values have special meanings. If the probability of an event happening is 0, then that event never happens. If you purchase a lottery ticket after all the prizes have been given out, the probability is 0 that you will win one of those prizes. If the probability of an event happening is 1, then that event always happens. If you flip a coin, the probability of a coin landing heads or tails is 1.

Another useful property to remember is that the probability that an event will *not* happen is 1 minus the probability that it will happen. If there is a 0.90 chance that it will rain, then there is a $1 - 0.9 = 0.10$ chance that it will not. If there is a 1/6 chance of rolling a "1" on a die, then there is a $1 - (1/6) = 5/6$ probability that you will not get a 1.

We call such a "not event" a **complement**. The complement of the event "it rains today" is the event "it does not rain today." The complement of the event "coin lands heads" is "coin lands tails." The complement of the event "a die lands with a 1 on top" is "a die lands with a 2, 3, 4, 5, or 6 on top"; in other words, the die lands with a number that is *not* 1 on top.

Events are usually represented by uppercase letters: A, B, C, and so on. For example, we might let A represent the event "it rains tomorrow." Then the notation P(A) means "the probability that it will rain tomorrow." In sentence form, the notation $P(A) = 0.50$ translates into English as "The probability that it will rain tomorrow is 0.50, or 50%."

A common misinterpretation of probability is to think that large probabilities mean that the event will certainly happen. For example, suppose your local weather reporter predicts a 90% chance of rain tomorrow. Tomorrow, however, it doesn't rain. Was the weather reporter wrong? Not necessarily. When the weather reporter says there is a 90% chance of rain, it means that on 10% of the days like tomorrow it does not rain. Thus a 90% chance of rain means that on 90% of all days just like tomorrow, it rains, but on 10% of those days it does not.

Summary of Probability Rules

Rule 1: A probability is always a number from 0 to 1 (or 0% to 100%) inclusive (which means 0 and 1 are allowed). It may be expressed as a *fraction,* a *decimal,* or a *percent*.

In symbols: For any event A,

$$0 \leq P(A) \leq 1$$

Rule 2: The probability that an event will not occur is 1 minus the probability that the event will occur. In symbols, for any event A,

$$P(\text{A does } \textit{not} \text{ occur}) = 1 - P(\text{A does occur})$$

The symbol A^c is used to represent the complement of A. With this notation, we can write Rule 2 as

$$P(A^c) = 1 - P(A)$$

Finding Theoretical Probabilities with Equally Likely Outcomes

In some situations, all of the possible outcomes of a random experiment occur with the same frequency. We call these situations "equally likely outcomes." For example, when you flip a coin, heads and tails are equally likely. When you roll a die, 1, 2, 3, 4, 5, and 6 are all equally likely. Assuming, of course, that the die is balanced correctly.

When we are dealing with equally likely outcomes, it is sometimes helpful to list all of the possible outcomes. A list that contains *all* possible (and equally likely) outcomes is called the **sample space**. We often represent the sample space with the letter S. An **event** is any collection of outcomes in the sample space. For example, the sample space S for rolling a die is the numbers 1, 2, 3, 4, 5, 6. The event "get an even number on the die" consists of the even outcomes in the sample space: 2, 4, and 6.

When the outcomes are equally likely, the probability that a particular event occurs is just the number of outcomes that make up that event, divided by the total number of equally likely outcomes in the sample space. In other words, it is the number of outcomes resulting in the event divided by the number of outcomes in the sample space.

Details

Precision Dice
The dice that casinos use are very different from the dice you use to play board games at home. Casino dice are precisely made. One casino claims, for instance, that each hole is exactly 17/1000 inch deep and filled with a material that is exactly the same density as the cube itself.

Summary of Probability Rules

Rule 3:

$$\text{Probability of A} = \text{P(A)} = \frac{\text{Number of outcomes in A}}{\text{Number of all possible outcomes}}$$

This is true *only* for equally likely outcomes.

For example, suppose 30 people are in your class, and one person will be selected at random by a raffle to win a prize. What is the probability that you will win? The sample space is the list of the names of the 30 people. The event A is the event that contains only one outcome: your name. The probability that you win is 1/30, because there is only 1 way for you to win and there are 30 different ways that this raffle can turn out. We write this using mathematical notation as follows:

$$\text{P(you win prize)} = 1/30$$

We can be even more compact:

Let *A* represent the event that you win the raffle. Then

$$\text{P(A)} = 1/30$$

One consequence of Rule 3 is that the probability that *something* in the sample space will occur is 1. In symbols, P(S) = 1. This is because

$$\text{P(S)} = \frac{\text{Number of outcomes in S}}{\text{Number of outcomes in S}} = 1$$

EXAMPLE 2 Ten Dice in a Bowl

Reach into a bowl that contains 5 red dice, 3 green dice, and 2 white dice (Figure 5.2). But assume that, unlike what you see in Figure 5.2, the dice have been well mixed.

FIGURE 5.2 Ten dice in a bowl.

QUESTION What is the probability of picking (a) a red die? (b) a green die? (c) a white die?

SOLUTIONS The bowl contains 10 dice, so we have ten possible outcomes. All are equally likely (assuming all the dice are equal in size, they are mixed up within the bowl, and we do not peek when choosing).

a. Five dice are red, so the probability of picking a red die is 5/10, 1/2, 0.50, or 50%. That is,

$$P(\text{red die}) = 1/2, \text{ or } 50\%$$

b. Three dice are green, so the probability of picking a green die is 3/10, or 30%. That is,

$$P(\text{green die}) = 3/10, \text{ or } 30\%$$

c. Two dice are white, so the probability of picking a white die is 2/10, 1/5, or 20%. That is,

$$P(\text{white die}) = 1/5, \text{ or } 20\%$$

Note that the probabilities add up to 1, or 100%, as they must.

TRY THIS! Exercise 5.11

Example 3 shows that it is important to make sure the outcomes in your sample space are equally likely.

EXAMPLE 3 Adding Two Dice

Roll two dice and add the outcomes. Assume each side of each die is equally likely to appear face up when rolled. Event A is the event that the sum of the two dice is 7.

QUESTION What is the probability of event A? In other words, find P(A).

SOLUTION This problem is harder because it takes some work to list all of the equally likely outcomes, which are shown in Table 5.2.

TABLE 5.2 Possible outcomes for two six-sided dice.

Die 1	1	1	1	1	1	1	2	2	2	2	2	2
Die 2	1	2	3	4	5	6	1	2	3	4	5	6
Die 1	3	3	3	3	3	3	4	4	4	4	4	4
Die 2	1	2	3	4	5	6	1	2	3	4	5	6
Die 1	5	5	5	5	5	5	6	6	6	6	6	6
Die 2	1	2	3	4	5	6	1	2	3	4	5	6

Table 5.2 lists 36 possible equally likely outcomes. Here are the outcomes in event A:

$$(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)$$

There are six outcomes for which the dice add to 7.

CONCLUSION The probability of rolling a sum of 7 is 6/36, or 1/6.

TRY THIS! Exercise 5.15

It is important to make sure that the outcomes in the sample space are equally likely. A common mistake when solving Example 3 is listing all the possible *sums*, instead of listing all the equally likely outcomes of the two dice. If we made that mistake here, our list of sums would look like this:

2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

This list has 11 sums, and only 1 of them is a 7, so we would incorrectly conclude that the probability of getting a sum of 7 is 1/11.

Why didn't we get the correct answer of 1/6? The reason is that the outcomes we listed—2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12—are not equally likely. For instance, we can get a sum of 2 in only one way: roll two "aces," for 1 + 1. Similarly, we have only one way to get a 12: roll two 6's, for 6 + 6.

However, there are six ways of getting a 7: (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), and (6, 1). In other words, a sum of 7 happens more often than a sum of 2 or a sum of 12. The outcomes 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 are not equally likely.

Usually it is not practical to list all the outcomes in a sample space—or even just those in the event you're interested in. For example, if you are dealing out 5 playing cards from a 52-card deck, the sample space has 2,598,960 possibilities. You really do not want to list all of those outcomes. Mathematicians have developed rules for counting the number of outcomes in complex situations such as these. These rules do not play an important role in introductory statistics, and we do not include them in this text.

! Caution

"Equally Likely Outcomes" Assumption

Just wishing it were true doesn't make it true. The assumption of equally likely outcomes is not always true. And if this assumption is wrong, your theoretical probabilities will not be correct.

Combining Events with "AND" and "OR"

As you saw in Chapter 4, we are often interested in studying more than one variable at a time. Real people, after all, often have several attributes we want to study, and we frequently want to know the relationship among these variables. The words AND and OR can be used to combine events into new, more complex events. The real people in Figure 5.3a, for example, have two attributes we decided to examine. They are either wearing a hat, or not. Also, they are either wearing glasses, or not. In the photo, the people who are wearing hats AND glasses are raising their hands.

Another way to visualize this situation is with a **Venn diagram**, as shown in Figure 5.3b. The rectangle represents the sample space, which consists of all possible outcomes if we were to select a person at random. The ovals represent events—for example, the event that someone is wearing glasses. The people who "belong" to *both* events are in the intersection of the two ovals.

The word **AND** creates a new event out of two other events. The probability that a randomly selected person in this photo is wearing a hat is 3/6, because three of the six people are wearing a hat. The probability that a randomly selected person

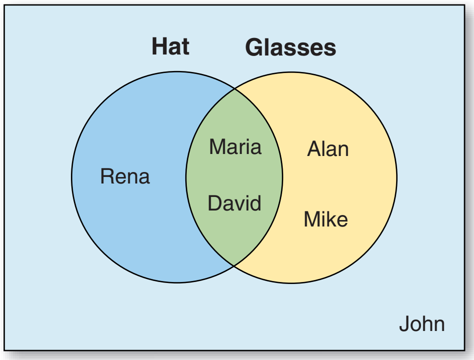
! Caution

Venn Diagrams

The areas of the regions in Venn diagrams have no numerical meaning. A large area does not contain more outcomes than a small area.

(a)

(b)



▲ **FIGURE 5.3** **(a)** Raise your hand if you are wearing glasses AND a hat. **(b)** The people wearing both glasses AND a hat (Maria and David) appear in the intersection of the two circles in this Venn diagram.

wears glasses is 4/6. The probability that a randomly selected person is wearing a hat AND glasses is 2/6, because only two people are in both groups. We could write this, mathematically, as

$$P(\text{wears glasses AND wears a hat}) = 2/6$$

KEY POINT The word AND creates a new event out of two events A and B. The new event consists of *only* those outcomes that are in *both* event A and event B.

In most situations, you will not have a photo to rely on. A more typical situation is given in Table 5.3, which records frequencies of two attributes for the people in a random sample of a recent U.S. census (www.census.gov). The two attributes are highest educational level and current marital status. ("Single" means never married. All other categories refer to the respondent's current status. Thus a person who was divorced but then remarried is categorized as "Married.")

▶ **TABLE 5.3** Education and marital status for 665 randomly selected U.S. residents. "Less HS" means did not graduate from high school.

Education Level	Single	Married	Divorced	Widow/Widower	Total
Less HS	17	70	10	28	125
High school	68	240	59	30	397
College or higher	27	98	15	3	143
Total	112	408	84	61	665

EXAMPLE 4 Education AND Marital Status

Suppose we will select a person at random from the collection of 665 people categorized in Table 5.3.

QUESTIONS

a. What is the probability that the person is married?

b. What is the probability that the person has a college education or higher?

c. What is the probability that the person is married AND has a college education or higher?

SOLUTIONS The sample space has a total of 665 equally likely outcomes.

a. In 408 of those outcomes, the person is married. So the probability that a randomly selected respondent is married is 408/665, or 61.4% (approximately).

b. In 143 of those outcomes, the person has a college education or higher. So the probability that the selected person has a college education or higher is 143/665, or 21.5%.

c. There are 665 possible outcomes. In 98 of them, the respondents both are married AND have a college degree or higher. So the probability that the selected person is married AND has a college degree is 98/665, or 14.7%.

TRY THIS! Exercise 5.21

! Caution

AND

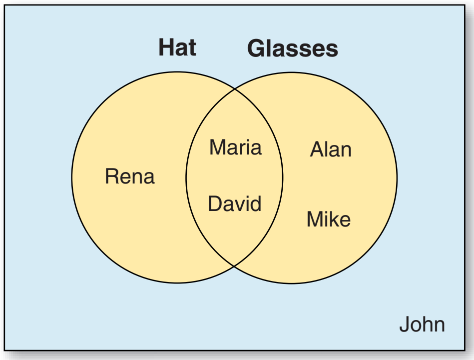
P(A AND B) will always be less than (or equal to) P(A) and also less than (or equal to) P(B). If this isn't the case, you've made a mistake!

Using "OR" to Combine Events

The people in Figure 5.4a were asked to raise their hands if they were wearing glasses OR wearing a hat. Note that people who are wearing both also have their hands raised. If we selected one of these people at random, the probability that this person is wearing

(a)

(b)



▲ **FIGURE 5.4** **(a)** Raise your hand if you are wearing a hat OR glasses. This photograph illustrates the inclusive OR. **(b)** In this Venn diagram, note the orange region for "raise your hand if you are wearing a hat OR glasses."

glasses OR wearing a hat would be 5/6, because we would count people who wear glasses, people who wear a hat, and people who wear both glasses AND a hat.

In a Venn diagram, OR events are represented by shading all relevant events. Here Mike, Rena, Maria, Alan, and David appear in the yellow area because each is wearing glasses OR wearing a hat.

The last example illustrates a special meaning of the word OR. This word is used slightly differently in mathematics and probability than you may use it in English. In statistics and probability, we use the **inclusive OR**. For example, the people in the photo shown in Figure 5.4a were asked to raise their hands if they had a hat OR glasses. This means that the people who raise their hands have a hat only, or have glasses only, or have both a hat AND glasses.

KEY POINT The word OR creates a new event out of the events A and B. The new event consists of all outcomes that are only in A, that are only in B, or that are in both.

EXAMPLE 5 OR with Marital Status

Again, select someone at random from Table 5.3.

QUESTION What is the probability the person is single OR married?

SOLUTION The event of interest occurs if we select a person who is married, a person who is single, or a person who is both married AND single. There are 665 possible equally likely outcomes. Of these, 112 are single and 408 are married (and none are both!). Thus there are $112 + 408 = 520$ people who are single OR married.

CONCLUSION The probability that the selected person is single OR married is 520/665, or 78.2%.

TRY THIS! Exercise 5.23

EXAMPLE 6 Education OR Marital Status

Select someone at random from Table 5.3 on page 238 (which is shown again in Table 5.4 on the next page).

QUESTION What is the probability that the person is married OR has a college degree?

SOLUTION Table 5.3 gives us 665 possible outcomes. The event of interest occurs if we select someone who is married, or someone who has a college degree, or someone who both is married AND has a college degree. There are 408 married people and 143 people with a college degree.

But wait a minute: There are not 408 + 143 different people who are married OR have a college degree—some of these people got counted twice! The people who are both married AND have a college degree were counted once when we looked at the married people, and they were counted a second time when we looked at the college graduates. We can see from Table 5.4 that 98 people *both* are married AND have a college degree. So we counted 98 people too many. Thus, we have 408 + 143 − 98 = 453 different outcomes in which the person is married or has a college degree. Table 5.4 is the same as Table 5.3 except for the added ovals, which are meant to be interpreted as in a Venn diagram.

▶ **TABLE 5.4** Here we reprint Table 5.3, with ovals for Married and for College added.

Education Level	Single	Married	Divorced	Widow/Widower	Total
Less HS	17	**70**	10	28	125
High school	68	**240**	59	30	397
College or higher	**27**	**98**	**15**	**3**	**143**
Total	112	**408**	84	61	665

The numbers in bold type represent the people who are married OR have a college degree. This Venn-like treatment emphasizes that one group (of 98 people) is in both categories and reminds us not to count them twice.

Another way to say this is that we count the 453 distinct outcomes by adding the numbers in the ovals in the table, but not adding any of them more than once:

$$70 + 240 + 98 + 27 + 15 + 3 = 453$$

CONCLUSION The probability that the randomly selected person is married OR has a college degree is 453/665, or 68.1%.

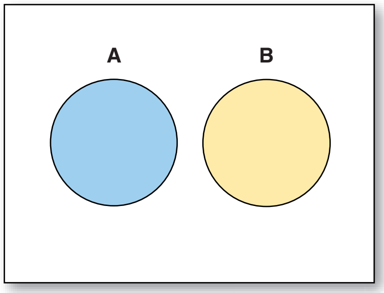
TRY THIS! Exercise 5.25

Mutually Exclusive Events

Did you notice that the first example of an OR (single OR married) was much easier than the second (married OR has a college degree)? In the second example, we had to be careful not to count some of the people twice. In the first example, this was not a problem. Why?

The answer is that in the first example, we were counting people who were married OR single, and no person can be in both categories. It is impossible to be simultaneously married AND single. When two events have no outcomes in common—that is, when it is impossible for both events to happen at once—they are called **mutually exclusive events**. The events “person is married” and “person is single” are mutually exclusive.

The Venn diagram in Figure 5.5 shows two mutually exclusive events. There is no intersection between these events; it is impossible for both event A AND event B to happen at once. This means that the probability that both events occur at the same time is 0.



◀ FIGURE 5.5 In a Venn diagram, two mutually exclusive events have no overlap.

EXAMPLE 7 Mutually Exclusive Events: Education and Marital Status

Imagine selecting a person at random from those represented in Table 5.4.

QUESTION Name two mutually exclusive events, and name two events that are not mutually exclusive. Remember that "marital status" means a person's *current* marital status.

SOLUTION For mutually exclusive events, we can choose any two columns or any two rows. It is impossible for someone's current status to be both divorced AND married. It is impossible for someone to have less than a high school education AND to have a high school education. The events "person has a HS education" and "person has less than a HS education" are mutually exclusive, because no one can be in both categories at once. The probability that a randomly selected person has a HS education AND has less than a HS education is 0. We could have chosen other pairs of events as well.

To find two events that are not mutually exclusive, find events that have outcomes in common. There are 30 people who have a HS education AND are widows/widowers. Therefore, these events are *not* mutually exclusive. The events "person has HS education" AND "person is a widow/ widower" are not mutually exclusive.

TRY THIS! Exercise 5.27

Summary of Probability Rules

Rule 4: The probability that event A happens OR event B happens is

(the probability that A happens) plus (the probability that B happens) minus (the probability that both A AND B happen)

If A and B are mutually exclusive events (for example, A is the event that the selected person is single, and B is the event that the person is married), then P(A AND B) = 0. In this case, the rule becomes simpler:

Rule 4a: If A and B are mutually exclusive events, the probability that event A happens OR event B happens is the sum of the probability that A happens and the probability that B happens.

Rule 4 in symbols:

$$\text{Always: } P(A \text{ OR } B) = P(A) + P(B) - P(A \text{ AND } B)$$

Rule 4a in symbols:

$$\text{Only if A and B are mutually exclusive: } P(A \text{ OR } B) = P(A) + P(B)$$

EXAMPLE 8 Rolling a Six-Sided Die

Roll a fair, six-sided die.

QUESTIONS

a. Find the probability that the die shows an even number OR a number greater than 4 on top.

b. Find the probability that the die shows an even number OR the number 5 on top.

SOLUTIONS

a. We could do this in two ways. First, we note that six equally likely outcomes are possible. The even numbers are (2, 4, 6) and the numbers greater than 4 are (5, 6). Thus the event "even number OR number greater than 4" has four different ways of happening: roll a 2, 4, 5, or 6. We conclude that the probability is 4/6.

The second approach is to use Rule 4. The probability of getting an even number is 3/6. The probability of getting a number greater than 4 is 2/6. The probability of getting both an even number AND a number greater than 4 is 1/6 (because the only way for this to happen is to roll a 6). So

$$\begin{aligned} P(\text{even OR greater than 4}) &= P(\text{even}) + P(\text{greater than 4}) - P(\text{even AND greater than 4}) \\ &= 3/6 + 2/6 - 1/6 \\ &= 4/6 \end{aligned}$$

b. $P(\text{even OR roll 5}) = P(\text{even}) + P(\text{roll 5}) - P(\text{even AND roll 5})$

It is impossible for the die to be both even AND a 5, because 5 is an odd number. So the events "get a 5" and "get an even number" are mutually exclusive. Therefore, we get

$$P(\text{even number OR a 5}) = 3/6 + 1/6 - 0 = 4/6$$

CONCLUSIONS

a. The probability of rolling an even number OR a number greater than 4 is 4/6 (or 2/3).

b. The probability of rolling an even number OR a 5 is 4/6 (or 2/3).

TRY THIS! Exercise 5.33

SECTION 5.3

Associations in Categorical Variables

Judging on the basis of our sample in Table 5.3, is there an association between marital status and having a college education? If so, we would expect the proportion of married people to be different for those who had a college education and those who did not have a college education. (Perhaps we would find different proportions of marital status for each category of education.)

In other words, if there is an association, we would expect the probability that a randomly selected college-educated person is married to be different from the probability that a person with less than a college education is married.

Conditional Probabilities

Language is important here. The probability that "a college-educated person is married" is different from the probability that "a person is college-educated AND is married." In the AND case, we're looking at everyone in the sample and wondering how many have both a college degree AND are married. But when we ask for the probability that a college-educated person is married, we're taking it as a given that the person is college-educated. We're *not* saying "choose someone from the whole collection." We're saying, "Just focus on the people with the college degrees. What proportion of those people are married?"

Probabilities such as these, where we focus on just one group of objects and imagine taking a random sample from that group alone, are called **conditional probabilities**.

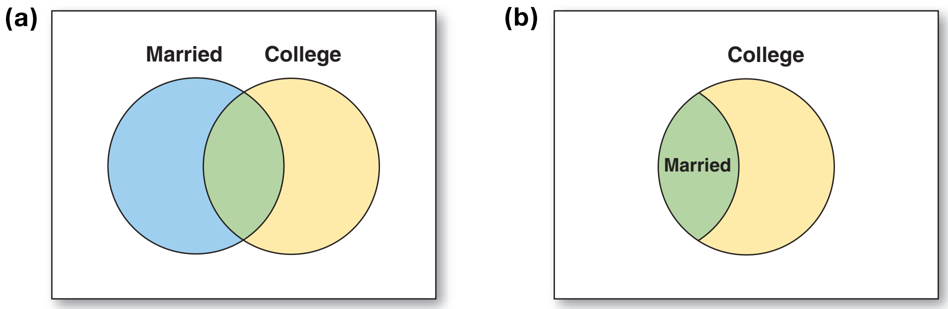
For example, in Table 5.5 (which repeats Table 5.3), we've highlighted in red the people with college degrees. In this row, there are 143 people. If we select someone at random from among those 143 people, the probability that the person will be married is 98/143 (or about 0.685). We call this a conditional probability because we're finding the probability of being married *conditioned* on having a college education (that is, we are assuming we're selecting only from college-educated people).

Education Level	Single	Married	Divorced	Widow/Widower	Total
Less HS	17	70	10	28	125
HS	68	240	59	30	397
College or higher	27	98	15	3	143
Total	112	408	84	61	665

◀ **TABLE 5.5** What's the probability that a person with a college degree or higher is married? To find this, focus on the row shown in red and imagine selecting a person from this row.

"Given That" vs. "AND" Often, conditional probabilities are worded with the phrase *given that*, as in "Find the probability that a randomly selected person is married given that the person has a college degree." But you might also see it phrased as in the last paragraph: "Find the probability that a randomly selected person with a college degree is married." The latter phrasing is more subtle, because it implies that we're supposed to assume the selected person has a college degree: We must assume we are *given that* the person has a college degree.

Figure 5.6a shows a Venn diagram representing all of the data. The green overlap region represents the event of being married AND having completed college. By way of contrast, Figure 5.6b shows only those with college educations; it emphasizes that if we wish to find the probability of being married, given that the person has a college degree, we need to focus on only those with college degrees.



◀ **FIGURE 5.6** **(a)** The probability of being married AND having a college degree; **(b)** the probability that a person with a college degree is married.

The mathematical notation for a conditional probability might seem a little unusual. We write

$$P(\text{person is married} \mid \text{person has college degree}) = 98/143$$

The vertical bar inside the probability notation is *not* a division sign. You should think of pronouncing it as "given that." This sentence reads, "The probability that the person is married, given that we know this person has a college degree, is 98/143." Some statisticians like to think of the vertical bar as meaning "within," and would translate this as "The probability that we randomly select a married person from within those who have a college degree." Either way you think about it is fine; use whichever makes the most sense to you.

KEY POINT In the study of conditional probabilities, P(A | B) means to find the probability that event A occurs, but to restrict your consideration to those outcomes of A that occur within event B. It means "the probability of A occurring, given that event B has occurred."

EXAMPLE 9 Teens and the Internet

Consider the following statements, which are based on a Pew Foundation report.

a. The probability that a randomly selected teenager (12–17 years old) from the United States will go online at some point during the month is about 93%. Event A: the selected person is a teenager. Event B: the selected person goes online during the month.

b. The probability that a randomly selected adult age 65 or older will go online during the month is 38%. Event A: the selected person is age 65 or older. Event B: the selected person will go online during the month.

c. The probability that a randomly selected resident of the United States is a teenager who will go on line this month is about 7%. Event A: the selected person is a teenager. Event B: the person will go online this month.

We wish to use this information to find the probability that a randomly selected resident of the United States will be 12–17 years old and will go online this month (Pew Foundation, July 2010).

QUESTION For each of these three statements, determine whether the events in the question are used in a conditional probability or an AND probability. Explain. Write the statement using probability notation.

SOLUTIONS

a. This statement is asking about a conditional probability. It says that among all teenagers, 93% go online. We are "given that" the group we're sampling from are all teenagers.

b. This statement is also a conditional probability.

c. This statement, on the other hand, is asking us to assume nothing and, instead, once the person is selected from the entire United States, to determine whether that person has these two characteristics.

Using probability notation, these statements are

a. P(person goes online | person is teenager)

b. P(person goes online | person is adult aged 65 or older)

c. P(person goes online AND person is teenager)

TRY THIS! Exercise 5.47

Finding Conditional Probabilities If you are given a table like Table 5.5, you can find conditional probabilities as we did above: by isolating the group from which you are sampling. However, a formula exists that is useful for times when you do not have such complete information.

The formula for calculating conditional probabilities is

$$P(A \mid B) = \frac{P(A \text{ AND } B)}{P(B)}$$

EXAMPLE 10 Education and Marital Status

Suppose a person is randomly selected from those represented in Table 5.3 on page 238.

QUESTION Find the probability that a person with less than a high school degree (and no higher degrees) is married. Use the table, then confirm your calculation with the formula.

SOLUTION We are asked to find P(married | less than high school degree)—in other words, the probability a person with less than a HS degree is married. We are told to imagine taking a random sample from only those who have less than a high school degree. There are 125 such people, of whom 70 are married.

$$P(\text{married} \mid \text{less HS}) = 70/125 = 0.560$$

The formula confirms this:

$$P(\text{married} \mid \text{less HS}) = \frac{P(\text{married AND less HS})}{P(\text{less HS})} = \frac{\frac{70}{665}}{\frac{125}{665}} = \frac{70}{125} = 0.560$$

Interestingly, the probability that a college graduate is married (0.685) is greater than the probability that someone with less than a high school education is married (0.560).

TRY THIS! Exercise 5.49

With a little algebra, we can discover that this formula can serve as another way of finding AND probabilities:

$$P(A \text{ AND } B) = P(A)P(B \mid A)$$

We'll make use of this formula later.

Summary of Probability Rules

Rule 5a: $P(A \mid B) = \frac{P(A \text{ AND } B)}{P(B)}$

Rule 5b: $P(A \text{ AND } B) = P(B)\,P(A \mid B)$ and also $P(A \text{ AND } B) = P(A)\,P(B \mid A)$

Both forms of Rule 5b are true, because it doesn't matter which event is called A and which is called B.

Flipping the Condition A common mistake with conditional probabilities is thinking that $P(A|B)$ is the same as $P(B|A)$.

$$P(B|A) \neq P(A|B)$$

A second common mistake is to confuse conditional probabilities with fractions and think that $P(B|A) = 1/P(A|B)$.

$$P(B|A) \neq \frac{1}{P(A|B)}$$

For example, using the data in Table 5.5, we earlier computed that P(married | college) = 98/143 = 0.685. Let's flip the condition and compare this to P(college | married):

$$P(\text{college} \mid \text{married}) = 98/408, \text{ or about } 0.240$$

Clearly, P(college | married) does not equal P(married | college).

Also, it is *not* true that P(married | college) = 1/P(college | married) = 408/98 = 4.16, a number bigger than 1! It is impossible for a probability to be bigger than 1, so obviously,

$$P(A|B) \text{ does not equal } \frac{1}{P(B|A)}.$$

KEY POINT $P(B|A) \neq P(A|B)$

Independent and Dependent Events

We saw that the probability that a person is married differs, depending on whether we know that person is a college grad or that she or he has less than a high school education. If we randomly select from different educational levels, we get a different probability of marriage. Another way of putting this is to say that marital status and education level are **associated**. We know they are associated because the conditional probabilities change depending on which educational level we condition on.

We call variables or events that are *not* associated **independent events**. Independent variables, and independent events, play a very important role in statistics. Let's first talk about independent events.

Two events are independent if knowledge that one event has happened tells you nothing about whether or not the other event has happened. In mathematical notation,

$$A \text{ and } B \text{ are independent events means } P(A|B) = P(A).$$

In other words, if the event "a person is married" is independent from "a person has a college degree," then the probability that a married person has a college degree, P(college | married), is the same as the probability that any person selected from the sample has a college degree. We already found that

$$P(\text{college} \mid \text{married}) = 0.240, \text{ and } P(\text{college}) = 143/665 = 0.215$$

Because these probabilities are not equal, we conclude that completing college and being married are not independent.

It doesn't matter which event you call A and which B, so events are also independent if $P(B|A) = P(B)$.

Details

Equal Probabilities
Sometimes, when real data are used to check for independence (as we're doing here), the probabilities come out very close, but not exactly equal. You'll learn in Chapter 10 how to decide when the probabilities are close enough for you to conclude that events are independent. For now, let's agree that the probabilities have to be *exactly* the same for us to conclude independence.

KEY POINT To say that events A and B are independent means that $P(A|B) = P(A)$. In words: Knowledge that event B occurred does not change the probability of event A occurring.

EXAMPLE 11 Dealing a Diamond

Figure 5.7 shows a standard deck of playing cards. When playing card games, players nearly always try to avoid showing their cards to the other players. The reason for this is that knowing the other player's cards can sometimes give you an advantage. Suppose you are wondering whether your opponent has a diamond. If you find out that one of the cards he holds is red, does this provide useful information?

QUESTION Suppose a deck of cards is shuffled and one card is dealt facedown on the table. Are the events "the card is a diamond" and "the card is red" independent?

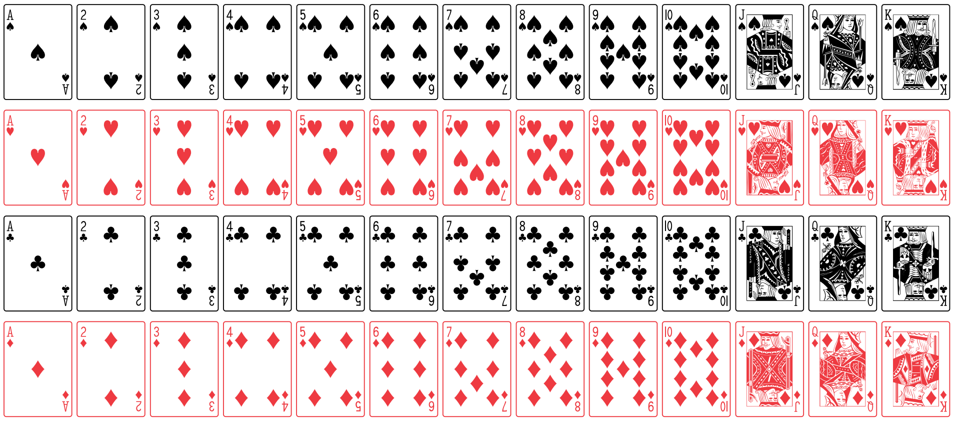


FIGURE 5.7 Fifty-two playing cards in a standard deck.

SOLUTION To answer this, we must apply the definition of independent events and find

$$P(\text{card is a diamond})$$

and compare it to

$$P(\text{card is a diamond} \mid \text{card is red})$$

If these probabilities are different, then the events are not independent; they are associated events.

Out of a total of 52 cards, 13 are diamonds. Therefore,

$$P(\text{card is a diamond}) = 13/52, \text{ or } 1/4$$

Now suppose we know the card is red. What's the probability that a red card is a diamond? That is, find

$$P(\text{card is a diamond} \mid \text{card is red})$$

Because there are 26 red cards, you are now limited to the 26 cards in the second and fourth rows in Figure 5.7. There are still 13 diamonds. Therefore, the probability that the card is a diamond, given that it is red, is $13/26 = 1/2$.

$$P(\text{card is diamond}) = 1/4$$
$$P(\text{card is diamond} \mid \text{card is red}) = 1/2$$

These probabilities are *not* equal.

CONCLUSION The events “select a diamond” and “select a red card” are associated, because P(select a diamond | color is red) is not the same as P(select a diamond). This means that if you learn, somehow, that the opponent’s card is red, then you have gained some information that will be useful in deciding whether he has a diamond.

Note that we could also have compared P(card is red) to P(card is red | card is a diamond), and we would have reached the same conclusion.

TRY THIS! Exercise 5.55

EXAMPLE 12 Dealing an Ace

A playing card is dealt facedown. This time, you are interested in knowing whether your opponent holds an ace. You have discovered that his card is a diamond. Is this helpful information?

QUESTION Are the events “card is a diamond” and “card is an ace” independent?

SOLUTION Now we must find P(card is an ace) and compare it to P(card is an ace | card is a diamond).

P(card is an ace):
Of the 52 cards, 4 are aces.
Therefore, P(card is an ace) $= 4/52 = 1/13$.

P(card is an ace | card is a diamond):
There are 13 diamonds in the deck, and only one of these 13 is an ace.
Therefore, P(card is an ace | card is a diamond) $= 1/13$.

We find that P(card is an ace) $= 1/13 =$ P(card is an ace | card is a diamond).

CONCLUSION The events “card is a diamond” and “card is an ace” are independent. This means the information that your opponent’s card is a diamond will not help you determine whether it is an ace.

Note that we could also have compared P(card is a diamond) to P(card is a diamond | card is an ace), and we would have reached the same conclusion.

TRY THIS! Exercise 5.57

Intuition about Independence

Sometimes you can use your intuition to make an educated guess about whether two events are independent. For example, flip a coin twice. You should know that P(second flip is heads) $= 1/2$. But what if you know that the first flip was a head? Then you need to find

P(second flip is heads | first flip was heads)

Intuitively, we know that the coin always has a 50% chance of coming up heads. The coin doesn’t know what happened earlier. Thus

P(second is heads | first is heads) $= 1/2 =$ P(second is heads)

The two events “second flip comes up heads” and “first flip comes up heads” are independent.

Although you can sometimes feel very confident in your intuition, you should check your intuition with data whenever possible.

EXAMPLE 13 Education and Widows

Suppose we select a person at random from the sample of people asked about education and marital status in Table 5.3. Is the event "person selected has HS education" independent from the event "person selected is widowed"? Intuitively, we would think so. After all, why should a person's educational level affect whether his or her spouse dies first?

QUESTION Check, using the data in Table 5.3 on page 238, whether these two events are independent.

SOLUTION To check independence, we need to check whether $P(A|B) = P(A)$. It doesn't matter which event we call A and which we call B, so let's check to see whether

$$\text{P(person selected has HS education} \mid \text{person is widowed)} = \text{P(person has HS education)}$$

From the table, we see that there are only 61 widows, so

$$P(\text{HS} \mid \text{widowed}) = 30/61 = 0.492$$

$$P(\text{HS}) = 397/665 = 0.597$$

The two probabilities are not equal.

CONCLUSION The events are associated. If you know the person is widowed, then the person is less likely to have a high school education than he or she would be if you knew nothing about his or her marital status. Our intuition was wrong, at least as far as these data are concerned. It is possible, of course, that these data are not representative of the population as a whole, or that our conclusion of association is incorrect because of chance variation in the people sampled.

TRY THIS! Exercise 5.59

Sequences of Independent and Associated Events

A common challenge in probability is to find probabilities for sequences of events. By *sequence,* we mean events that take place in a certain order. For example, a married couple plans to have two children. What is the probability that the first will be a boy and the second a girl? When dealing with sequences, it is helpful to determine first whether the events are independent or associated.

If the two events are associated, then our knowledge of conditional probabilities is useful, and we should use Probability Rule 5b:

$$P(\text{A AND B}) = P(\text{A})P(\text{B} \mid \text{A})$$

If the events are independent, then we know $P(B|A) = P(B)$, and this rule simplifies to $P(\text{A AND B}) = P(A)\,P(B)$. This formula is often called the **multiplication rule**.

Summary of Probability Rules

Rule 5c: Multiplication Rule. If A and B are independent events, then

$$P(\text{A AND B}) = P(\text{A})\,P(\text{B})$$

Independent Events

When two events are independent, the multiplication rule speeds up probability calculations for events joined by AND.

For example, suppose that 51% of all babies born in the United States are boys. Then P(first child is boy) = 51%. What is the probability that a family planning to have two children will have two boys? In other words, how do we find the sequence probability

P(first child is boy AND second child is boy)?

Caution

False Assumptions of Independence
The mere fact that your intuition tells you events are independent doesn't necessarily mean that they really are. If you are wrong, then using Rule 5c to compute P(A and B) will produce tragically wrong values!

Researchers have good reason to suspect that the genders of children in a family are independent (if you do not include identical twins). Because of this, we can apply the multiplication rule:

$$\begin{aligned}\text{P(first child is boy AND second child is boy)} &= \text{P(first is boy) P(second is boy)}\\ &= 0.51 \times 0.51 = 0.2601\end{aligned}$$

The same logic could be applied to finding the probability that the first child is a boy and the second is a girl:

$$\begin{aligned}\text{P(first child is boy AND second child is girl)} &= \text{P(first is boy) P(second is girl)}\\ &= 0.51 \times 0.49 = 0.2499\end{aligned}$$

EXAMPLE 14 Three Coin Flips

Toss a fair coin three times. A fair coin is one in which each side is equally likely to land up when the coin is tossed.

QUESTION What is the probability that all three tosses are tails? What is the probability that the first toss is heads AND the next two are tails?

SOLUTION Using mathematical notation, we are trying to find P(first toss is tails AND second is tails AND third is tails). We know these events are independent (this is theoretical knowledge; we "know" this because the coin cannot change itself on the basis of its past). This means the probability is

$$\text{P(first is tails)} \times \text{P(second is tails)} \times \text{P(third is tails)} = 1/2 \times 1/2 \times 1/2 = 1/8$$

Also, P(first is heads AND second is tails AND third is tails) is

$$\text{P(heads)} \times \text{P(tails)} \times \text{P(tails)} = 1/2 \times 1/2 \times 1/2 = 1/8$$

CONCLUSION The probability of getting three tails is the same as that of getting first heads and then two tails: 1/8.

TRY THIS! Exercise 5.61

EXAMPLE 15 Ten Coin Flips

Suppose I toss a coin 10 times and record whether each toss lands heads or tails. Assume that each side of the coin is equally likely to land up when the coin is tossed.

QUESTION Which sequence is the more likely outcome?

Sequence A: HTHTHTHTHT

Sequence B: HHTTTHTHHH

SOLUTION Because these are independent events, the probability that sequence A happens is

$$\begin{aligned}\text{P(H)P(T)P(H)P(T)P(H)P(T)P(H)P(T)P(H)P(T)} &= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}\\ &\times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}\\ &= \left(\frac{1}{2}\right)^{10} = 0.0009766\end{aligned}$$

The probability that sequence B happens is

$$P(H)P(H)P(T)P(T)P(T)P(H)P(T)P(H)P(H)P(H) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \left(\frac{1}{2}\right)^{10} = 0.0009766$$

CONCLUSION Even though sequence A looks improbable, because it alternates between heads and tails, both outcomes have the same probability!

TRY THIS! Exercise 5.63

Another common probability question asks about the likelihood of "at least one" of a sequence happening a certain way.

EXAMPLE 16 Teacher Satisfaction and "at Least One"

According to a 2012 MetLife survey, 44% of U.S. public school teachers reported that they were satisfied with their jobs. (This is the lowest percentage in more than twenty years.) Suppose we select three teachers randomly and with replacement from the population of all teachers in the United States. (Selecting "with replacement" means that, once a teacher is selected, they are eligible to be selected again.)

QUESTIONS

a. What is the probability that all three are satisfied with their careers?

b. What is the probability that none of the three is satisfied?

c. What is the probability that at least one teacher is satisfied?

SOLUTIONS

a. We are asked to find P(first is satisfied AND second is satisfied AND third is satisfied). Because the teachers were selected at random from the population, these are independent events. (One teacher's answer won't affect the probability that the next one will answer one way or the other.) Because these are independent events, the probability we seek is just P(first is satisfied) $\times$ P(second is satisfied) $\times$ P(third is satisfied) $= 0.44 \times 0.44 \times 0.44 = 0.0852$.

b. The probability that none is satisfied is trickier to determine. This event occurs if the first teacher is not satisfied AND the second is not AND the third is not. So we need to find P(first not satisfied AND second not satisfied AND third not satisfied) $=$ P(first not) $\times$ P(second not) $\times$ P(third not) $= (1 - 0.44)(1 - 0.44)(1 - 0.44) = 0.1756$.

c. The probability that at least one teacher is satisfied is the probability that one is satisfied OR two are satisfied OR all three are satisfied. The calculation is easier if you realize that "at least one is satisfied" is the complement of "none is satisfied" because it includes all categories except "none." And so

$$1 - 0.1756 = 0.8244$$

CONCLUSION The probability that all three randomly selected teachers are satisfied with their careers is 0.0852. The probability that none is satisfied is 0.1756. The probability that at least one is satisfied is 0.8244.

TRY THIS! Exercise 5.65

Caution

Non-independent Events
If events are not independent, then the multiplication rule does not apply. Multiplying probabilities for non-independent events can lead to disaster. (Read the Case Study!)

SECTION 5.4

Finding Empirical Probabilities

Empirical probabilities are based on observations of real-life events. For instance, a softball player's batting average—the number of successful hits divided by the number of attempts—can be thought of as the empirical probability that she will get a hit next time at the plate. The percentage of times you've forgotten an item on your shopping list is the empirical probability that you will forget an item on your next shopping trip. If we've collected data—that is, if for each shopping trip we've recorded whether or not we forgot something—then we can find the empirical probability.

But sometimes we can't find the data or the situation is too complex for us to find the empirical probability of some random event. If so, we can simulate the situation. The key component is to be able to simulate the randomness with something that has the same probability of happening as that real-life event we are interested in.

Coin Flip Simulation

Let's consider this very simple real-life event: We wish to find the probability that we will get three heads if we flip a coin three times. This is a probability we can figure out theoretically if we assume that each side of the coin is equally likely (the solution is 1/8), so we can compare our empirical probability to the theoretical. As you will see from the Law of Large Numbers at the end of this section, when you do a large enough number of repetitions, the empirical probability gets close to the theoretical result.

At one level, simulations are straightforward. An empirical probability is simply the proportion of times an event with a random outcome happens out of so many tries. All we have to do is perform some action many times and count how often the event we're interested in occurs.

To create a simulation to find an empirical probability, follow these steps:

1. Identify the random action and the probability of a successful outcome.
2. Determine how to simulate this random action.
3. Determine the event that you're interested in. (Often, this will be some summary of the outcomes of several actions).
4. Explain how you will simulate one trial.
5. Carry out a trial, and record whether or not the event you are interested in occurred.
6. Repeat a trial many times, at least 100, and count the number of times your event occurred.
7. The number of times the event occurred, divided by the number of trials, is your empirical probability.

To apply this to our coin flip example, we identify the random action to be a single flip of the coin. The probability of success (getting heads), is 0.50. We *could* flip a coin to simulate this action, but this is slow if we wish to do many flips. Instead, we will use the random number table (Appendix A). Each flip of a coin has two possible outcomes—heads or tails—and each outcome is equally likely. Therefore, we need to assign digits to represent "heads" and "tails" in such a way that these represented outcomes are equally likely.

There are many ways to do this, but here's one: Assign even numbers (0, 2, 4, 6, 8) to represent "tails" and odd numbers (1, 3, 5, 7, 9) to represent "heads," The probability of getting an odd number is 0.50.

Line						
01	2 1 0 3 3	3 2 5 2 2	1 9 3 0 5	9 0 6 3 3	8 0 8 7 3	1 9 1 6 7
02	1 7 5 1 6	6 9 3 2 8	8 8 3 8 9	1 9 7 7 0	3 3 1 9 7	2 7 3 3 6
03	2 6 4 2 7	4 0 6 5 0	7 0 2 5 1	8 4 4 1 3	3 0 8 9 6	2 1 4 9 0
04	4 5 5 0 6	4 4 7 1 6	0 2 4 9 8	1 5 3 2 7	7 9 1 4 9	2 8 4 0 9
05	5 5 1 8 5	7 4 8 3 4	8 1 1 7 2	8 9 2 8 1	4 8 1 3 4	7 1 1 8 5

◀ **TABLE 5.6** The first five lines of the random number table, Appendix A.

We are now ready for step 3, where we state that the event we are interested in is whether or not we get three heads after three flips. To simulate this (step 4), we read off three digits in a row from the random number table; each digit represents a flip of the coin. We start at the beginning of line 01. If we see an even number we record T, and if we see an odd number, H.

Table 5.7 shows steps 5 and 6 for ten trials. We record "Yes" whenever we see the event of interest, HHH, occur.

Trial	Random Numbers	Translation	Number of Heads	Did Event Occur?
1	2 1 0	THT	1	No
2	3 3 3	HHH	3	Yes
3	2 5 2	THT	1	No
4	2 1 9	THH	2	No
5	3 0 5	HTH	2	No
6	9 0 6	HTT	1	No
7	3 3 8	HHT	2	No
8	0 8 7	TTH	1	No
9	3 1 9	HHH	3	Yes
10	1 6 7	HTH	2	No

◀ **TABLE 5.7** Simulating heads and tails with random numbers. Trials resulting in three heads are shown in red.

The event we are interested in happened twice in the ten trials (shown in color in Table 5.7), so our empirical probability is 2/10, or 0.20 (step 7). This is quite a bit different from the theoretical probability of 1/8 (which is 0.125). As you will soon see, this is not surprising, because with only ten trials, there's quite a bit of variability in the empirical probabilities we might get. That is why it is best to do a large number of trials.

EXAMPLE 17 Dice Simulation

Use a simulation to find the approximate probability that a fair, six-sided die will land with the 6 showing on top. Do ten trials, using these random numbers:

4 4 6 8 7	7 5 0 3 2	8 3 4 0 8	1 0 2 3 9
8 0 0 1 6	5 8 2 5 0	9 1 4 1 9	5 6 3 1 5

Trial	Translation	Did Event Occur?
1	4	No
2	4	No
3	6	Yes
4	5	No
5	3	No
6	2	No
7	3	No
8	4	No
9	1	No
10	2	No

▲ **TABLE 5.8** Simulations for rolling a six-sided die.

QUESTION What is your empirical probability of getting a 6? Compare this to the theoretical probability. Show all steps.

SOLUTION

Step 1: The random action is throwing a die, and the probability of a successful outcome is $1/6$, because the probability that a fair, six-sided die lands with the 6 on top is $1/6$ (assuming the die is well balanced so that each outcome is equally likely.)

Step 2: We simulate this random action using digits in the random number table. Each number in the table will represent the roll of a die. The number in the table will represent the number that comes up on the die. We'll ignore the digits 0, 7, 8, and 9, because these are impossible outcomes when rolling a six-sided die.

Step 3: The event we're interested in is whether we see a 6 after a single toss.

Step 4: A trial consists of reading a single digit from the table.

Step 5: The first line of Table 5.8 shows the results of one trial. Our simulated die landed on a 4. We record "No" because the event we are studying did not happen. Note that we simply skip the digits 0, 7, 8, and 9.

Step 6: The remaining nine trials are shown in the table.

Step 7: A 6 occurred on one of the ten trials (the third trial).

CONCLUSION Our empirical probability of getting a 6 on the roll of a balanced die is 1/10, or 0.10. In contrast, the theoretical probability is $1/6$, or about 0.167.

TRY THIS! Exercise 5.67

The Law of Large Numbers

The **Law of Large Numbers** is a famous mathematical theorem that tells us that if our simulation is designed correctly, then the more trials we do, the closer we can expect our empirical probability to come to the true probability. The Law of Large Numbers shows that as we approach infinitely many trials, the true probability and the empirical probability approach the same value.

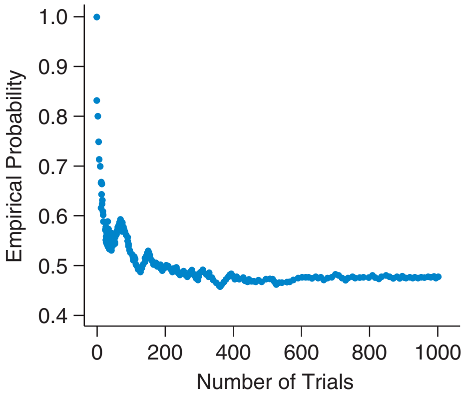
KEY POINT The Law of Large Numbers states that if an experiment with a random outcome is repeated a large number of times, the empirical probability of an event is likely to be close to the true probability. The larger the number of repetitions, the closer together these probabilities are likely to be.

The Law of Large Numbers is the reason why simulations are useful: Given enough trials, and assuming that our simulations are a good match to real life, we can get a good approximation of the true probability.

Table 5.9 shows the results of a very simple simulation. We used a computer to simulate flipping a coin, and we were interested in observing the frequency at which the "coin" comes up heads. We show the results at the end of each trial so that you can see how our approximation gets better as we perform more trials. For example, on the first trial we got a head, so our empirical probability is $1/1 = 1.00$. On the second and third trials we got heads, so the empirical probability of heads is still 1.00. On the fourth trial we got tails, and up to that point we've had 3 heads in 4 trials. Therefore, our empirical probability is $3/4 = 0.75$.

Trial	Outcome	Empirical Probability of Heads
1	H	$1/1 = 1.00$
2	H	$2/2 = 1.00$
3	H	$3/3 = 1.00$
4	T	$3/4 = 0.75$
5	H	$4/5 = 0.80$
6	H	$5/6 = 0.83$
7	T	$5/7 = 0.71$
8	H	$6/8 = 0.75$
9	T	$6/9 = 0.67$
10	H	$7/10 = 0.70$

▲ **TABLE 5.9** Simulations of heads and tails with cumulative empirical probabilities



◀ **FIGURE 5.10** The Law of Large Numbers predicts that after many flips, the proportion of heads we get from flipping a real coin will get close to the true probability of getting heads. Because these empirical probabilities are "settling down" to about 0.50, this supports the theoretical probability of 0.50.

We can continue this way, but it is easier to show you the results by making a graph. Figure 5.10 shows a plot of the empirical probabilities against the number of trials.

Note that with a small number of trials (say, less than 75 or so), our empirical probability was relatively far away from the theoretical value of 0.50. Also, when the number of trials is small, the empirical probabilities can change a lot with each additional "coin flip." But eventually things settle down, and the empirical probabilities get close to what we know to be the theoretical probability. If you were to simulate a coin flip just 20 times, you might not expect your empirical probability to be too close to the theoretical probability. But after 1000 flips, you'd expect it to be very close.

How Many Trials Should I Do in a Simulation? In our two examples of simulations we did 10 trials, which is not very many. As you can tell from Figure 5.10, if we had stopped at 5 flips, we would have had an empirical probability that was pretty far from the true value.

The number of trials you need to be confident that your empirical probability is close to the theoretical value can be fairly large. In general, the more rare the event, the more trials you'll need. Fortunately, simulations are very quick when done with computers, and a variety of software tools are available to help you do fast simulations. We recommend that you do at least 100 trials in most cases.

What If My Simulation Doesn't Give the Theoretical Value I Expect? You have no guarantee that your empirical probability will give you exactly the theoretical value. As you noticed if you did the above activity, even though it's not unusual to get 50% heads in 10 tosses, more often you get some other proportion.

When you don't get the "right" value, two explanations are possible:

1. Your theoretical value is incorrect.
2. Your empirical probability is just varying—that's what empirical probabilities do. You can make it vary less and get closer to the theoretical value by doing more trials.

How do we choose between these two alternatives? Well, that's what statistics is all about! This is one of the central questions of statistics: Are the data consistent with our expectations, or do they suggest that our expectations are wrong? We will return to this question in almost every chapter in this text.

KEY POINT When based on a small number of trials, empirical probabilities can stray quite far from their true values.

Some Subtleties with the Law

The Law of Large Numbers (LLN) is one law that cannot be broken. Nevertheless, many people think the law has been broken when it really hasn't, because interpreting the LLN takes some care.

The Law of Large Numbers tells us, for example, that with many flips of a coin, our empirical probability of heads will be close to 0.50. It tells us nothing about the *number* of heads we will get after some number of tosses and nothing about the order in which the heads and tails appear.

Streaks: Tails Are Never "Due" Many people mistakenly believe the LLN means that if they get a large number of heads in a row, then the next flip is more likely to come up tails. For example, if you just flipped five heads in a row, you might think that the sixth is more likely to be a tail than a head so that the empirical probability will work out to be 0.50. Some people might incorrectly say, "Tails are due."

This misinterpretation of the LLN has put many a gambler into debt. This is a misinterpretation for two reasons. First, the Law of Large Numbers is patient. It says that the empirical probability will equal the true probability after infinitely many trials. That's a lot of trials. Thus a streak of 10 or 20 or even 100 heads, though extremely rare, does not contradict the LLN.

CASE STUDY REVISITED

Sally Clark was convicted of murdering two children on the basis of the testimony of physician/expert Dr. Meadow. Dr. Meadow testified that the probability of two children in the same family dying of SIDS was extremely low and that, therefore, murder was a more plausible explanation. In Dr. Meadow's testimony, he assumed that the event "One baby in a family dies of SIDS" AND "a second baby in the family dies of SIDS" were independent. For this reason, he applied the multiplication rule to find the probability of two babies dying of SIDS in the same family. The probability of one baby dying was $1/8543$, so the probability of two independent deaths was $(1/8543) \times (1/8543)$, or about 1 in 73 million.

However, as noted in a press release by the Royal Statistical Society, these events may not be independent. "This approach (multiplying probabilities) . . . would only be valid if SIDS cases arose independently within families, an assumption that would need to be justified empirically. Not only was no such empirical justification provided in the case, but there are very strong *a priori* reasons for supposing that the assumption will be false. There may well be unknown genetic or environmental factors that predispose families to SIDS, so that a second case within the family becomes much more likely."

Dr. Meadow made several other errors in statistical reasoning, which are beyond the scope of this chapter but quite interesting nonetheless. There's a nice discussion that's easy to read at http://www.richardwebster.net/cotdeaths.html. For a summary and a list of supporting references, including a video by a statistician explaining the statistical errors, see http://en.wikipedia.org/wiki/Sally_Clark.

Sally Clark was released from prison but died in March 2007 of alcohol poisoning at the age of 43. Her family believes her early death was caused in part by the stress inflicted by her trial and imprisonment.

EXPLORING STATISTICS
CLASS ACTIVITY

Let's Make a Deal: Stay or Switch?

GOALS	MATERIALS
Simulate an experiment to estimate complex (and counterintuitive) probabilities.	Three playing cards—two of one color, the third of the other color—for each group of three students.

ACTIVITY

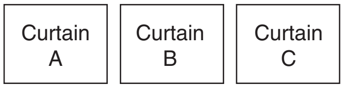
Legend has it that in the popular game show "Let's Make a Deal," a contestant was shown three curtains. Behind one curtain was a high-value prize—for example, a car—and behind the other two were less desirable prizes—for example, goats. The contestant picked one of the three curtains. Next the host (who knew what was behind each curtain) raised another curtain to reveal a goat. The host, before revealing what was behind the contestant's chosen curtain, then asked the contestant whether she wanted to stay with the curtain that she had chosen or switch to the other curtain.

What should the contestant's strategy be? There are three possibilities: The contestant should stay; the contestant should switch; it doesn't matter whether the contestant stays or switches.

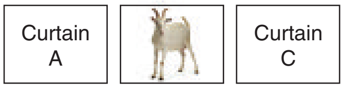
Form groups of three and choose roles: host, contestant, and recorder. The host lays the cards facedown, with the two cards of the same color representing goats and the other card representing the car. The contestant chooses a card. The host then turns over a "goat" card (but not the one just selected). The contestant now must choose a strategy: stay or switch. After that decision is made, the host reveals whether the contestant won a goat or a car. Repeat this several times, giving the contestant several opportunities to try both strategies. For each trial, the recorder will put a tally mark in one of four categories:

Stayed and Won; Stayed and Lost; Switched and Won; Switched and Lost

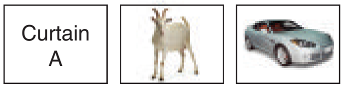
After a few games, change roles so that everyone gets to try all roles. Your instructor will compile the results.



▲ **FIGURE A** Three curtains, and the contestant picked curtain A.



▲ **FIGURE B** The host, who knows there is a goat behind curtain B, reveals curtain B. Should the contestant stick with curtain A or switch to curtain C?



▲ **FIGURE C** The contestant switched and won.

BEFORE THE ACTIVITY

Which strategy do you think will be better: staying or switching? Or does it make no difference whatsoever? Why? (Bonus points: What is the probability of winning if you always switch? It's probably not what you think!)

AFTER THE ACTIVITY

Judging on the basis of your class data, what is the empirical probability of winning if the contestant stays? What is it if the contestant switches? Do these probabilities convince you to change your strategy? Explain.

CHAPTER REVIEW

KEY TERMS

random, *231*
probability, *232*
theoretical probabilities, *232*
empirical probabilities, *232*
simulation, *233*
complement, *234*
sample space, *234*
event, *234*
Venn diagram, *237*
AND, *237*
inclusive or, *239*
mutually exclusive events, *240*
conditional probabilities, *243*
associated, *246*
independent events, *246*
multiplication rule, *249*
Law of Large Numbers, *254*

LEARNING OBJECTIVES

After reading this chapter and doing the assigned homework problems, you should

- Understand that humans can't reliably create random numbers or sequences.
- Understand that a probability is a long-term relative frequency.
- Know the difference between empirical and theoretical probabilities—and know how to calculate them.
- Be able to determine whether two events are independent or associated, and understand the implications of making incorrect assumptions about independent events.
- Understand that the Law of Large Numbers enables us to use empirical probabilities to estimate and test theoretical probabilities.
- Know how to design a simulation to estimate empirical probabilities.

SUMMARY

Random samples or random experiments must be generated with the use of outside mechanisms such as computer algorithms or by relying on random number tables. Human intuition cannot be relied on to produce reliable "randomness."

Probability is based on the concept of long-run relative frequencies: If an action is repeated infinitely many times, how often does a particular event occur? To find theoretical probabilities, we calculate these relative frequencies on the basis of assumptions about the situation and rely on mathematical rules. In finding empirical probabilities, we actually carry out the action many times or, alternatively, rely on a simulation (using a computer or random number table) to quickly carry out the action many times. The empirical probability is the proportion of times a particular event was observed to occur. The Law of Large Numbers tells us that the empirical probability becomes closer to the true probability as the number of repetitions is increased.

Theoretical Probability Rules

Rule 1: A probability is always a number from 0 to 1 (or 0% to 100%) inclusive (which means 0 and 1 are allowed). It may be expressed as a fraction, a decimal, or a percent.

$$0 \leq P(A) \leq 1$$

Rule 2: For any event A,

$$P(\text{A does } \textit{not} \text{ occur}) = 1 - P(\text{A does occur})$$

A^c is the complement of A:

$$P(A^c) = 1 - P(A)$$

Rule 3: For equally likely outcomes:

$$P(A) = \frac{\text{Number of outcomes in A}}{\text{Number of all possible outcomes}}$$

Rule 4: Always: $P(\text{A OR B}) = P(A) + P(B) - P(\text{A AND B})$
Rule 4a: Only if A and B are mutually exclusive:

$$P(\text{A OR B}) = P(A) + P(B)$$

Rule 5a: Conditional probabilities

$$\text{Probability of A given that B occurred: } P(A \mid B) = \frac{P(\text{A AND B})}{P(B)}$$

Rule 5b: Always: $P(\text{A AND B}) = P(B)\,P(A \mid B)$
Rule 5c: Multiplication Rule. If A and B are independent events, then

$$P(\text{A AND B}) = P(A)\,P(B)$$

This applies for any (finite) number of events. For example, $P(\text{A AND B AND C AND D}) = P(A)\,P(B)\,P(C)\,P(D)$ if A, B, C, D are independent of each other.

SOURCES

MetLife Survey of the American teacher. 2013. http://www.harrisinteractive.com

Starr, N. 1997. Nonrandom risk: The 1970 draft lottery. *Journal of Statistics Education* 5(2). Vietnam-era draft data can be found, along with supporting references, at www.amstat.org/publications/jse/datasets/draft.txt.

Pew Foundation. *Social Media and Young Adults* report. http://pewinternet.org (accessed July 2010).

SECTION EXERCISES

SECTION 5.1

TRY **5.1 Simulation (Example 1)** If we flip a coin 10 times, how often do we get 6 or more heads? A first step to answering this question would be to simulate 10 flips. Use the random number table in Appendix A to simulate flipping a coin 10 times. Let odd digits (1, 3, 5, 7, 9) represent heads, and let even digits (0, 2, 4, 6, 8) represent tails. Begin with the first digit in the third row.

a. Write the sequence of 10 random digits.
b. Write the sequence of 10 "heads" and "tails." Write H for heads and T for tails.
c. How many heads did you get? Did you get 6 or more heads?

5.2 Simulation Suppose you are carrying out a randomized experiment to test whether loud music interferes with memorizing numbers. You have 20 college student participants. You want each participant to have a 50% chance of being assigned to the experimental group (memorizes numbers while music plays) and a 50% chance of being assigned to the control group (memorizes numbers with no music). Let the digits 0, 1, 2, 3, and 4 represent assignment to the experimental group (music) and the digits 5, 6, 7, 8, and 9 represent assignment to the control group. Begin with the first digits in the fifth line of the random number table in the back of the text.

a. Write the 20 random numbers. For each number, write M under it if it represents a student randomized to the music group and C if it represents a student randomized to the control group.
b. What percentage of the 20 participants were assigned to the music group?
c. Would it be appropriate to assign all the even numbers (0, 2, 4, 6, 8) to the music group and all the odd numbers (1, 3, 5, 7, 9) to the control group? Why or why not?

(In real life, researchers might sample from the random number table without replacement, to make sure that both groups have the same sample size.)

5.3 Empirical vs. Theoretical A Monopoly player claims that the probability of getting a 4 when rolling a six-sided die is 1/6 because the die is equally likely to land on any of the six sides. Is this an example of an empirical probability or a theoretical probability? Explain.

5.4 Empirical vs. Theoretical A person was trying to understand the probability of drawing a black card from a fair deck of cards. He drew a card 20 times, and in these 20 times, a black card was drawn 12 times. On the basis of this, he claims that the probability of drawing a black card from a fair deck of cards is 60%. Is this an example of empirical probability or theoretical probability? Explain.

5.5 Empirical vs. Theoretical A friend flips a coin 10 times and says that the probability of getting a head is 60% because he got six heads. Is the friend referring to an empirical probability or a theoretical probability? Explain.

5.6 Empirical vs. Theoretical A magician claims that he has a fair coin—"fair" because both sides, heads and tails, are equally likely to land face up when the coin is flipped. He tells you that if you flip the coin three times, the probability of getting three tails is 1/8. Is this an empirical probability or a theoretical probability? Explain.

SECTION 5.2

5.7 Magistrate's Court District Judges Cases of a particular district are assigned to district judges (Magistrate's Court) randomly. The list of the district judges for Western District (UK) (taken from the Courts and Tribunals Judiciary website, www.judiciary.gov.uk), is given in the table. Assume that only Callaway, Goozee, and Lorrain Morgan are females and the rest are males. If you were a defense attorney in the Western District, you might be interested in whether the judge assigned to your case was a male or a female.

Arnold
Bopa Rai
Brown
Callaway
Simon Cooper
Gillibrand
Goozee
Lorrain Morgan
David Huw Parsons

Suppose the names are in a pot, and a clerk pulls a name out at random.

a. List the equally likely outcomes that could occur; last names are enough.
b. Suppose the event of interest, event A, is that a judge is a man. List the outcomes that make up event A.
c. What is the probability that one case will be assigned to a male judge.
d. List the outcomes that are in the complement of event A.

5.8 Random Assignment of Professors A study randomly assigned students attending the Air Force Academy to different professors for Calculus I, with equal numbers of students assigned to each professor. Some professors were experienced, and some were relatively inexperienced. Suppose the names of the professors are Peters, Parker, Diaz, Nguyen, and Black. Suppose Diaz and Black are inexperienced and the others are experienced. The researchers reported that the students who had the experienced teachers for Calculus I did better in Calculus II. (Source: Scott E. Carrell and James E. West, *Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors*, 2010)

a. List the equally likely outcomes that could occur for assignment of one student to a professor.
b. Suppose the event of interest, event A, is that a teacher is experienced. List the outcomes that make up event A.
c. What is the probability that a student will be assigned to an experienced teacher?
d. List the outcomes in the complement of event A. Describe this complement in words.
e. What is the probability that a student will be assigned to an inexperienced teacher?

5.9 Which of the following numbers could *not* be probabilities, and why?

a. -0.85
b. 8.50
c. 8.5%
d. 0.85
e. 850%

5.10 Which of the following numbers could *not* be probabilities, and why?

a. 38.4 b. 3.84% c. −3.84 d. 384% e. 0.00384

TRY **5.11 Playing Cards (Example 2)** There are four suits: clubs (♣), diamonds (♦), hearts (♥), and spades (♠), and the following cards appear in each suit: ace, 2, 3, 4, 5, 6, 7, 8, 9, 10, jack, queen, king. The jack, queen, and king are called face cards because they have a drawing of a face on them. Diamonds and hearts are red, and clubs and spades are black.

If you draw 1 card randomly from a standard 52-card playing deck, what is the probability that it will be:

a. A heart?
b. A red card?
c. An ace?
d. A face card (jack, queen, or king)?
e. A three?

5.12 Playing Cards Refer to Exercise 5.11 for information about cards. If you draw 1 card randomly from a standard 52-card playing deck, what is the probability that it will be:

a. A black card?
b. A diamond?
c. A face card (jack, queen, or king)?
d. A nine?
e. A king or queen?

5.13 Guessing on Tests

a. On a true/false quiz in which you are guessing, what is the probability of guessing correctly on one question?
b. What is the probability that a guess on one true/false question will be incorrect?

5.14 Guessing on Balls Consider a bag containing five balls of different colors (green, blue, red, white, and yellow) for each of these questions.

a. What is the probability of guessing the draw of a blue ball if a ball is to be drawn only once?
b. What is the probability of guessing the draw of any other ball in one draw?

TRY **5.15 Four Children (Example 3)** The sample space given here shows all possible sequences for a family with 4 children, where B stands for boy and G stands for girl.

GGGG	GGGB	GGBB	GBBB	BBBB
	GGBG	GBGB	BGBB	
	GBGG	GBBG	BBGB	
	BGGG	BGGB	BBBG	
		BGBG		
		BBGG		

Assume that all of the 16 outcomes are equally likely. Find the probability of having the following numbers of girls out of 4 children: (a) exactly 0 girls, (b) exactly 1 girl, (c) exactly 2 girls, (d) exactly 3 girls, (e) exactly 4 girls.

(*Hint:* The probability of having 3 girls and a boy is 4/16, or 25%, because the second column shows that there are 4 ways to have 3 girls and 1 boy.)

5.16 Three Coins The sample shows the possible sequences for flipping three fair coins or flipping one coin three times, where H stands for heads and T stands for tails.

HHH	HHT	HTT	TTT
	HTH	THT	
	THH	TTH	

Assume that all of the 8 outcomes are equally likely. Find the probability of having exactly the following numbers of heads out of the 3 coins: (a) exactly 0 heads, (b) exactly 1 head, (c) exactly 2 heads, (d) exactly 3 heads. (e) What do the four probabilities add up to and why?

5.17 Anniversary What is the probability that the anniversary of a randomly selected couple will fall on a weekday if all the days of the week are equally likely?

5.18 Playing Cards If *one* card is selected from a well-shuffled deck of 52 cards, what is the probability that the card will be a club OR a diamond OR a heart? What is the probability of the complement of this event? (Refer to Exercise 5.11 for information about cards.)

5.19 College Poll A StatCrunch poll asked people if college was worth the financial investment. They also asked the respondent's gender. The table shows a summary of the responses. (Source: StatCrunch: *Responses to Is college worth it?* Owner: scsurvey)

	Female	Male	All
No	45	56	101
Unsure	100	96	196
Yes	577	401	978
All	722	553	1275

a. If a person is chosen randomly from the group, what is the probability that the person is male?
b. If a person is chosen randomly from the group, what is the probability that the person said Yes?

5.20 College Poll Refer to the table given for Exercise 5.19.

a. If a person is chosen randomly, what is the probability that the person is female?
b. If a person is chosen randomly, what is the probability that the person said No?

TRY **5.21 College Poll: "AND" (Example 4)** Refer to the table given for Exercise 5.19. If a person is chosen randomly from the group, what is the probability that the person is female AND said Yes?

5.22 College Poll: "AND" Refer to the table given for Exercise 5.19. If a person is chosen randomly from the group, what is the probability that the person is male AND said No?

TRY **5.23 College Poll: "OR" (Example 5)** Refer to the table given for Exercise 5.19.

a. If a person is chosen randomly from the group, what is the probability that the person said Yes OR No?
b. Are saying Yes and saying No complementary in this data set? Explain.

5.24 College Poll: OR Refer to the table given for Exercise 5.19.

a. If a person is chosen randomly from the group, what is the probability that the person is male OR female?
b. Are the events being male and being female complementary? Explain.

TRY g **5.25 College Poll: "OR" (Example 6)** Refer to the table given for Exercise 5.19. If a person is chosen randomly from the group, what is the probability that the person is male OR said Yes (or both)? The question was whether college was worth the financial investment. *See page 269 for guidance.*

5.26 College Poll: OR Refer to the table given for Exercise 5.19. If a person is chosen randomly from the group, what is the probability that the person is female OR said No (or both)?

TRY **5.27 College Poll: Mutually Exclusive (Example 7)** Referring to the table given in Exercise 5.19, name a pair of mutually exclusive events that could result when one person was selected at random from the entire group.

5.28 College Poll: Not Mutually Exclusive Refer to the table given in Exercise 5.19. Suppose we select one person at random from this group. Name a pair of events that are *not* mutually exclusive.

5.29 Mutually Exclusive Suppose a student is selected at random in a college. Label each pair of events as mutually exclusive or not mutually exclusive.

a. The student studies economics; the student studies statistics.

b. The student is pursuing a graduate degree; the student is pursuing a post-graduate degree.

5.30 Mutually Exclusive Suppose a student is selected at random in a college. Label each pair of events as mutually exclusive or not mutually exclusive.

a. The student is specializing in finance; the student is specializing in computers.

b. The student gets a distinction; the student plays basketball.

5.31 "OR" for Farmers In a village, the percentage of farmers who use chemical fertilizers is 80%. About 60% of the farmers practice crop rotation. From this information, is it possible to find the percentage of farmers who use chemical fertilizers OR crop rotation (or both)? Why or why not?

5.32 "OR" for Dinner Suppose a family says the probability that they will dine out on Friday is 60% and the probability that they will dine out on Saturday is 80%. From this information, is it possible to find the probability that the family will dine out on Friday OR Saturday (or both)? Why or why not?

TRY **5.33 Fair Die (Example 8)** Roll a fair six-sided die.

a. What is the probability that the die shows an odd number OR a number less than 3 on top?

b. What is the probability that the die shows an odd number OR a number less than 2 on top?

5.34 Roll a Die Roll a fair six-sided die.

a. What is the probability that the die shows an odd number OR a number greater than 5 on top?

b. What is the probability that the die shows an odd number OR a number greater than 4 on top?

5.35 Grades Assume that the only grades possible in a history course are A, B, C, and lower than C. The probability that a randomly selected student will get an A in a certain history course is 0.18, the probability that a student will get a B in the course is 0.25, and the probability that a student will get a C in the course is 0.37.

a. What is the probability that a student will get an A OR a B?

b. What is the probability that a student will get an A OR a B OR a C?

c. What is the probability that a student will get a grade lower than a C?

5.36 Changing Multiple-Choice Answers One of the authors did a survey to determine the effect of students changing answers while taking a multiple-choice test on which there is only one correct answer for each question. Some students erase their initial choice and replace it with another. It turned out that 61% of the changes were from incorrect answers to correct and that 26% were from correct to incorrect. What percent of changes were from incorrect to incorrect?

5.37 Driving License Suppose that according to transport authority, adults are classified as having a regular driving license, having a temporary/learner's license, and not having a driving license. In a city, 78% of the adults had a regular driving license and 14% did not have a driving license. What percentage of adults held a learner's driving license?

5.38 Education The children of a locality go to primary school, middle school, and high school. Suppose that 38% go to primary school and 42% go to middle school. What percentage of children go to high school?

5.39 "AND" and "OR" Consider these categories of players, assuming that we are talking about all the players in a national cricket team:

Category 1: Players who can bat AND bowl well.

Category 2: Players who can only bat well.

Category 3: Players who can only bowl well.

Category 4: Players who can bat OR bowl well.

a. Which of the four categories has the most players?

b. Which category has the fewest players?

5.40 "AND" and "OR" Assume that we are talking about all residents in a colony.

a. Which group is larger: residents having children below 10 years OR children above 10 years, or residents having children below 10 years?

b. Which group is larger: residents having children below 10 years AND children above 10 years, or residents having children below 10 years?

5.41 "AND" and "OR" Considering all the students in a college, which group is larger: students who can sing AND dance or students who can sing OR dance?

5.42 "AND" and "OR" Considering all the employees of a company, which group is larger: employees coming late OR leaving early or employees coming late AND leaving early?

* **5.43 Thumbtacks** When a certain type of thumbtack is tossed, the probability that it lands tip up is 60%. All possible outcomes when two thumbtacks are tossed are listed. U means the tip is up, and D means the tip is down.

UU UD DU DD

a. What is the probability of getting two Ups?

b. What is the probability of getting exactly one Up?

c. What is the probability of getting at least one Up (one or more Ups)?

d. What is the probability of getting at most one Up (one or fewer Ups)?

* **5.44 Thumbtacks** When a certain type of thumbtack is tossed, the probability that it lands tip up is 60%, and the probability that it lands tip down is 40%. All possible outcomes when two thumbtacks are tossed are listed. U means the tip is Up, and D means the tip is Down.

UU UD DU DD

a. What is the probability of getting exactly one Down?

b. What is the probability of getting two Downs?

c. What is the probability of getting at least one Down (one or more Downs)?

d. What is the probability of getting at most one Down (one or fewer Downs)?

* **5.45 Online Test** An online test consists of 20 multiple-choice questions. Each of the 20 answers is either right or wrong. Suppose the probability that an examinee gets fewer than 5 answers correct is 0.42 and the probability that an examinee gets from 5 to 12 (inclusive) answers correct is 0.38. Find the probability that an examinee gets:

a. More than 12 answers correct
b. At most 12 answers correct
c. 50 or more answers correct
d. Which of the two events in a-c are complements of each other, and why?

* **5.46 True or False Exam** A true/false exam consists of 50 questions. Each of the questions can be answered as either true or false. Suppose that the probability of getting fewer than 20 answers correct is 0.24 and the probability of getting 20 to 40 answers correct is 0.64. Find the probability of getting:

a. At most 40 answers correct
b. 41 or more answers correct
c. 20 or more answers correct
d. Which of the two events in a-c are complements of each other? Explain.

SECTION 5.3

TRY **5.47 College Poll Again: Is College Worth It? (Example 9)**

	Female	Male	All
No	45	56	101
Unsure	100	96	196
Yes	577	401	978
All	722	553	1275

A person is selected randomly from the men in the group whose responses are summarized in the table. We want to find the probability that a male said Yes. Which of the following statements best describes the problem?

i. P(Yes|Male)
ii. P(Male|Yes)
iii. P(Male AND Yes)

5.48 College Poll A person is selected randomly from the entire group whose responses are summarized in the table for Exercise 5.47. We want to find the probability that the person selected is a male who said yes. Which of the following statements best describes the problem?

i. P(Yes|Male)
ii. P(Male|Yes)
iii. P(Male AND Yes)

TRY **5.49 College Poll (Example 10)** Use the data given in Exercise 5.47.

a. Find the probability that a randomly chosen person said Yes given that the person is female. In other words, what percentage of the females said Yes?
b. Find the probability that a randomly chosen person said Yes given that the person is male. In other words, what percentage of the males said Yes?
c. Were the males or were the females more likely to say Yes?

5.50 College Poll Use the data given in Exercise 5.47.

a. Find the probability that a randomly chosen person was female given that the person said Yes. In other words, what percentage of the people who said Yes were female?
b. Find the probability that a randomly chosen person who reported being Unsure was female. In other words, what percentage of the people who were Unsure were female?
c. Find the probability that a randomly selected person from the entire group was a female who said she was Unsure.

5.51 Independent? Suppose a person is chosen at random. Use your understanding of commercial vehicle driving to decide whether the event that the person holds a valid commercial driving license and the event that the person drives a cab are independent or associated? Explain.

5.52 Independent? Suppose a person is chosen at random. Use your knowledge about literacy to decide whether the event that the person is above 20 years of age and the event that the person is illiterate are independent or associated? Explain.

5.53 Independent? About 12% of boys and 19% of girls in a school wear glasses. If we select a student at random, are the event that the student is a girl and the event that the student wears spectacles independent or associated?

5.54 Independent? Shoe sizes typically range from 4 to 12. Based on what you know about gender differences, if we randomly select a person, are the event that the shoe size is smaller than 6 and the event that the person is female independent or associated? Explain.

TRY **5.55 College Poll (Example 11)** Refer to the table in Exercise 5.47. Suppose a person is randomly selected from this group. Is being female independent of answering YES?

* **5.56 College Poll** Assume a person is selected randomly from the group of people represented in the table in Exercise 5.47. The probability that the person says Yes given that the person is a woman is 577/722, or 79.9%. The probability that the person is a woman given that the person says Yes is 577/978, or 59.0%, and the probability that the person says Yes and is a woman is 577/1275, or 45.3%. Why is the last probability the smallest?

TRY g **5.57 Hand Folding (Example 12)** When people fold their hands together with interlocking fingers, most people are more comfortable with one of two ways. In one way, the right thumb ends up on top and in the other way, the left thumb is on top. The table shows the data from one group of people. M means man, and W means woman; Right means the right thumb is on top, and Left means the left thumb is on top. Judging on the basis of this data set, are the events "right thumb on top" and male independent or associated? Data were collected in a class taught by one of the authors but were simplified for clarity. The conclusion remains the same as that derived from the original data. *See page 269 for guidance.*

	M	W
Right	18	42
Left	12	28

5.58 Dice When two dice are rolled, is the event "the first die shows a 1 on top" independent of the event "the second die shows a 1 on top"?

TRY **5.59 Happiness and Traditional Views (Example 13)** In the 2012 General Social Survey (GSS), people were asked about their happiness and were also asked whether they agreed with the following statement: "In a marriage the husband should work, and the wife should take care of the home." The table summarizes the data collected.

	Agree	Don't Know	Disagree
Happy	242	65	684
Unhappy	45	30	80

a. Include the row totals, the column totals, and the grand total in the table. Show the complete table with the totals.

b. Determine whether, for this sample, being happy is independent of agreeing with the statement.

5.60 Happiness Using the table in Exercise 5.59, determine whether being unhappy is independent of disagreeing with the statement for this sample.

TRY **5.61 Coin (Example 14)** Imagine flipping three fair coins.

a. What is the theoretical probability that all three come up heads?

b. What is the theoretical probability that the first toss is tails AND the next two are heads?

5.62 Die Imagine rolling a fair six-sided die three times.

a. What is the theoretical probability that all three rolls of the die show a 1 on top?

b. What is the theoretical probability that the first roll of the die shows a 6 AND the next two rolls both show a 1 on the top.

TRY **5.63 Die Sequences (Example 15)** Roll a fair six-sided die five times, and record the number of spots on top. Which sequence is more likely? Explain.

Sequence A: 66666

Sequence B: 16643

5.64 Babies Assume that babies born are equally likely to be boys (B) or girls (G). Assume a woman has 6 children, none of whom are twins. Which sequence is more likely? Explain.

Sequence A: GGGGGG

Sequence B: GGGBBB

5.65 Recidivism (Example 16) Norway's recidivism rate is one of the lowest in the world at 20%. This means that about 20% of released prisoners end up back in prison (within three years). Suppose three randomly selected prisoners who have been released are studied.

a. What is the probability that all three of them go back to prison? What assumptions must you make to calculate this?

b. What is the probability that neither of them goes back to prison?

c. What is the probability that at least two go back to prison?

5.66 Replacement of Helmets Use of obsolete helmets by bikers in the United Kingdom in 2016 is estimated at 40%, which means 40% of bikers use helmets that have become obsolete. Suppose two independent bikers have been randomly selected.

a. What is the probability that neither of them is using obsolete helmets?

b. What is the probability that both of them are using obsolete helmets?

c. What is the probability that at least one of them is using an obsolete helmet?

SECTION 5.4

TRY **5.67 Simulating Coin Flips (Example 17)**

a. Simulate flipping a coin 20 times. Use the line of random numbers below to obtain and report the resulting list of heads and tails. Use odd numbers (1, 3, 5, 7, 9) for heads and even numbers for tails (0, 2, 4, 6, 8).

1 4 7 0 9	9 3 2 2 0	8 9 5 4 7	9 5 3 2 0

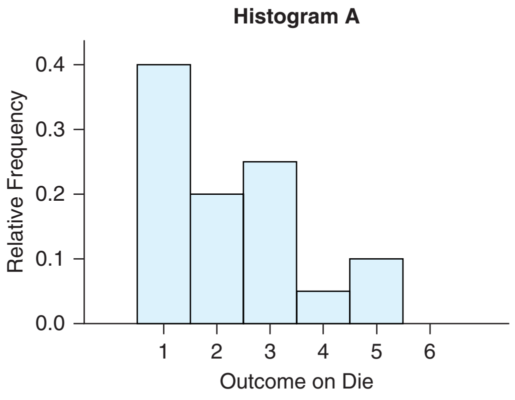
b. Judging on the basis of these 20 trials, what is the empirical probability of getting heads?

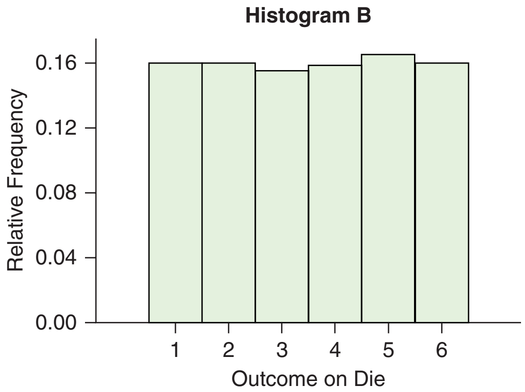
* **5.68 Simulation**

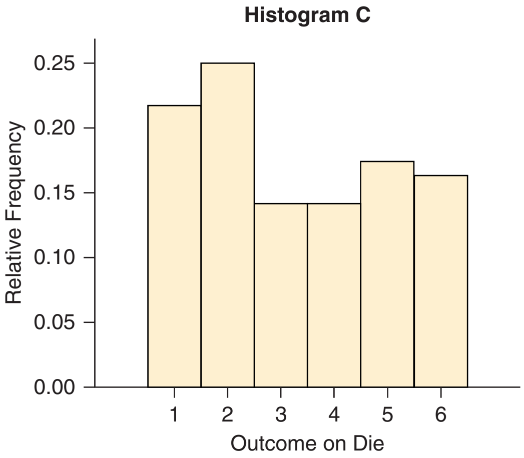
a. Explain how you could use digits from a random number table to simulate rolling a fair eight-sided die with outcomes 1, 2, 3, 4, 5, 6, 7, and 8 equally likely. Assume that you want to know the probability of getting a 1.

b. Carry out your simulation, beginning with line 5 of the random number table in Appendix A. Perform 20 repetitions of your trial. Using your results, report the empirical probability of getting a 1, and compare it with the theoretical probability of getting a 1.

5.69 Law of Large Numbers Refer to Histograms A, B, and C, which show the relative frequencies from experiments in which a fair six-sided die was rolled. One histogram shows the results for 20 rolls, one the results for 100 rolls, and another the results for 10,000 rolls. Which histogram do you think was for 10,000 rolls, and why?







5.70 Law of Large Numbers The table shows the results of rolling a fair six-sided die.

Outcome on Die	20 Trials	100 Trials	1000 Trials
1	8	20	167
2	4	23	167
3	5	13	161
4	1	13	166
5	2	16	172
6	0	15	167

Using the table, find the empirical probability of rolling a 1 for 20, 100, and 1000 trials. Report the theoretical probability of rolling a 1 with a fair six-sided die. Compare the empirical probabilities to the theoretical probability, and explain what they show.

5.71 Coin Flips Imagine flipping a fair coin many times. Explain what should happen to the proportion of heads as the number of coin flips increases.

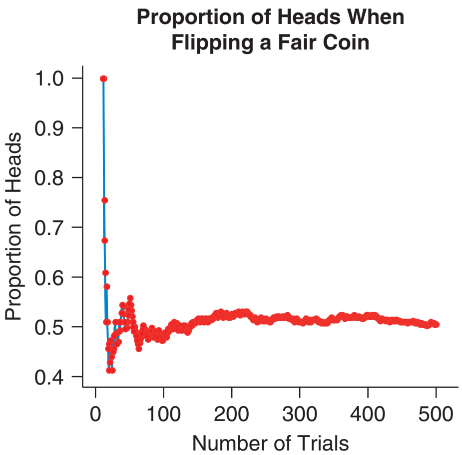
5.72 Coin Flips, Again Refer to the figure.

a. After a large number of flips, the overall proportion of heads "settles down" to nearly what value?

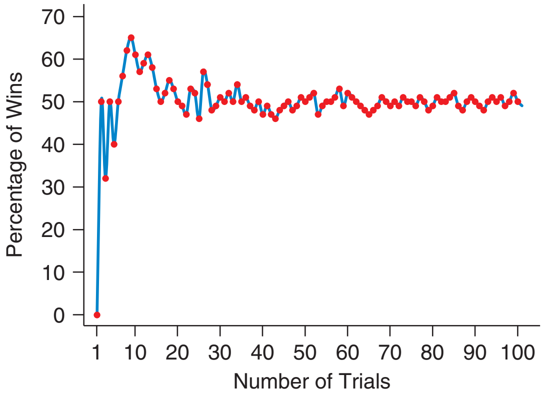
b. Approximately how many coin flips does it take before the proportion of heads settles down?

c. What do we call the law that causes this settling down of the proportion?

d. From the graph, determine whether the first flip was heads or tails.



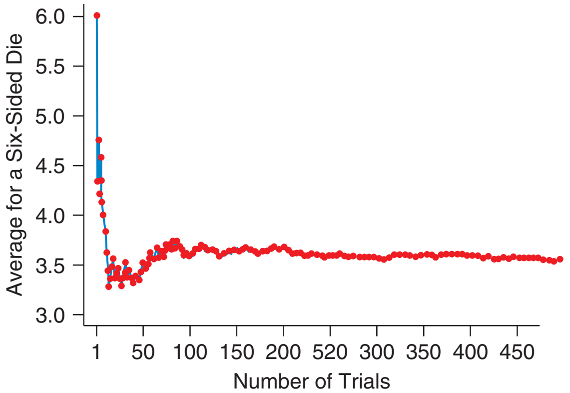
5.73 Law of Large Numbers Alfred and David are rolling dice. Whoever rolls the lower number wins the bet. If both roll the same number (for example, both roll five), they try again. Alfred and David do this 100 times. Laura and June are doing the same thing but only 20 times. Is it Alfred or June who is more likely to end having very close to 50% wins? Explain. You may refer to the graph to help you decide. It is one simulation based on 100 trials.



* **5.74 LLN: Organizations** Consider two corporate organizations. The first organization has 124 employees and the second organization has 15 employees. Which of the two organizations is more likely to have between 40% and 60% female employees, assuming that both males and females have equal employment opportunities in both the organizations? Why?

5.75 LLN: Card If you draw a card from a fair deck of cards and the first five draws (after replacement each time) are black cards, are you more likely to get a red card on the next draw, more likely to draw a black card again, or equally likely to get a red or a black card?

5.76 LLN: Die The graph shows the average when a six-sided die is rolled repeatedly. For example, if the first two rolls resulted in a 6 and a 2, the average would be 4. If the next trial resulted in a 1, the new average would be $(6 + 2 + 1)/3 = 3$. Explain how the graph demonstrates the Law of Large Numbers.



5.77 State Assembly A state assembly is supposed to represent the population. We wish to perform a simulation to determine an empirical probability that an assembly of 50 representatives has 25 or fewer males. Assume that about 50% of the population is male, so the probability that a person who has been elected to the state assembly is a male is 50%. Using a random number table, we decide that each digit will represent an assembly member. The digits 0–25, we decide, will represent a male member, and 26–50 will represent a female. Why is this a bad choice for this simulation?

5.78 Eye Color Some estimates say that 60% of the population has brown eyes. We wish to design a simulation to find an empirical probability that if 10 babies are born on a single day, at least 6 will have brown eyes. Suppose we decide that the numbers 0–4 will represent babies with brown eyes and the numbers 5–9 will represent babies with eyes of other colors. Explain what is wrong with the stated simulation method, and provide a correct method.

*** 5.79 Simulation: Four-Sided Die**

a. Explain how you could use a random number table (or the random numbers generated by software or a calculator) to simulate rolling a fair *four-sided* die 20 times. Assume you are interested in the probability of rolling a 1. Then report a line or two of the random number table (or numbers generated by a computer or calculator) and the values that were obtained from it.

b. Report the empirical probability of rolling a 1 on the four-sided die from part a, and compare it with the theoretical probability of rolling a 1.

*** 5.80 Simulation: Six-Sided Die**

a. Explain how you could use a random number table to simulate rolling a fair six-sided die 20 times. Assume you wish to find the probability of rolling a 1. Then report a line or two of the random number table (or numbers generated by a computer or calculator) and the values that were obtained from it.

b. Report the empirical probability of rolling a 1 from part a, and compare it with the theoretical probability of rolling a 1.

CHAPTER REVIEW EXERCISES

5.81 Capitalism According to a Pew poll conducted in 2012, 228 out of 380 Republicans viewed "Capitalism" as positive. If one Republican is randomly chosen from those 380, what is the probability that the person has a positive view of "Capitalism"?

5.82 Socialism According to a Pew poll conducted in 2012, 235 out of 489 Democrats viewed "Socialism" as positive. If one Democrat is randomly chosen from those 489, what is the probability that the person has a positive view of "Socialism"?

5.83 Independent Variables Use your general knowledge to label the following pairs of variables as independent or associated. Explain.

a. For a sample of artists, number of visitors visiting their art exhibitions and nationality of the artists.

b. For a sample of horses, their breed and body weight.

5.84 Independent Variables Use your general knowledge to label the following pairs of variables as independent or associated. Explain.

a. For a sample of adults, gender and ring size.

b. The outcome on rolls of two separate, fair dices.

5.85 Horse Racing According to CNN statistics, in British flat racing, 63% of winning horses are males while 37% are females. Now suppose that one male and one female horse are selected.

a. What is the probability that both the horses are race winners?

b. What is the probability that neither of them is a race winner?

c. What is the probability that one is a race winner and the other is not?

d. What is the probability that at least one of them is a race winner?

5.86 Literacy in 2015 The UNESCO Institute for Statistics reported that the literacy rate in Zimbabwe was 88.5% for males and 84.6% for females. Suppose these are accurate percentages. Now suppose a random man and a random woman meet.

a. What is the probability that both of them are literate?

b. What is the probability that neither of them is literate?

c. What is the probability that one of them is literate?

d. What is the probability that at least one of them is literate?

5.87 Internet Access A 2013 Pew poll said that 93% of young adults in the United States have Internet access. Assume that this is still correct.

a. If two people are randomly selected, what is the probability that they both have Internet access?

b. If the two people chosen were a married couple living in the same residence, explain why they would not be considered independent with regard to Internet access.

5.88 GPA The probability of a randomly selected person having a GPA of 8.5 or above in all subjects is 0.25.

a. If two students are chosen randomly and independently, what is the probability that they both have a GPA of 8.5 or above?

b. If two students are selected from the same high school statistics class, do you think the probability of their both having a GPA of 8.5 or above is different from your answer to part a? Explain.

*** 5.89 Marriage Anniversaries** Suppose all the months of the year are equally likely as marriage anniversaries. Glen and Shahid are two randomly selected married males (unrelated).

a. What is the probability that they were both married in August?

b. What is the probability that Glen OR Shahid was married in August? *Hint:* The answer is not 2/12 or 1/6. Refer to Guided Exercise 5.25 if you need help.

*** 5.90 Use of Seatbelts** According to the National Highway Traffic Safety Administration (NHTSA), about 87% of the total population of driving adults used seatbelts in 2014 (www.exchange.aaa.com). Suppose that Wilson and Dannis are randomly selected adults driving a car.

a. What is the probability that they both are wearing seatbelts?

b. What is the probability that Wilson OR Dannis is wearing a seatbelt?

5.91 Rich Happier: 2012 A Gallup poll asked, "Do you think that rich people in America today are happier than you, less happy, or about the same?" In 2012, 27% said less happy, 11% said happier, and 57% said about the same. The reason these don't add to 100% is that there were some people who had no opinion. Suppose Gallup were to do another survey polling 1500 people, and the percentages were the same as those in 2012.

a. How many respondents would say less happy?
b. How many would say happier?
c. How many would say about the same?

5.92 Rich Happier: 1990 A Gallup poll asked, "Do you think that rich people in America today are happier than you, less happy, or about the same?" In 1990, 36% said less happy, 11% said happier, and 50% said about the same. The reason these don't add up to 100% is that there were some people who had no opinion. Suppose Gallup were to do another survey polling 1500 people, and the percentages were the same as those in 1990.

a. How many would say less happy?
b. How many would say happier?
c. How many would say about the same?

5.93 Drinking Coffee The data collected in 2011–2012 by Australian Bureau of Statistics asked whether people consumed coffee and coffee substitutes on a daily basis. The table gives the total number of people in each age range (rounded) and the percentage who said they consumed coffee and coffee substitutes.

Age	Total Count	Consumed Coffee and Coffee Substitutes
9–18	300	16.4%
19–30	500	33.3%
31–50	100	62.6%
51+	200	66.8%

a. Make a two-way table of counts (not percentages). The table is started below:

	Consumed	Did not Consume
9–18	49	251
19–30		
31		

b. What is the tendency? Which group is most likely to consume coffee and coffee substitutes, and which group is least likely to consume coffee and coffee substitutes? Does this make sense to you?
c. From your table, if one person is selected, what is the probability that this person is in the age range of 19–30 AND consumes coffee and coffee substitutes on a daily basis?
d. If you select a person from the people who drink coffee, what is the probability that this person is in the age range of 19–30 years?
e. If you select a person in the age range of 19–30, what is the probability that this person consumes coffee?
f. Why is the probability in part c smaller than those in parts d and e?

5.94 Satisfaction with FDI In a CRISIL Survey conducted in India in April 2003, laborers were asked, "Are you satisfied with the Foreign Direct Investment (FDI) in the country?" In response, 62% of senior management, 48% of middle management, and 28% of wage earners said Yes. Assume that anyone who did not answer Yes answered No. Suppose the number of senior management employees polled was 700, the number of middle management employees was 900, and the number of wage earners was 400.

a. Create a two-way table with counts (not percentages) that starts as shown here.

	Top Level	Middle Level	Wage Earners
Yes	434		
No			

b. What is the probability that a person randomly selected is a wage earner given they said Yes?
c. What is the probability that a person randomly selected is a top-level manager given they said Yes?
d. What is the probability that a person randomly selected said Yes given they are a wage earner?
e. What is the probability that a person randomly selected from the entire group is a middle-level manager AND said No?

5.95 Car Thefts In a city, 28% of new cars were stolen within one year of purchase. Suppose two new cars were randomly selected. Assume that thefts are independent for the selected cars.

a. What is the probability that both of these cars will be stolen?
b. What is the probability that neither of these cars will be stolen?
c. What is the probability that one OR the other car will be stolen?

5.96 California Recidivism In California, the recidivism rate for prisoners is 67.5%. That is, 67.5% of those released from prison go back to prison within three years. This is one of the highest recidivism rates in the nation.

a. Suppose two independent prisoners are released. What is the probability that they will both go back to prison within three years?
b. What is the probability that neither will go back to prison within three years?
c. Suppose two independent prisoners are released. What is the probability that one OR the other (or both) will go back to prison within three years?

5.97 UK Recidivism and Gender Men return to prison at a higher rate than women (27.5% for men, compared to 19.1% for women) in the United Kingdom. For a randomly chosen prisoner, are the event that the person returns to prison and the event that the person is a woman independent?

5.98 Left-handedness Let's assume that around 13% of the total men and 6% of the total women in the world are left-handed. If we randomly select a person, are the event that the person is left-handed and the event that the person is female independent?

* **5.99** Construct a two-way table with 60 women and 80 men in which both groups show equal percentages of right-handedness.

* **5.100** Construct a two-way table with 60 women and 80 men in which there is a higher percentage of right-handed women.

* **5.101 Law of Large Numbers** A famous study by Amos Tversky and Nobel laureate Daniel Kahneman asked people to consider two hospitals. Hospital A is small and has 15 babies born per day. Hospital B has 45 babies born each day. Over one year, each hospital recorded the number of days that it had more than 60% girls born. Assuming that 50% of all babies are girls, which hospital had the most such

days? Or do you think both will have about the same number of days with more than 60% girls born? Answer, and explain. (Source: Amos Tversky. 2004. *Preference, Belief, and Similarity: Selected Writings*, ed. Eldar Shafir. Cambridge, Mass.: MIT Press, p. 205)

* **5.102 Law of Large Numbers** A certain professional basketball player typically makes 80% of his basket attempts, which is considered to be good. Suppose you go to several games at which this player plays. Sometimes the player attempts only a few baskets, say 10. Other times, he attempts about 60. On which of those nights is the player most likely to have a "bad" night, in which he makes much fewer than 80% of his baskets?

* **5.103 Simulating Guessing on a Multiple-Choice Test** Suppose a student takes a 10-question multiple-choice quiz, and for each question on the quiz there are five possible options. Only one option is correct. Now suppose the student, who did not study, guesses at random for each question. A passing grade is 3 (or more) correct. We wish to design a simulation to find the probability that a student who is guessing can pass the exam.

a. In this simulation, the random action consists of a student guessing on a question that has five possible answers. We will simulate this by selecting a single digit from the random number table given in this exercise.
In this table, we will let 0 and 1 represent correct answers, and 2 through 9 will represent incorrect answers. Explain why this is a correct approach for the exam questions with five possible answers. (This completes the first two steps of the simulation summary given in Section 5.4.)

b. A trial, in this simulation, consists of picking 10 digits in a row. Each digit represents one guess on a question on the exam. Write the sequence of numbers from the first trial. Also translate this to correct and incorrect answers by writing R for right and W for wrong. (This completes step 4.)

c. We are interested in knowing whether there were 3 or more correct answers chosen. Did this occur in the first trial? (This completes step 5.)

d. Perform a second simulation of the student taking this 10-question quiz by guessing randomly. Use the second line of the table given. What score did your student get? Did the event of interest occur this time?

e. Repeat the trial twice more, using lines 3 and 4 of the table. For each trial, write the score and whether or not the event occurred.

f. On the basis of these four trials, what is the empirical probability of passing the exam by guessing?

1 1 3 7 3	9 6 8 7 1
5 2 0 2 2	5 9 0 9 3
1 4 7 0 9	9 3 2 2 0
3 1 8 6 7	8 5 8 7 2

* **5.104 Simulating Guessing on a True/False Test** Perform a simulation of a student guessing on a true/false quiz with 10 questions. Use the same four lines of the random number table that are given for the preceding question. Write out each of the seven steps outlined in Section 5.4. Be sure to explain which numbers you will use to represent correct answers and which numbers for incorrect answers. Explain why your choice is logical. Do four repetitions, each trial consisting of 10 questions. Find the empirical probability of getting more than 5 correct out of 10.

5.105 Red Light/Green Light A busy street has three traffic lights in a row. These lights are not synchronized, so they run independently of each other. At any given moment, the probability that a light is green is 60%. Assuming that there is no traffic, follow the steps below to design a simulation to estimate the probability that you will get three green lights.

a. Identify the action with a random outcome, and give the probability that it is a success.

b. Explain how you will simulate this action using the random number table in Appendix A. Which digits will represent green and which non-green? If you want to get the same results we did, use all of the possible one-digit numbers (0, 1, 2, 3, 4, 5, 6, 7, 8, and 9), and let the first few represent the green lights. How many and what numbers would represent green lights, and what numbers would represent non-green lights?

c. Describe the event of interest.

d. Explain how you will simulate a single trial.

e. Carry out 20 repetitions of your trial, beginning with the first digit on line 11 of the random number table. For each trial, list the random digits, the outcomes they represent, and whether or not the event of interest happened.

f. What is the empirical probability that you get three green lights?

5.106 Soda A soda-bottling plant has a flaw in that 20% of the bottles it fills do not have enough soda in them. The sodas are sold in six-packs. Follow these steps to carry out a simulation to find the probability that three or more bottles in a six-pack will not have enough soda.

a. Identify the action with a random outcome, and explain how you will simulate this outcome using the random number table in Appendix A. If you want to get the same answers we got, use all the possible one digit numbers (0, 1, 2, 3, 4, 5, 6, 7, 8, and 9), and use some at the beginning of the list of numbers to represent bad and the rest to represent good. What numbers would represent bad and what numbers would represent good, and why?

b. Describe how you will simulate a single trial.

c. Describe the event of interest—that is, the event for which you wish to estimate a probability.

d. Carry out 10 trials, beginning with the first digit on line 15 of the random number table in Appendix A. For each trial, list the digits chosen, the outcomes they represent, and whether or not the event of interest occurred.

e. What is the experimental probability that you get three or more "bad" bottles in a six-pack?

5.107 GSS: Political Party The General Social Survey (GSS) is a survey done nearly every year at the University of Chicago. One survey, summarized in the table, asked each respondent to report her or his political party affiliation and whether she or he was liberal, moderate, or conservative. (Dem stands for Democrat, and Rep stands for Republican.)

	Dem	Rep	Other	Total
Liberal	306	26	198	530
Moderate	279	134	322	735
Conservative	104	309	180	593
Total	689	469	700	1858

a. If one person is chosen randomly from the group, what is the probability that the person is liberal?

b. If one person is chosen randomly from the group, what is the probability that the person is a Democrat?

5.108 GSS: Political Party Refer to the table given in Exercise 5.107.

a. If one person is chosen randomly from the group of 1858 people, what is the probability that the person is conservative?

b. If one person is chosen randomly from the group of 1858 people, what is the probability that the person is a Republican?

5.109 GSS: AND Refer to the table given in Exercise 5.107. Suppose we select a person at random from this collection of 1858 people. What is the probability the person is conservative AND a Democrat? In other words, find P(person is conservative AND person is a Democrat).

5.110 GSS: AND Refer to the table given in Exercise 5.107. Suppose we select a person at random from this collection of 1858 people. What is the probability that the person is liberal AND a Republican? In other words, find P(person is liberal AND person is a Republican).

5.111 GSS: OR Select someone at random from the 1858 people in the table given in Exercise 5.107. What is the probability that the person is a Democrat OR a Republican?

5.112 GSS: OR Select someone at random from the 1858 people in the table given in Exercise 5.107. What is the probability that the person is liberal OR conservative?

5.113 GSS: OR Assume one person is chosen randomly from the 1858 people in the table given in Exercise 5.107. What is the probability that the person is liberal OR a Democrat?

5.114 GSS: OR Assume that one person is chosen randomly from the table given in Exercise 5.107. What is the probability that the person is conservative OR a Republican?

5.115 GSS: Mutually Exclusive Referring to the table given in Exercise 5.107, name a pair of mutually exclusive events that could result from selecting an individual at random from this sample.

5.116 GSS: Mutually Exclusive Referring to the table given in Exercise 5.107, name a pair of events that are not mutually exclusive that could result from selecting an individual at random from this sample.

5.117 Political Party, Again A person is selected randomly from the sample summarized in the table for Exercise 5.107. We want to find the probability that a conservative person is a Democrat. Which of the following statements best describes the problem? (Choose one.)

i. P(conservative | Democrat) "conservative given Democrat"

ii. P(Democrat | conservative) "Democrat given conservative"

iii. P(Democrat AND conservative)

5.118 Political Party Use the table in Exercise 5.107. A person is selected randomly from the sample summarized in the table. We want to determine the probability that the person is a moderate Republican. Which of the following statements best describes the problem?

i. P(moderate | Republican)

ii. P(Republican | moderate)

iii. P(moderate AND Republican)

5.119 Political Party, Again Refer to the table for Exercise 5.107.

a. Find the probability that a randomly chosen respondent is a Democrat given that he or she is liberal. In other words, what percentage of the liberals are Democrats?

b. Find the probability that a randomly chosen respondent is a Democrat given he or she is conservative? In other words, what percentage of the conservatives are Democrats?

c. Which respondents are more likely to be Democrats: the liberal or the conservative respondents?

5.120 Party, Again Refer to the table for Exercise 5.107.

a. Find the probability that a randomly chosen respondent is conservative given that she or he is a Republican.

b. Find the probability that a randomly chosen respondent is a Republican given that he or she is conservative.

c. Find the probability that a randomly chosen respondent is both conservative AND a Republican.

5.121 Coin Flips Let H stand for heads and let T stand for tails in an experiment where a fair coin is flipped twice. Assume that the four outcomes listed are equally likely outcomes:

HH, HT, TH, TT

What are the probabilities of getting:

a. 0 heads?

b. Exactly 1 head?

c. Exactly 2 heads?

d. At least 1 head?

e. Not more than 2 heads?

5.122 Cubes A hat contains a number of cubes: 15 red, 10 white, 5 blue, and 20 black. One cube is chosen at random. What is the probability that it is:

a. A red cube?

b. Not a red cube?

c. A cube that is white OR black?

d. A cube that is neither white nor black?

e. What do the answers to part a and part b add up to and why?

5.123 Mutually Exclusive Suppose a person is selected at random. Label each pair of events as *mutually exclusive* or *not mutually exclusive*.

a. The person has brown eyes; the person has blue eyes.

b. The person is 50 years old; the person is a U.S. senator.

5.124 Mutually Exclusive Suppose a person is selected at random. Label each pair of events as *mutually exclusive* or *not mutually exclusive*.

a. The person is a parent; the person is a toddler.

b. The person is a woman; the person is a CEO (chief executive officer).

5.125 "OR" The *Times of India* reported that 21% of urban households in Delhi own one or more four-wheelers and 38.9% of urban households own one or more two-wheelers. From this information, is it possible to find the percentage of urban households that own a four-wheeler OR a two-wheeler? Why or why not?

5.126 "OR" Suppose you discovered that in a college campus, 40% of the female students were pursuing law and 30% of the female students were pursuing literature.

a. From this information, is it possible to determine the percentage of female students who were pursuing law OR literature?

b. If your answer to part a is no, what additional information would you need to answer this question?

5.127 UFOs When two people meet, they are sometimes surprised that they have similar beliefs. A survey of 1003 random adults conducted by the Scripps Survey Research Center at Ohio University found that 62 percent of men and 50 percent of women believe in intelligent life on other planets. Well, actually, they said it is either "very likely" or "somewhat likely" that intelligent life exists on other planets (www.reporternews.com).

a. If a man and a woman meet, what is the probability that they both believe in intelligent life on other planets?
b. If a man and a woman meet, what is the probability that neither believes in intelligent life on other planets?
c. What is the probability that the man and woman agree about life on other planets?
d. If a man and a woman meet, what is the probability that they have opposite beliefs on this issue?

5.128 Seat Belt Use In 2009, the National Highway Traffic Safety Administration said that 84% of drivers buckled their seat belts. Assume that this percentage is still accurate. If four drivers are randomly selected, what is the probability that they are all wearing their seat belts?

* **5.129 Independent** Imagine rolling a red die and a blue die. From this trial, name a pair of independent events.

* **5.130 Mutually Exclusive** Imagine rolling a red die and a blue die. From this trial, name a pair of mutually exclusive events.

5.131 Opinion about Music A Heartbeats International's survey in 2011 estimated that 40% of the world's population thought music is the most difficult to live without in their daily lives. If this rate is still correct and a new poll of 10,000 people were obtained, how many out of those 10,000 would you expect to think the same?

5.132 Greenhouse Effect An environmental issues survey asked the students of a college whether they felt that the main cause of the greenhouse effect is CFCs. It was estimated that 84% of all students felt that it was not so. If another poll was taken and there were 200 participants, how many would you expect to say that CFCs are not the main cause of the greenhouse effect, assuming the percentage remained the same?

GUIDED EXERCISES

g **5.25 College Poll OR** Refer to the table. Assume one person is chosen from the 1275 people in the table. Answer the question below by following the numbered steps. The question was whether college was worth the financial investment.

	Female	Male	All
No	45	56	101
Unsure	100	96	196
Yes	577	401	978
All	722	553	1275

QUESTION What is the probability that the person selected from the entire group is male OR said Yes?

Step 1 ▶ What is the probability that the person from the table is male?

Step 2 ▶ What is the probability that the person said Yes?

Step 3 ▶ If being male and saying YES were mutually exclusive, you could just add the probabilities from step 1 and 2 to find the probability that a person is male OR says YES. Are they mutually exclusive? Why or why not?

Step 4 ▶ What is the probability that a person is male AND said Yes?

Step 5 ▶ To find the probability that a person is male OR said Yes, why should you subtract the probability that a person is male AND said Yes from the sum as shown below?

$$P(\text{Male OR SaidYES}) = P(\text{Male}) + P(\text{SaidYES}) - P(\text{Male AND SaidYES})$$

Step 6 ▶ Do the calculation using the formula given in step 5.

Step 7: ▶ Report the answer in a sentence.

g **5.57 Hand Folding** When people fold their hands together with interlocking fingers, most people are more comfortable doing it in one of two ways. In one way, the right thumb ends up on top, and in the other way, the left thumb is on top. The table shows the data from one group of people.

	M	W
Right	18	42
Left	12	28

M means man, W means woman, Right means the right thumb is on top, and Left means the left thumb is on top.

QUESTION Say a person is selected from this group at random. Are the events "right thumb on top" and "male" independent or associated?

To answer, we need to determine whether the probability of having the right thumb on top given that you are a man is equal to the probability of having the right thumb on top (for the entire group). If so, the variables are independent.

Step 1 ▶ Figure out the marginal totals and put them into the table.

Step 2 ▶ Find the overall probability that the person's right thumb is on top.

Step 3 ▶ Find the probability that the right thumb is on top given that the person is a man. (What percentage of men have the right thumb on top?)

Step 4 ▶ Finally, are the variables independent? Why or why not?

TechTips

For All Technology

EXAMPLE: GENERATING RANDOM INTEGERS ▶ Generate four random integers from 1 to 6, for simulating the results of rolling a six-sided die.

TI-84

Seed First before the Random Integers

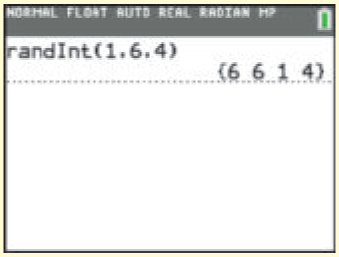
If you do not seed the calculator, everyone might get the same series of "random" numbers.

1. Enter the last four digits of your Social Security number or cell phone number and press **STO>**.
2. Then press **MATH**, choose **PROB**, and Press **ENTER** (to choose 1:rand). Press **ENTER** again.

You only need to seed the calculator once, unless you **Reset** the calculator. (If you want the same sequence later on, you can seed again with the same number.)

Random Integers

1. Press **MATH**, choose **PROB**, and press **5** (to choose 5:randInt).
2. Press **1**, **ENTER**, **6**, **ENTER**, **4**, **ENTER, ENTER**, and **ENTER**.



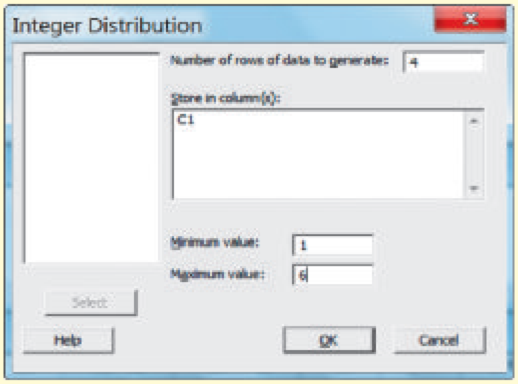
▲ **FIGURE 5A** TI-84

The first two digits (1 and 6 in Figure 5A) determine the smallest and largest integers, and the third digit (4 in Figure 5A) determines the number of random integers generated. The four numbers in the braces in Figure 5A are the generated random integers. Yours will be different. To get four more random integers, press **ENTER** again.

MINITAB

Random Integers

1. **Calc** > **Random Data** > **Integer**
2. See Figure 5B. Enter:
   Number of rows of data to generate, **4**
   Store in column(s), **c1**
   Minimum value, **1**
   Maximum value, **6**
3. Click **OK**.

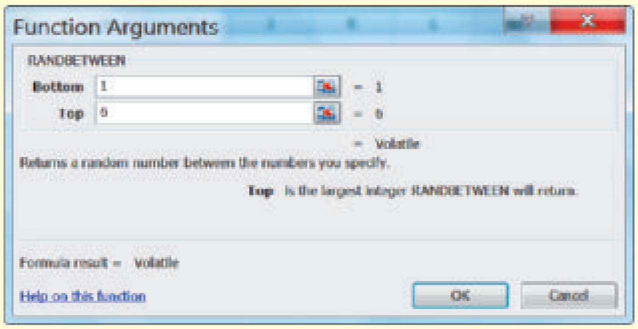


▲ **FIGURE 5B** Minitab

EXCEL

Random Integers

1. Click ***fx***, select a category **All**, and **RANDBETWEEN**.
2. See Figure 5C.
   Enter: **Bottom**, 1; **Top**, **6**.
   Click **OK**.
   You will get one random integer in the active cell in the spreadsheet.
3. To get more random integers, put the cursor at the lower right corner of the cell containing the first integer until you see a black cross (+), and drag downward until you have as many as you need.



▲ **FIGURE 5C** Excel

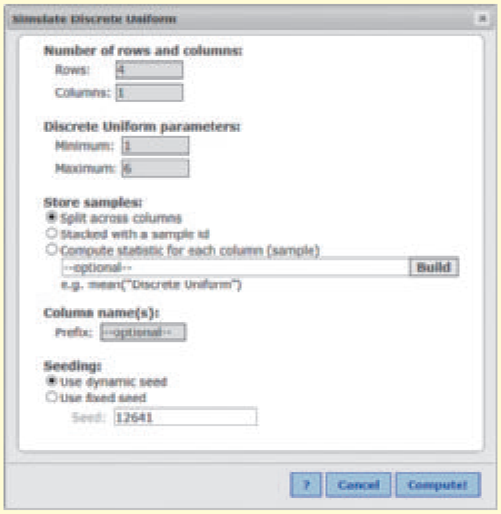
STATCRUNCH

Random Integers

1. **Data** > **Simulate** > **Discrete Uniform**
2. See Figure 5D.

   Enter **Rows**, **4**; **Columns**, **1**; **Minimum**, **1**; **Maximum**, **6** and leave **Split across columns** and **Use dynamic seed**.
3. Click **Compute!**

   You will get four random integers (from 1 to 6) in the first empty column.



▲ **FIGURE 5D** StatCrunch

6

Modeling Random Events: The Normal and Binomial Models

THEME

Probability distributions describe random outcomes in much the same way as the distributions we discussed in Chapter 2 describe samples of data. A probability distribution tells us the possible outcomes of a random trial and the probability that each of those outcomes will occur.

Random events can seem chaotic and unpredictable. If you flip a coin 10 times, there's no way you can say with absolute certainty how many heads will appear. There's no way your local weather forecaster can tell you with certainty whether it will rain tomorrow. Still, if we watch enough of these random events, patterns begin to emerge. We begin to learn how often different outcomes occur and to gain an understanding of what is common and what is unusual.

Science fiction writer Isaac Asimov is often quoted as having said, "The most exciting phrase to hear in science, the one that heralds new discoveries, is not Eureka! (I found it!) but rather 'hmm … that's funny . . . .'" When an outcome strikes us "funny" or unusual, it is because something unlikely occurred. This is exciting because it means a discovery has been made; the world is not the way we thought it was!

However, to know whether something unusual has happened, we first have to know how often it usually occurs. In other words, we need to know the probability. In the case study, you'll see a "homemade" example, in which one of the authors found that five ice cream cones ordered at a local McDonald's weighed more than the advertised amount. Was this a common occurrence, or did McDonald's perhaps deliberately understate the weight of the cones so that no one would ever be disappointed?

In this chapter we'll introduce a new tool to help us characterize probabilities for random events: the probability distribution. We'll examine the Normal probability distribution and the binomial probability distribution, two very useful tools for answering the question "Is this unusual?"

CASE STUDY

You Sometimes Get More Than You Pay For

A McDonald's restaurant near the home of one of the authors sells ice cream cones that, according to the "fact sheet" provided, weigh 3.18 ounces (converted from grams) and contain 150 calories. Do the ice cream cones really weigh exactly 3.18 ounces? To get 3.18 ounces for every cone would require a very fine-tuned machine or an employee with a *very* good sense of timing. Thus, we expect some natural variation in the weight of these cones. In fact, one of the authors bought five ice cream cones on different days. She found that each of the five cones weighed more than 3.18 ounces. Such an outcome, five out of five over the advertised weight, might occur just by chance. But how often? If such an outcome (five out of five) rarely happens, it's pretty surprising that it happened to us. In that case, perhaps McDonald's actually puts in more than 3.18 ounces. By the end of this chapter, after learning about the binomial and Normal probability distributions, you will be able to measure how surprising this outcome is—or is not.

SECTION 6.1

Probability Distributions Are Models of Random Experiments

A **probability model** is a description of how a statistician thinks data are produced. We use the word *model* to remind us that our description does not really explain how the data came into existence, but we hope that it describes the actual process fairly closely. We can tell whether a model is good by noting whether the probabilities that it predicts are matched by real-life outcomes. Thus, if a model says that the probability of getting heads when we flip a coin is 0.54, but in fact we get heads 50% of the time, we suspect that the model is not a good match.

A **probability distribution**, sometimes called a **probability distribution function (pdf)**, is a tool that helps us by keeping track of the outcomes of a random experiment and the probabilities associated with those outcomes. For example, suppose the playlist on your mp3 player has 10 songs: 6 are Rock, 2 are Country, 1 is Hip-hop, and 1 is Opera. Put your player on shuffle. What is the probability that the first song chosen is Rock?

The way the question is worded ("What is the probability . . . is Rock"?) means that we care about only two outcomes: Is the song classified as Rock or is it not? We could write the probabilities as shown in Table 6.1.

Outcome	Probability
Rock	6/10
Not Rock	4/10

▲ **TABLE 6.1** Probability distribution of songs.

Table 6.1 is a very simple probability distribution. It has two important features: It tells us all the possible outcomes of our random experiment (Rock or Not Rock), and it tells us the probability of each of these outcomes. All probability distributions have these two features, though they are not always listed so clearly.

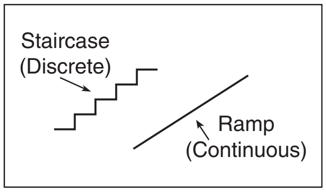
KEY POINT A probability distribution tells us (1) all the possible outcomes of a random experiment, and (2) the probability of each outcome.

Looking Back

Distributions of a Sample
Probability distributions are similar to distributions of a sample, which were introduced in Chapter 2. A distribution of a sample tells us the values in the sample and their frequency. A probability distribution tells us all possible values of the random experiment and their probability.

In Chapter 1, we classified variables as either numerical or categorical. It is now useful to break the numerical variables down into two more categories. **Discrete outcomes** (or discrete variables) are numerical values that you can list or count. An example is the number of phone numbers stored on the phones of your classmates. **Continuous outcomes** (or continuous variables) cannot be listed or counted because they occur over a range. The length of time your next phone call will last is a continuous variable. Refer to Figure 6.1 for a visual comparison of these terms.

This distinction is important because if we can list the outcomes, as we can for a discrete variable, then we have a nice way of displaying the probability distribution. However, if we are working with a continuous variable, then we can't list the outcomes, and we have to be a bit more clever in describing the probability distribution function. For this reason, we treat discrete values separately from continuous variables.



▲ **FIGURE 6.1** Visual representation of discrete and continuous changes in elevation. Note that for the staircase (discrete outcomes), you can count the stairs.

EXAMPLE 1 Discrete or Continuous?

Consider these variables:

a. The weight of a submarine sandwich you're served at a deli.

b. The elapsed time from when you left your house to when you arrived in class this morning.

c. The number of people in the next passing car.

d. The blood-alcohol level of a driver pulled over by the police in a random sobriety check. (Blood-alcohol level is measured as the percent of the blood that is alcohol.)

e. The number of eggs laid by a randomly selected salmon as observed in a fishery.

QUESTION Identify each of these numerical variables as continuous or discrete.

SOLUTION The continuous variables are variables a, b, and d. Continuous variables can take on any value in a spectrum. For example, the sandwich might weigh 6 ounces or 6.1 ounces or 6.0013 ounces. The amount of time it takes you to get to class could be 5000 seconds or 5000.4 seconds or 5000.456 seconds. Blood-alcohol content can be any value between 0 and 1, including 0.0013 (or 0.13%), 0.0013333, 0.001357, and so on.

Variables c and e, on the other hand, are discrete quantities—that is, numbers that can be counted.

TRY THIS! Exercise 6.1

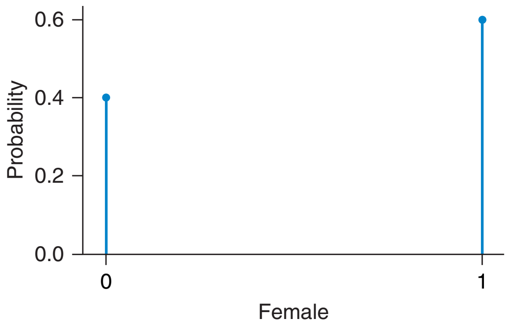
Discrete Probability Distributions Can Be Tables or Graphs

A statistics class at UCLA was approximately 40% male and 60% female. Let's arbitrarily code the males as 0 and the females as 1. If we select a person at random, what is the probability that the person is female?

Creating a probability distribution for this situation is as easy as listing both outcomes (0 and 1) and their probabilities (0.40 and 0.60). The easiest way to do this is in a table (see Table 6.2). However, we could also do it in a graph (Figure 6.2).

Female	Probability
0	0.40
1	0.60

▲ **TABLE 6.2** Probability distribution of gender in a class.

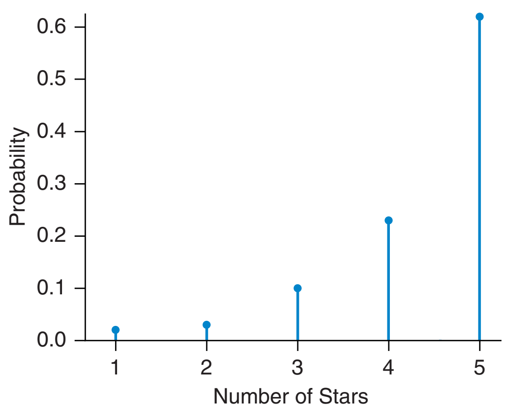


◀ **FIGURE 6.2** Probability distribution for selecting a person at random from a particular statistics class and recording whether the person is male (0) or female (1).

Amazon.com invites customers to rate books on a scale of 1 to 5 stars. At one visit to the site, *A Tale for the Time Being*, by Ruth Ozeki, had received 218 customer reviews. Suppose we randomly select a reviewer and base our decision whether to buy the book on how many stars that person gave to it. Table 6.3 and Figure 6.3 illustrate two different ways of representing the probability distribution for this random event. For example, we see that the most likely outcome is that the reviewer gave the book 5 stars. The probability that the reviewer gave the book 5 stars is 0.62, or 62%.

Number of Stars	Probability
5	0.62
4	0.23
3	0.10
2	0.03
1	0.02

▲ **TABLE 6.3** Probability distribution of number of stars.



◀ **FIGURE 6.3** Probability distribution for the number of stars given, as a rating, to a particular book on amazon.com by a randomly selected customer.

Note that the probabilities in Table 6.3 add to 1. This is true of all probability distributions: When you add up the probabilities for all possible outcomes, they add to 1. They have to, because there are no other possible outcomes.

Discrete Distributions Can Also Be Equations

What if we have too many outcomes to list in a table? For example, suppose a married couple decides to keep having children until they have a girl. How many children will they have, assuming that boys and girls are equally likely and that the gender of one birth doesn't depend on any of the previous births? It could very well turn out that their first child is a girl and they therefore have only one child. Or that the first is a boy but the second is a girl. Or, just possibly, they might never have a girl. The value of this experiment could be any number 1, 2, 3, … up to infinity. (Okay, in reality, it's impossible to have that many children. But we can imagine!)

We can't list all these values and probabilities in a table, and we can only hint at what the graph might look like. But we *can* write them in a formula:

The probability of having x children is $(1/2)^x$.

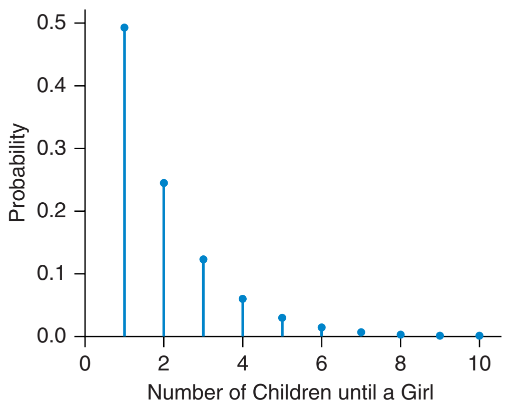
For example, the probability that they have 1 child (that is, the first is a girl) is $(1/2)^1 = 1/2$.

The probability that they have 4 children is $(1/2)^4 = 1/16$.

The probability that they have 10 children is small: $(1/2)^{10} = 0.00098$.

In this text, we will give the probabilities in a table or graph whenever convenient. In fact, even if it's not especially convenient, we will often provide a graph to encourage you to visualize what the probability distribution looks like. Figure 6.4 is part of the graph of the probability distribution for the number of children a couple can have if they continue to have children until the first girl. We see that the probability that the first child is a girl is 0.50 (50%). The probability that the couple has two children is half this: 0.25. The probabilities continue to decrease, and each probability is half the one before it.

FIGURE 6.4 Probability distribution of the number of children born until the first girl.



EXAMPLE 2 Playing Dice

Roll a fair six-sided die. A fair die is one in which each side is equally likely to end up on top. You will win \$4 if you roll a 5 or a 6. You will lose \$5 if you roll a 1. For any other outcome, you will win or lose nothing.

QUESTION Give a table that shows the probability distribution for the amount of money you will win. Draw a graph of this probability distribution function.

SOLUTION There are three outcomes: you win \$4, you win 0 dollars, you win −\$5. (Winning negative five dollars is the same as losing five dollars.)

You win \$4 if you roll a 5 or 6, so the probability is $2/6 = 1/3$.

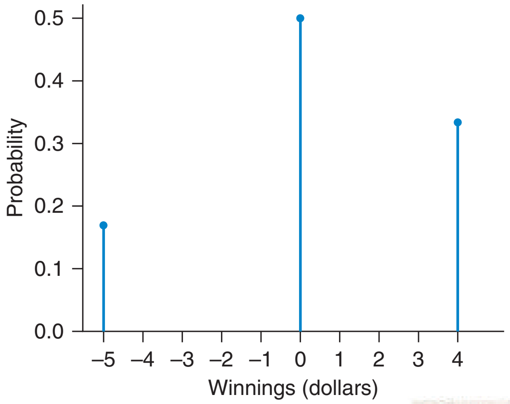
You win \$0 if you roll a 2, 3, or 4, so the probability is $3/6 = 1/2$

You "win" $-\$5$ if you roll a 1, so the probability is $1/6$.

We can put the probability distribution function in a table (Table 6.4), or we can represent the pdf as a graph (Figure 6.5).

Winnings	Probability
-5	$1/6$
0	$1/2$
4	$1/3$

▲ **TABLE 6.4** Probability distribution function of the dice game.



◀ **FIGURE 6.5** Probability distribution function of the dice game.

TRY THIS! Exercise 6.5

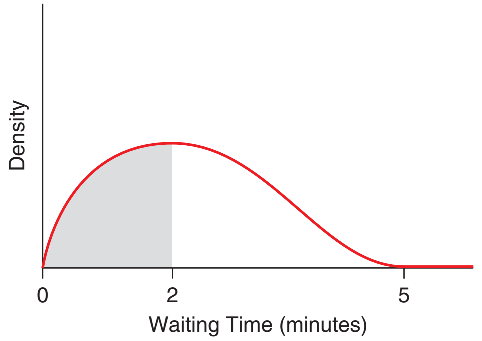
Continuous Probabilities Are Represented as Areas under Curves

Finding probabilities for continuous outcomes is more complicated, because we cannot simply list all the possible values we might see. What we can list is the *range of values* we might see.

For example, suppose you want to know the probability that you will wait in line for between 3 and 4 minutes when you go to the coffee shop. You can't list all possible outcomes that could result from your visit: 1.0 minute, 1.00032 minutes, 2.00000321 minutes. It would take (literally) an eternity. But you can specify a range. Suppose this particular coffee shop has done extensive research and knows that everyone gets helped in under 5 minutes. Therefore, all customers get helped within the range of 0 to 5 minutes.

If we want to find probabilities concerning a continuous-valued experiment, we also need to give a range for the outcomes. For example, the manager wants to know the probability that a customer will wait less than 2 minutes; this gives a range of 0 to 2 minutes.

The probabilities for a continuous-valued random experiment are represented as areas under curves. The curve is called a **probability density curve**. The total area under the curve is 1, because this represents the probability that the outcome will be somewhere on the x-axis. To find the probability of waiting between 0 and 2 minutes, we find the area under the density curve and between 0 and 2 (Figure 6.6).



◀ **FIGURE 6.6** Probability distribution of times waiting in line at a particular coffee shop.

The y-axis in a continuous-valued pdf is labeled "Density." How the density is determined isn't important. What *is* important to know is that the units of density are scaled so that the area under the entire curve is 1.

You might wonder where this curve came from. How did we know the distribution was exactly this shape? In practice, it is very difficult to know which curve is correct for a real-life situation. Statisticians call these curves "probability models" because they are meant to mimic a real-life probability, but we don't know—and can never know for sure—that the curve is correct. On the other hand, we can compare our probability predictions to the actual frequencies that we see. If they are close, then our probability model is good. For example, if this probability model predicts that 45% of customers get coffee within 2 minutes, then we can compare this prediction to an actual sample of customers.

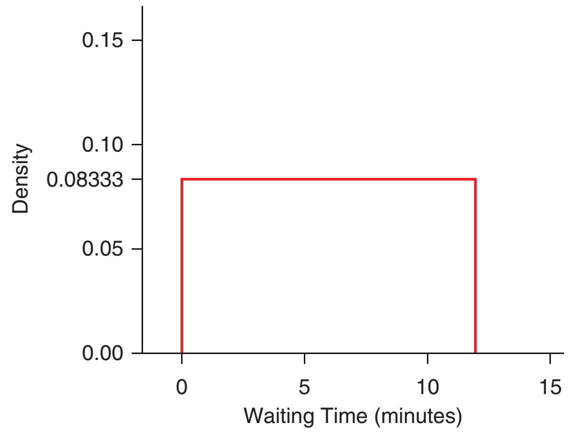
Finding Probabilities for Continuous-Valued Outcomes

Calculating the area under a curve is not easy. If you have a formula for the probability density, then you can sometimes apply techniques from calculus to find the area. However, for many commonly used probability densities, basic calculus is not helpful, and computer-based approximations are required.

In this book, you will always find areas for continuous-valued outcomes by using a table or by using technology. In Section 6.2 we introduce a table that can be used to find areas for one type of probability density that is very common in practice: the Normal curve.

EXAMPLE 3 Waiting for the Bus

The bus that runs near the home of one of the authors arrives every 12 minutes. If the author arrives at the bus stop at a randomly chosen time, then the probability distribution for the number of minutes he must wait for the bus is shown in Figure 6.7.

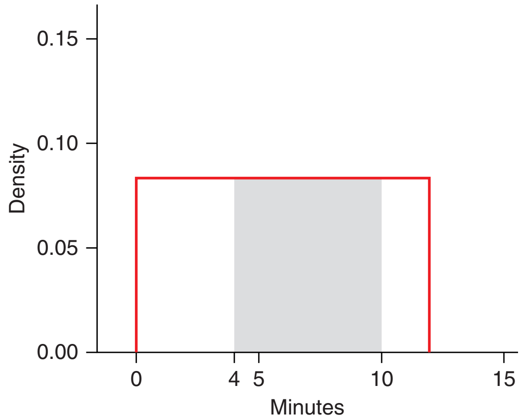


▶ **FIGURE 6.7** Probability distribution function showing the number of minutes the author must wait for the bus if he arrives at a randomly determined time.

QUESTION Find the probability that the author will have to wait between 4 and 10 minutes for the bus. (*Hint:* Remember that the area of a rectangle is the product of the lengths of the two sides.)

SOLUTION The distribution shown in Figure 6.7 is called a uniform distribution. Finding areas under this curve is easy because the curve is just a rectangular shape. The area we need to find is shown in Figure 6.8. The area of a rectangle is width times height. The width is $(10 - 4) = 6$, and the height is 0.08333.

The probability that the author must wait between 4 and 10 minutes is $6 \times 0.08333 = 0.49998$, or about 0.500. Visually, we see that about half of the area in Figure 6.8 is shaded.



◀ **FIGURE 6.8** The shaded area represents the probability that the author will wait between 4 and 10 minutes if he arrives at the bus stop at a randomly determined time.

There is approximately a 50% chance that the author must wait between 4 and 10 minutes.

TRY THIS! Exercise 6.11

SECTION 6.2

The Normal Model

The **Normal model** is the most widely used probability model for continuous numerical variables. One reason is that many numerical variables in which researchers have historically been interested have distributions for which the Normal model provides a very close fit. Also, an important mathematical theorem called the Central Limit Theorem (to be introduced in Chapter 7) links the Normal model to several key statistical ideas, which provides good motivation for learning this model.

We begin by showing you what the Normal model looks like. Then we discuss how to find probabilities by finding areas underneath the Normal curve. We will illustrate these concepts with examples and also discuss why the Normal model is appropriate for these situations.

Visualizing the Normal Distribution

Figure 6.9 on the next page shows several histograms of measurements taken from a sample of about 1400 adult men in the United States. All of these graphs have similar shapes: They are unimodal and symmetric. We have superimposed smooth curves over the histograms that capture this shape. You could easily imagine that if we continued to collect more and more data, the histogram would eventually fill in the curve and match the shape almost exactly.

The curve drawn on these histograms is called the **Normal curve**, or the **Normal distribution**. It is also sometimes called the Gaussian distribution, after Karl Friedrich Gauss (1777–1855), the mathematician who first derived the formula. Statisticians and scientists recognized that this curve provided a model that pretty closely described a good number of continuous-valued data distributions. Today, even though we have many other distributions to model real-life data, the Normal curve is still one of the most frequently used probability distribution functions in science.

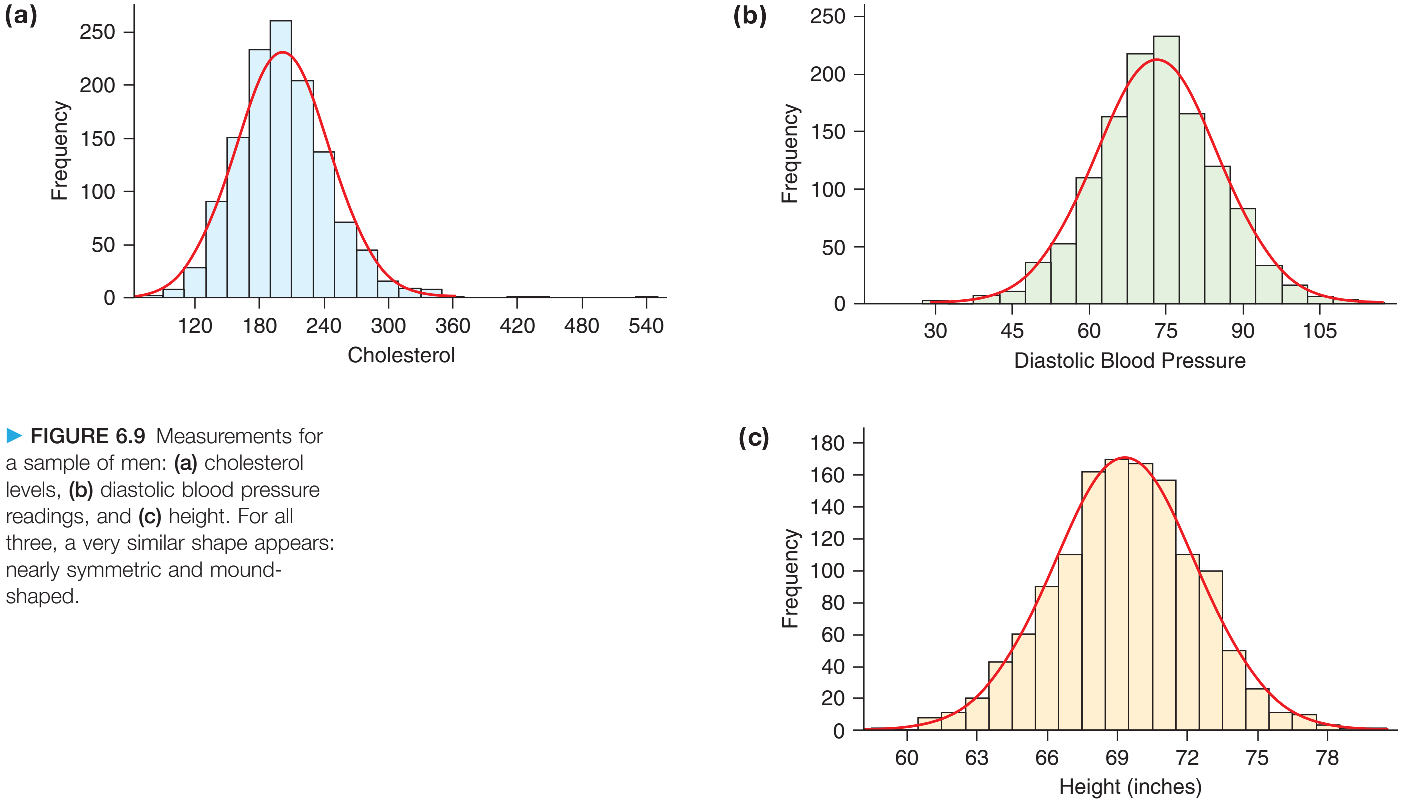
Looking Back

Unimodal and Symmetric Distributions
Symmetric distributions have histograms whose right and left sides are roughly mirror images of each other. Unimodal distributions have histograms with one mound.

Details

The Bell Curve
The Normal or Gaussian curve is also called the bell curve.

Center and Spread In Chapters 2 and 3 we discussed the center and spread of a distribution of *data*. These concepts are also useful for studying distributions of *probability*. The mean of a probability distribution sits at the balancing point of the



▶ **FIGURE 6.9** Measurements for a sample of men: **(a)** cholesterol levels, **(b)** diastolic blood pressure readings, and **(c)** height. For all three, a very similar shape appears: nearly symmetric and mound-shaped.

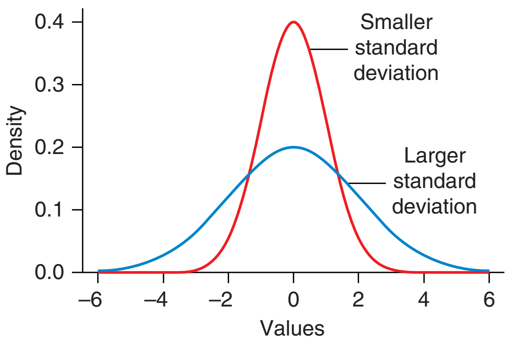
probability distribution. The standard deviation of a probability distribution measures the spread of the distribution by telling us how far away, typically, the values are from the mean. The conceptual understanding you developed for the mean and standard deviation of a sample still apply to probability distributions.

The notation we use is slightly different, so that we can distinguish means and standard deviations of probability distributions from means and standard deviations of data. The **mean of a probability distribution** is represented by the Greek character μ (mu, pronounced "mew"), and the **standard deviation of a probability distribution** is represented by the character σ (sigma). These Greek characters are used to avoid confusion of these concepts with their counterparts for samples of data, $\bar{x}$ and s.

Looking Back

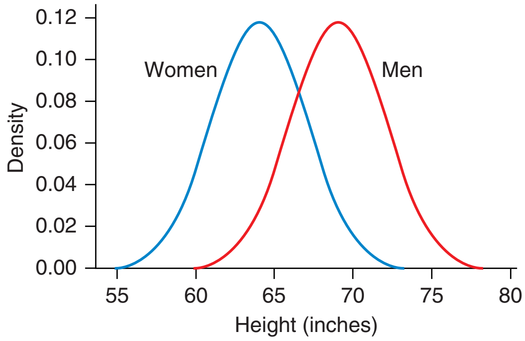
Mean and Standard Deviation
In Chapter 3, you learned that the symbol for the mean of a *sample* of data is $\bar{x}$, and the symbol for the standard deviation of a sample of data is s.

The Mean and Standard Deviation of a Normal Distribution The exact shape of the Normal distribution is determined by the values of the mean and the standard deviation. Because the Normal distribution is symmetric, the mean is in the exact center of the distribution. The standard deviation determines whether the Normal curve is wide and low (large standard deviation) or narrow and tall (small standard deviation). Figure 6.10 shows two Normal curves that have the same mean but different standard deviations.



▶ **FIGURE 6.10** Two Normal curves with the same mean but different standard deviations.

A Normal curve with a mean of 69 inches and a standard deviation of 3 inches provides a very good match to the distribution of heights of all adult men in the United States. Surprisingly, a Normal curve with the same standard deviation of 3 inches, but a smaller mean of about 64 inches, describes the distribution of adult women's heights. Figure 6.11 shows what these Normal curves look like.



◀ **FIGURE 6.11** Two Normal curves. The blue curve represents the distribution of women's heights; it has a mean of 64 inches and a standard deviation of 3 inches. The red curve represents the distribution of men's heights; it has the same standard deviation, but the mean is 69 inches.

The only way to distinguish among different Normal distributions is by their means and standard deviations. We can take advantage of this fact to write a short-hand notation to represent a particular Normal distribution. The notation $N(\mu, \sigma)$ represents a Normal distribution that is centered at the value of μ (the mean of the distribution) and whose spread is measured by the value of σ (the standard deviation of the distribution). For example, in Figure 6.11, the distribution of women's heights is $N(64, 3)$, and the distribution of men's heights is $N(69, 3)$.

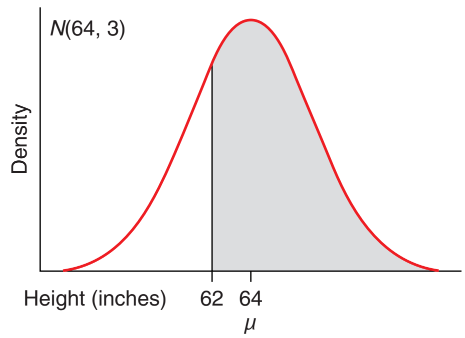
KEY POINT The Normal distribution is symmetric and unimodal ("bell-shaped"). The notation $N(\mu, \sigma)$ tells us the mean and standard deviation of the Normal distribution.

Finding Normal Probabilities

The Normal model $N(64, 3)$ gives a good approximation of the distribution of adult women's heights in the United States (where height is measured in inches). Suppose we were to select an adult woman from the United States at random and record her height. What is the probability that she is taller than a specified height?

Because height is a continuous numerical variable, we can answer this question by finding the appropriate area under the Normal curve. For example, Figure 6.12 shows a Normal curve that models the distribution of heights of women in the population—the same curve as in Figure 6.11. (We will often leave the numerical scale off the vertical axis from now on since it is not needed for doing calculations.) The area of the shaded region gives us the probability of selecting a woman taller than 62 inches. The entire area under the curve is 1.

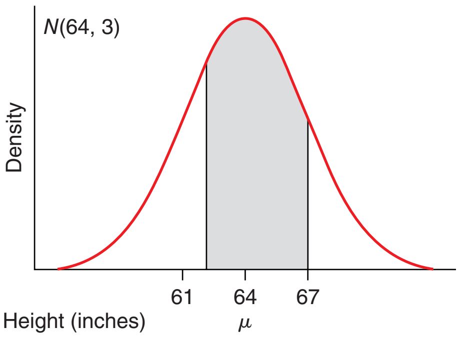
In fact, Figure 6.12 represents both the probability of selecting a woman taller than 62 inches and also the probability of selecting a woman *62 inches tall or taller*.



◀ **FIGURE 6.12** The area of the shaded region represents the probability of finding a woman taller than 62 inches from a $N(64, 3)$ distribution.

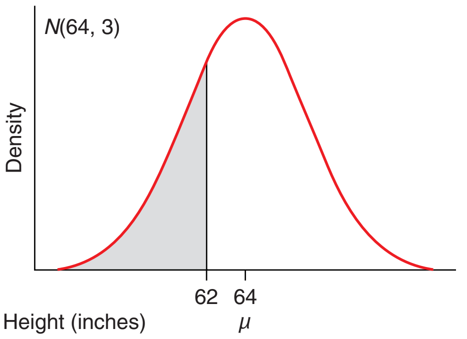
Because the areas for both regions (the one that is strictly greater than 62, and the other that includes 62) are the same, the probabilities are also the same. This is a convenient feature of continuous variables: We don't have to be too picky about our language when working with probabilities. This is in marked contrast with discrete variables, as you will soon see.

What if we instead wanted to know the probability that the chosen woman would be between 62 inches and 67 inches tall? That area would look like Figure 6.13.



▶ **FIGURE 6.13** The shaded area represents the probability that a randomly selected woman is between 62 and 67 inches tall. The probability distribution shown is the Normal distribution with mean 64 inches and standard deviation 3 inches.

Figure 6.14 shows the area corresponding to the probability that this randomly selected woman is less than 62 inches tall.



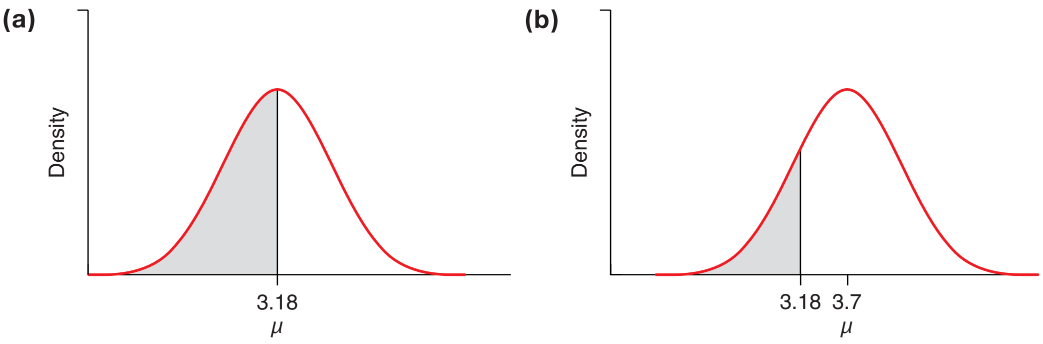
▶ **FIGURE 6.14** The shaded area represents the probability that the randomly selected woman is less than 62 inches tall.

KEY POINT When you are finding probabilities with Normal models, the first and most helpful step is to sketch the curve, label it appropriately, and shade in the region of interest.

We recommend that you always begin problems that concern Normal models by drawing a picture. One advantage of drawing a picture is that sometimes it is all you need. For example, according to the McDonald's "Fact Sheet," each serving of ice cream in a cone weighs 3.18 ounces. Now, we know that in real life it is not possible to serve exactly 3.18 ounces. Probably the employees operating the machines (or the machines themselves) actually dispense a little more or a little less than 3.18 ounces. Suppose that the amount of ice cream dispensed follows a Normal distribution with a mean of 3.18 ounces. What is the probability that a hungry customer will actually get less than 3.18 ounces?

Figure 6.15a shows the situation. From it we easily see that the area to the left of 3.18 is exactly half of the total area. Thus we know that the probability of getting less than 3.18 ounces is 0.50. (We also know this because the Normal curve is symmetric, so the mean—the balancing point—must sit right in the middle. Therefore, the probability of getting a value less than the mean is 0.50.)

What if the true mean is actually larger than 3.18 ounces? How will that affect the probability of getting a cone that weighs less than 3.18 ounces? Imagine "sliding" the Normal curve to the right, which corresponds to increasing the mean. Does the area to the left of 3.18 go up or down as the curve slides to the right? Figure 6.15b shows that



◀ **FIGURE 6.15** **(a)** A $N(3.18, 0.6)$ curve, showing that the probability of getting a cone that weighs less than 3.18 ounces is 50%. **(b)** A Normal curve with the same standard deviation (0.6) but a larger mean (3.7). The area below 3.18 is now much smaller.

the area below 3.18 is now smaller than 50%. The larger the mean amount of ice cream dispensed, the less likely it is that a customer will complain about getting too little.

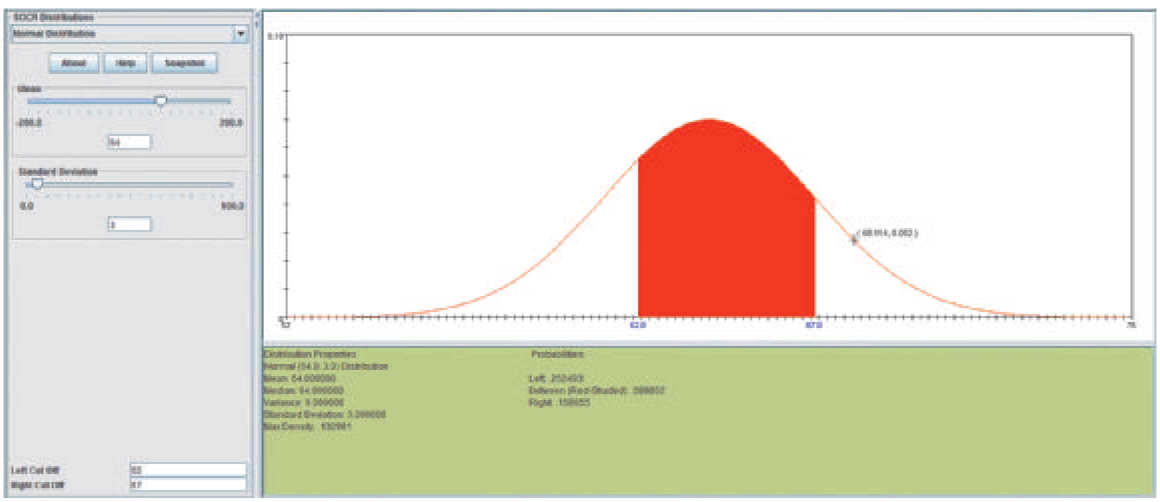
Finding Probability with Technology

Finding the area of a Normal distribution is best done with technology. Most calculators and many software packages will show you how to do this. We illustrate one such package, available free on the Internet, that you can use. We will also show you the "old-fashioned" way, which is useful when you do not have a computer handy. The old-fashioned way is also worth learning because it helps solidify your conceptual understanding of the Normal model.

Figure 6.16a is a screenshot from the SOCR (http://socr.stat.ucla.edu) calculator. It shows the probability that a randomly selected woman is between 62 and 67 inches tall if the $N(64, 3)$ model is a good description of the distribution of women's heights.

Tech

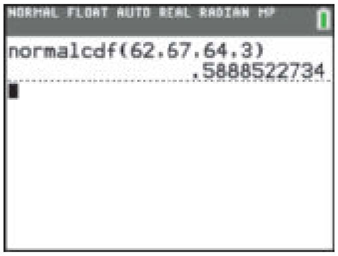
(a)



(b)

Distribution Properties	Probabilities
Normal (64.0, 3.0) Distribution	
Mean: 64.000000	Left: .252493
Median: 64.000000	Between (Red-Shaded): .588852
Variance: 9.000000	Right: .158655
Standard Deviation: 3.000000	
Max Density: .132981	

▲ **FIGURE 6.16** **(a)** Screenshot showing a calculation of the area under a $N(64, 3)$ curve between 62 and 67. **(b)** An enlarged view of the window below the graph, which shows the shaded and unshaded areas. The shaded area is given as .588852. The area to its left is given as .252493. (We will usually insert a 0 before the decimal point when we report or work with these values.)



▲ **FIGURE 6.17** TI-84 output showing that the probability that the woman is between 62 and 67 inches tall is about 59%.

To use this online calculator, the user moves sliders to set the mean and standard deviation and then uses the mouse to shade in the appropriate area. The actual probability is given in a window below the graph (shown on the previous page).

This particular calculator prints the probability next to the word "Between": .588852. Thus, the probability that the woman is *between* 62 and 67 inches tall is about 59%. Note that we can also, with no extra effort, get the probability that this woman will be shorter than 62 inches (.252493, or about 25%) or taller than 67 inches (about 16%.)

To use the SOCR calculator to find probabilities for the Normal model, go to http://socr.stat.ucla.edu, click on Distributions, and select Normal.

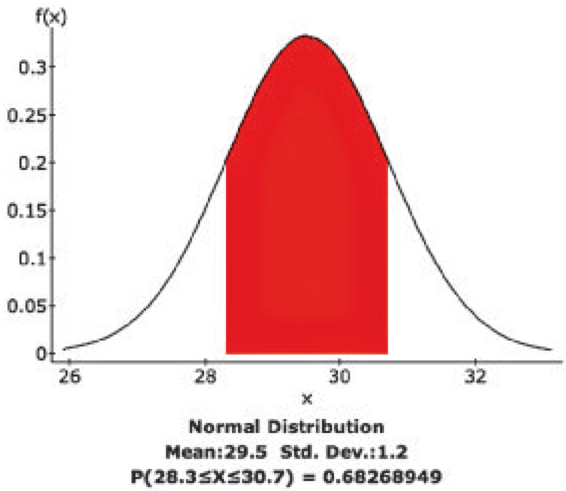
Figure 6.17 shows output from a TI-84 for the same calculation.

EXAMPLE 4 Baby Seals

Some research has shown that the mean length of a newborn Pacific harbor seal is 29.5 inches ($\mu = 29.5$ inches) and that the standard deviation is $\sigma = 1.2$ inches. Suppose that the lengths of these seal pups follow the Normal model.

QUESTION Using the StatCrunch output in Figure 6.18, find the probability that a randomly selected harbor seal pup is within 1 standard deviation of the mean length of 29.5 inches.

► **FIGURE 6.18** StatCruch Output: The shaded region represents the area under the curve between 28.3 inches and 30.7 inches. In other words, it represents the probability that the length of a randomly selected seal pup is within 1 standard deviation of the mean.



SOLUTION The phrase "within one standard deviation of the mean length" is one you will see often. (We used it in Chapter 3 when introducing the Empirical Rule.) It means that the pup's length will be somewhere between

mean minus 1 standard deviation

and

mean plus 1 standard deviation.

Because 1 standard deviation is 1.2 inches, this means the length must be between

$$29.5 - 1.2 = 28.3 \text{ inches}$$

and

$$29.5 + 1.2 = 30.7 \text{ inches}$$

From the results, we see that the probability that a randomly selected seal pup is within 1 standard deviation of mean length is about 68% (from Figure 6.18, it is 68.2689%).

TRY THIS! Exercise 6.17

Without Technology: The Standard Normal

Example 4 illustrates a principle that's very useful for finding probabilities from the Normal distribution without technology. This principle is the recognition that we don't need to refer to values in our distribution in the units in which they were measured. We can also refer to them in standard units. In other words, we can ask for the probability that the man's height is between 66 inches and 72 inches (measured units), but another way to say the same thing is to ask for the probability that his height is within 1 standard deviation of the mean, or between −1 and +1 standard units.

You can still use the Normal model if you change the units to standard units, but you must also convert the mean and standard deviation to standard units. This is easy, because the mean is 0 standard deviations away from itself, and any point 1 standard deviation away from the mean is 1 standard unit. Thus, if the Normal model was a good model, then when you convert to standard units, the $N(0, 1)$ model is appropriate.

This model—the Normal model with mean 0 and standard deviation 1—has a special name: the **standard Normal model**.

Looking Back

Standard Units
Standard units (Chapter 3) tell us how many standard deviations from the mean an observation lies. Values reported in standard units are called *z*-scores.

KEY POINT $N(0, 1)$ is the standard Normal model: a Normal model with a mean of 0 ($\mu = 0$) and a standard deviation of 1 ($\sigma = 1$).

The standard Normal model is a useful concept, because it allows us to find probabilities for any Normal model. All we need to do is first convert to standard units. We can then look up the areas in a published table that lists useful areas for the $N(0, 1)$ model. One such table is available in Appendix A.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	**.8413**	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015

TABLE 6.5 Excerpt from the Normal Table in Appendix A. This excerpt shows areas to the left of *z* in a standard Normal distribution. For example, the area to the left of $z = 1.00$ is 0.8413, and the area to the left of $z = 1.01$ is 0.8438.

Table 6.5 shows an excerpt from this table. The values within the table represent areas (probabilities). The numbers along the left margin, when joined to the numbers across the top, represent *z*-scores. For instance, the boldface value in this table represents the area under the curve *and to the left* of 1.00 standard unit. This represents the probability that a randomly selected person has a height *less than* 1 standard unit. Figure 6.19 shows what this area looks like.

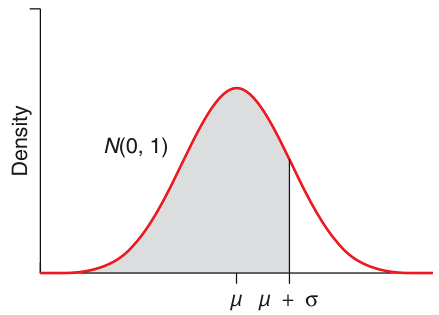


FIGURE 6.19 The area of the shaded region represents the probability that a randomly selected person (or thing) has a value less than 1.00 standard deviation above the mean, which is about 84%.

Details

***z*-Scores**

In Chapter 3 we gave the formula for a z-score in terms of the mean and standard deviation of a sample:

$$z = \frac{x - \bar{x}}{s}$$

The same idea works for probability distributions, but we change the notation to indicate that we are using the mean and standard deviation of a probability distribution:

$$z = \frac{x - \mu}{\sigma}$$

For example, imagine we want to find the probability that a randomly selected woman is shorter than 62 inches. This person is selected from a population of women whose heights follow a $N(64, 3)$ distribution. Our strategy is

1. Convert 62 inches to standard units. Call this number z.
2. Look up the area below z in the table for the $N(0, 1)$ distribution.

EXAMPLE 5 Small Pups

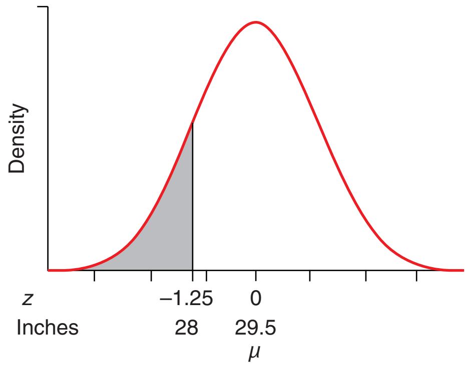
Small newborn seal pups have a lower chance of survival than larger newborn pups. Suppose that the length of a newborn seal pup follows a Normal distribution with a mean length of 29.5 inches and a standard deviation of 1.2 inches.

QUESTION What is the probability that a newborn pup selected at random is shorter than 28.0 inches?

SOLUTION Begin by converting the length 28.0 inches to standard units.

$$z = \frac{28 - 29.5}{1.2} = \frac{-1.5}{1.2} = -1.25$$

Next sketch the area that represents the probability we wish to find (see Figure 6.20). We want to find the area under the Normal curve and to the left of 28 inches, or, in standard units, to the left of −1.25.



▶ **FIGURE 6.20** A standard Normal distribution, showing the shaded area that represents the probability of selecting a seal pup shorter than −1.25 standard deviations below the mean.

We can look this up in the standard Normal table in Appendix A. Table 6.6 shows the part we are interested in. We see that the area to the left of a z-score of −1.25 is 10.56%.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	**.1056**	.1038	.1020	.1003	.0985

▲ **TABLE 6.6** Part of the Standard Normal Table. The value printed in boldface type is the area under the standard Normal density curve to the left of −1.25.

The probability that a newborn seal pup will be shorter than 28 inches is about 11% (rounding up from 10.56%).

TRY THIS! Exercise 6.25

EXAMPLE 6 A Range of Seal Pup Lengths

Tech

Again, suppose that the $N(29.5, 1.2)$ model is a good description of the distribution of seal pups' lengths.

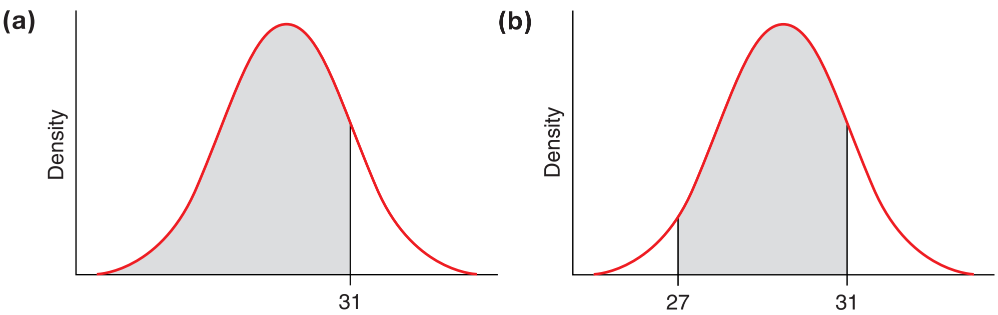
QUESTION What is the probability that this randomly selected seal pup will be between 27 inches and 31 inches long?

SOLUTION This question is slightly tricky. The table gives us only the area *below* a given value. How do we find the area *between* two values?

We proceed in two steps. First we find the area less than 31 inches. Second, we "chop off" the area below 27 inches. The area that remains is the region between 27 and 31 inches. This process is illustrated in Figure 6.21.

To find the area less than 31 inches, we convert 31 inches to standard units:

$$z = \frac{(31 - 29.5)}{1.2} = 1.25$$



◀ **FIGURE 6.21** Steps for finding the area between 27 and 31 inches under a $N(29.5, 1.2)$ distribution. **(a)** The area below 31. **(b)** We chop off the area below 27, and the remaining area (shaded) is what we are looking for.

Using the standard Normal table in Appendix A, we find that this probability is 0.8944. Next, we find the area below 27 inches:

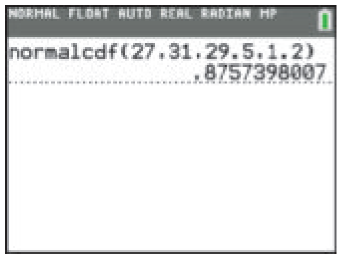
$$z = \frac{(27 - 29.5)}{1.2} = -2.08$$

Again, using the standard Normal table, we find this area to be 0.0188. Finally, we subtract (or "chop off") the smaller area from the big one:

$$\begin{array}{r} 0.8944 \\ -0.0188 \\ \hline 0.8756 \end{array}$$

CONCLUSION The probability that a newborn seal pup will be between 27 inches and 31 inches long is about 88%. Figure 6.22 confirms that answer.

TRY THIS! Exercise 6.29



▲ **FIGURE 6.22** TI-84 output for finding the probability that a newborn seal pup will be between 27 and 31 inches long.

Finding Measurements from Percentiles for the Normal Distribution

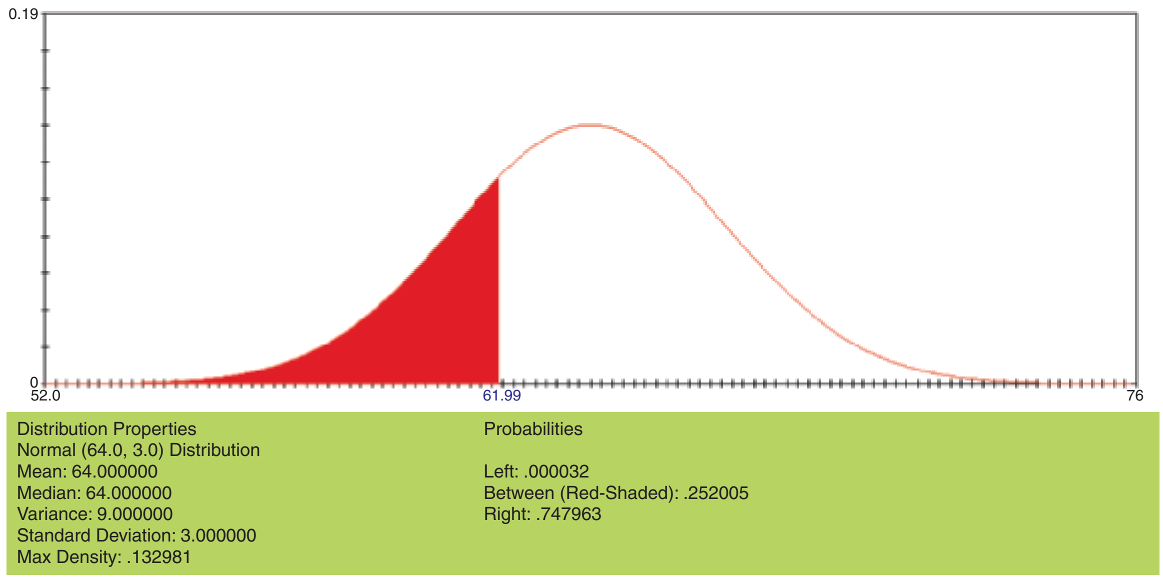
So far we have discussed how to find the probability that you will randomly select an object with a value that is within a certain range. Thus, in Example 5 we found that if newborn seal pups' lengths follow a $N(29.5, 1.2)$ distribution, then the probability that we will randomly select a pup shorter than 28 inches is (roughly) 11%. We found this by finding the area under the Normal curve that is to the left of 28 inches.

Details

Inverse Normal
Statisticians sometimes refer to finding measurements from percentiles of Normal distributions as "finding inverse Normal values."

Sometimes, though, we wish to turn this around. We are given a probability, but we want to find the value that corresponds to that probability. For instance, we might want to find the length of a seal pup such that the probability that we'll see any pups shorter than that is 11%. Such a number is called a **percentile**. The 11th percentile for seal pup lengths is 28 inches, the length that has 11% of the area under the Normal curve to its left.

Finding measurements from percentiles is simple with the right technology. The screenshot in Figure 6.23 shows how to use SOCR to find the height of a woman in the 25th percentile, assuming that women's heights follow a *N*(64, 3) distribution. Simply move the cursor until the shaded region represents the lower 25% of the curve. (This area is given after the words "Between (Red-Shaded)" below the graph.) We see that the value that has 25.2% (as close as we could get to 0.2500) below it is 61.99. So we say that 61.99 (or 62) is the 25th percentile for this distribution.



▲ **FIGURE 6.23** Technology shows that 61.99 inches is the 25th percentile, because it has 25% of the area below it (to the left of it). Thus, if a woman is about 62 inches tall, there's a 25 percent chance that another woman will be shorter than she.

Things are a little trickier if you don't have technology. You first need to use the standard Normal curve, *N*(0, 1), to find the *z*-score from the percentile. Then you must convert the *z*-score to the proper units. So without technology, finding a measurement from a percentile is a two-step process:

Step 1. Find the *z*-score from the percentile.

Step 2. Convert the *z*-score to the proper units.

Example 8 illustrates these steps.

EXAMPLE 7 Inverse Normal or Normal?

Suppose that the amount of money people keep in their online PayPal account follows a Normal model. Consider these two situations:

a. A PayPal customer wonders how much money he would have to put into the account to be in the 90th percentile.

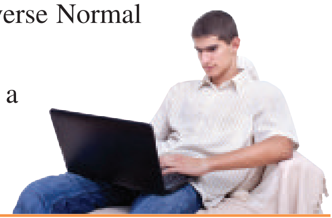
b. A PayPal employee wonders what the probability is that a randomly selected customer will have less than $150 in his account.

QUESTION For each situation, identify whether the question asks for a measurement or a Normal probability.

SOLUTIONS

a. This situation gives a percentile (the 90th) and asks for the measurement (in dollars) that has 90% of the other values below it. This is an inverse Normal question.

b. This situation gives a measurement ($150) and asks for a Normal probability.



TRY THIS! Exercise 6.39

EXAMPLE 8 Finding Measurements from Percentiles by Hand

Assume that women's heights follow a Normal distribution with mean 64 inches and standard deviation 3 inches: $N(64, 3)$. Earlier, we used technology to find that the 25th percentile was approximately 62 inches.

QUESTION Using the Normal table in Appendix A, confirm that the 25th percentile height is 62 inches.

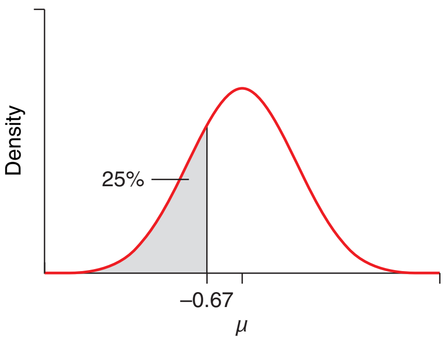
SOLUTION The question asks us to find a measurement (the height) that has 25% of all women's heights below it. Use the Normal table in Appendix A. This gives probabilities and percentiles for a Normal distribution with mean 0 and standard deviation 1: a $N(0, 1)$ or standard Normal distribution.

Step 1: Find the z-score from the percentile. To do this, you must first find the probability within the table. For the 25th percentile, use a probability of 0.25. Usually, you will not find exactly the value you are looking for, so settle for the value that is as close as you can get. This value, 0.2514, is underlined in Table 6.7, which is an excerpt from the Normal table in Appendix A.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	<u>.2514</u>	.2483	.2451

▲ **TABLE 6.7** Part of the Standard Normal Table.

You can now see that the z-score corresponding to a probability of 0.2514 is −0.67. This relation between the z-score and the probability is shown in Figure 6.24.



◀ **FIGURE 6.24** The percentile for 0.25 is −0.67 standard unit, because 25% of the area under the standard Normal curve is below −0.67.

Step 2: Convert the z-score to the proper units.

A z-score of -0.67 tells us that this value is 0.67 standard deviation below the mean. We need to convert this to a height in inches.

One standard deviation is 3 inches, so 0.67 standard deviation is

$$0.67 \times 3 = 2.0 \text{ inches}$$

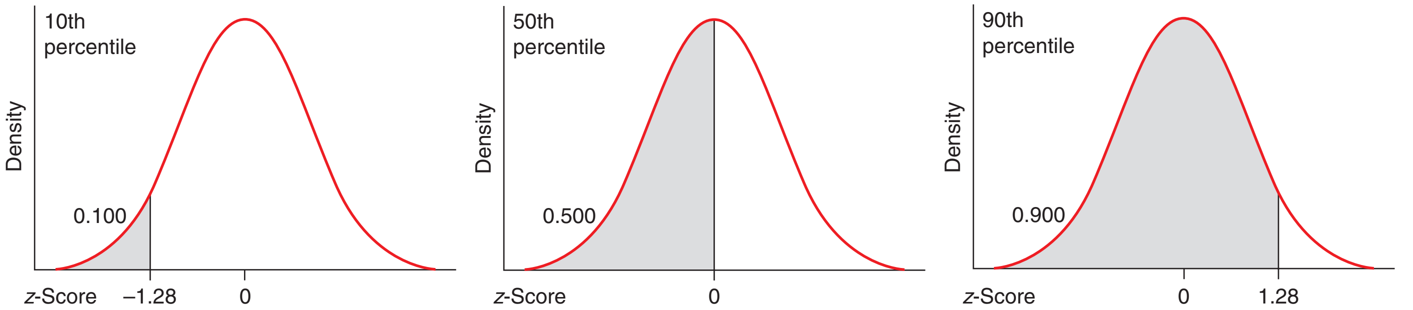
The height is 2.0 inches below the mean. The mean is 64.0 inches, so the 25th percentile is

$$64.0 - 2.0 = 62.0$$

CONCLUSION The woman's height at the 25th percentile is 62 inches, assuming that women's heights follow a $N(64, 3)$ distribution.

TRY THIS! Exercise 6.45

Figure 6.25 shows some percentiles and the corresponding z-scores to help you visualize percentiles.



▲ **FIGURE 6.25** z-Scores and percentiles—the 10th, 50th, and 90th percentiles. A percentile corresponds to the percentage in the area to the left under the curve.

The Normal Model and the Empirical Rule

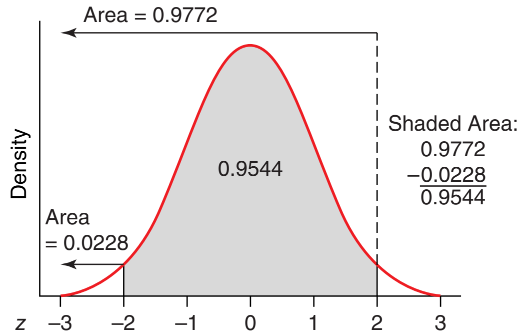
Looking Back

The Empirical Rule
The Empirical Rule says that if a distribution of a sample of data is unimodal and roughly symmetric, then about 68% of the observations are within 1 standard deviation of the mean, about 95% are within 2 standard deviations of the mean, and nearly all are within 3 standard deviations of the mean.

In Chapter 3 we mentioned the Empirical Rule, which is not so much a rule as a guideline for helping you understand how data are distributed. The Empirical Rule is meant to be applied to any symmetric, unimodal distribution. However, the Empirical Rule is based on the Normal model. For any arbitrary unimodal, symmetric distribution, the Empirical Rule is approximate. And sometimes very approximate. But if that distribution is the Normal model, the rule is (nearly) exact.

In Example 4, we found that the area between -1 and $+1$ in a standard Normal distribution was 68%. This is exactly what the Empirical Rule predicts. Similarly, we can find, as shown in Figure 6.26, that the area of the region between –2 and +2 in the $N(0, 1)$ model is 0.9544, or about 95%, just as the Empirical Rule predicts.

These facts from the Empirical Rule help us interpret the standard deviation in the context of the Normal distribution. For example, because the heights of women are Normally distributed and have a standard deviation of about 3 inches, we know that a majority of women (in fact, 68%) have heights within 3 inches of the mean: between 61 inches and 67 inches. Because nearly all women are within 3 standard deviations of the mean, we know we should not expect many women to be taller than $64 + 3 \times 3 = 73$ inches tall (73 inches is 6 feet and 1 inch). Such women are very rare.



◀ **FIGURE 6.26** The area between *z*-scores of −2.00 and +2.00 on the standard Normal curve.

SNAPSHOT THE NORMAL MODEL

WHAT IS IT? ▶ A model of a distribution for some numerical variables.

WHAT DOES IT DO? ▶ Provides us with a model of the distributions of probabilities for many real-life numerical variables.

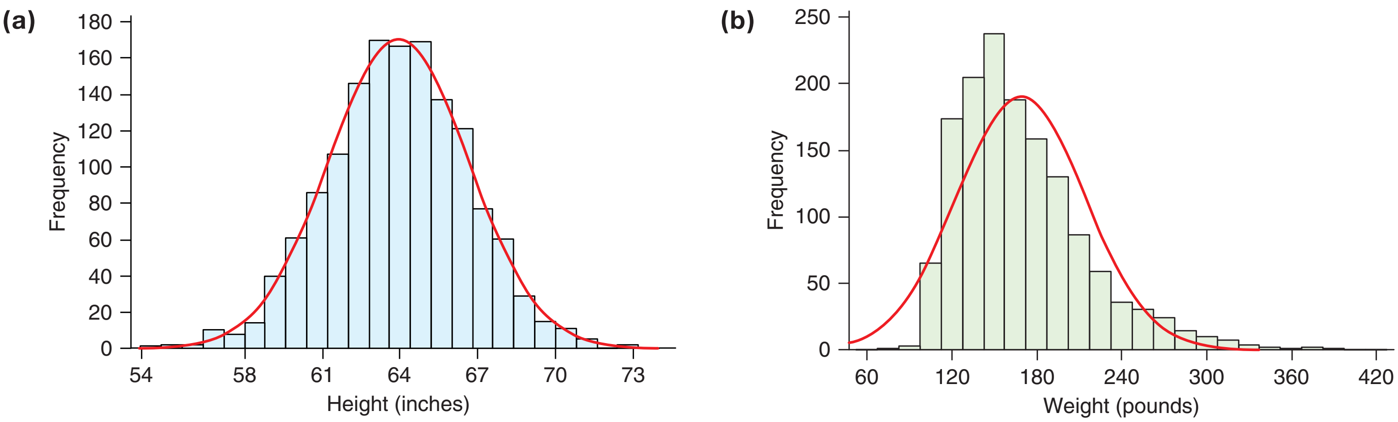
HOW DOES IT DO IT? ▶ The probabilities are represented by the area underneath the bell-shaped curve.

HOW IS IT USED? ▶ If the Normal model is appropriate, it can be used for finding probabilities or for finding measurements associated with particular percentiles.

Appropriateness of the Normal Model

The Normal model does not fit all distributions of numerical variables. For example, if we are randomly selecting people who submitted tax returns to the federal government, we cannot use the Normal model to find the probability that someone's income is higher than the mean value. The reason is that incomes are right-skewed, so the Normal model will not fit.

How do we know whether the Normal model is appropriate? If you've collected data, then you can check by making appropriate graphs. In short, the Normal model is appropriate if it produces results that match what we see in real life. If the data we collect match the Normal model fairly closely, then the model is appropriate. Figure 6.27a shows a histogram of the actual heights from a sample of more than 1400 women from the National Health and Nutrition Examination Survey (NHANES, www.cdc.gov/nchs/nhanes),



▲ **FIGURE 6.27** **(a)** A histogram of data from a large sample of adult women in the United States drawn at random from the National Health and Nutrition Examination Survey. The red curve is the Normal curve, which fits the shape of the histogram very well, indicating that the Normal model would be appropriate for these data. **(b)** A histogram of weights for the same women. Here the Normal model is a bad fit to the data.

with the Normal model superimposed over the histogram. Note that the model, though not perfect, is a pretty good description of the shape of the distribution. Compare this to Figure 6.27b, which shows the distribution of weights for the same women. The Normal model is not a very good fit for these data. The Normal model has the peak at the wrong place; specifically, the Normal model is symmetric, whereas the actual distribution is right-skewed.

Statisticians have several ways of checking whether the Normal model is a good fit to the population, but the easiest thing for you to do is to make a histogram of your data and see whether it looks unimodal and symmetric. If so, the Normal model is likely to be a good model.

SECTION 6.3

The Binomial Model (optional)

The Normal model applies to many real-life, continuous-valued numerical variables. The **binomial probability model** is useful in many situations with discrete-valued numerical variables (typically counts, whole numbers). As with the Normal model, we will explain what the model looks like and how to calculate probabilities with it. We will also provide examples and discuss why the binomial model is appropriate to the situations.

The classic application of the binomial model is counting heads when flipping a coin. Let's say I flip a coin 10 times. What is the probability I get 1 head? 2 heads? 10 heads? This is a situation that the binomial model fits quite well, and the probabilities of these outcomes are given by the binomial model. If we randomly select 10 people, what's the probability that exactly 5 will be Republicans? If we randomly select 100 students, what's the probability that 10 or fewer will be on the Dean's List? These are examples of situations where the binomial model applies.

How do you recognize a binomial model? The first sign that your random experiment is a candidate for the binomial model is that the outcome you are interested in is a count. If that's the case, then all four of the following characteristics must be present:

1. *A fixed number of trials*. We represent this number with the letter n. For example, if we flip a coin 10 times, then $n = 10$.
2. *Only two outcomes are possible at each trial*. We will call these two outcomes "success" and "failure." For example, we might consider the outcome of heads to be a success. Or we might be selecting people at random and counting the number of males; in this case, of the two outcomes "male" and "female," "male" might be considered a success.
3. *The probability of success is the same at each trial*. We represent this probability with the letter p. For example, the probability of getting heads after a coin flip is $p = 0.50$ and does not change from flip to flip.
4. *The trials are independent*. The outcome of one trial does not affect the outcome of any other trial.

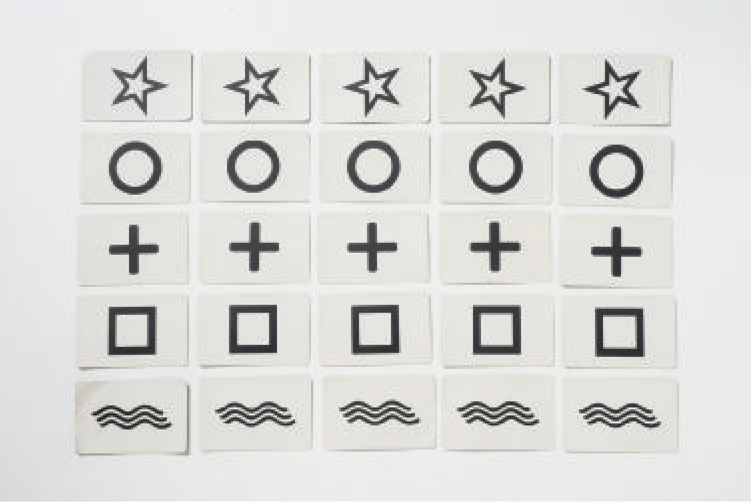
If all four of these characteristics are present, the binomial model applies and you can easily find the probabilities by looking at the binomial probability distribution.

KEY POINT The binomial model provides probabilities for random experiments in which you are counting the number of successes that occur. Four characteristics must be present:

1. Fixed number of trials: n
2. The only two outcomes are success and failure.
3. The probability of success, p, is the same at each trial.
4. The trials are independent.

EXAMPLE 9 Extrasensory Perception (Mind Reading)

Zener cards are special cards used to test whether people can read minds (telepathy). Each card in a Zener deck has one of five special designs: a star, a circle, a plus sign, a square, or three wavy lines (Figure 6.28). In an experiment, one person, the "sender," selects a card at random, looks at it, and thinks about the symbol on the card. Another person, the "receiver," cannot see the card (and in some studies cannot even see the sender) and guesses which of the symbols was chosen. A researcher records whether the guess was correct. The card is then placed back in the deck, the deck is shuffled, and another card is drawn. Suppose this happens 10 times (10 guesses are made). The receiver gets 3 guesses correct, and the researcher wants to know the probability of this happening if the receiver is simply guessing.



◀ **FIGURE 6.28** Zener cards (ESP cards) show one of five shapes. A deck has equal numbers of each shape.

QUESTION Explain why this is a binomial experiment.

SOLUTION First, we note that we are counting something: the number of successful guesses. We need to check that the experiment meets the four characteristics of a binomial model. (1) The experiment consists of a fixed number of trials: $n = 10$. (2) The outcome of each trial is success or failure: The receiver either gets the right answer or does not. (3) The probability of a success at a trial is $p = 1/5 = 0.20$, because there are 5 cards and so the receiver has a 1-in-5 chance of getting it correct, if we assume the receiver is guessing. (4) As long as the cards are put back in the deck and reshuffled (thoroughly), the probability of a success is the same for each trial, and each trial is independent.

All four characteristics are satisfied, so the experiment fits the binomial model.

TRY THIS! Exercise 6.57

EXAMPLE 10 Why Are They Not Binomial?

The following four experiments are almost, but not quite, binomial experiments.

a. Record the number of different eye colors in a group of 50 randomly selected people.

b. A married couple decide to have children until a girl is born, but to stop at five children if they do not have any girls. How many children will the couple have?

c. Suppose the probability that a flight will arrive on time (within 15 minutes of the scheduled arrival time) at O'Hare Airport in Chicago is 85%. How many flights arrive on time out of 300 flights scheduled to land on a day in January?

d. A student guesses on every question of a test that has 10 multiple-choice questions and 10 true/false questions. Record the number of questions the student gets right.

QUESTION For each situation, explain which of the four characteristics is not met.

SOLUTIONS

a. This is not a binomial experiment because there are more than two eye colors, so more than two outcomes may occur at each trial. However, if we reduced the eye colors to two categories by, say, recording whether the eye color was brown or not brown, then this would be a binomial experiment.

b. This is not a binomial experiment because the number of trials is not fixed before the children are born. The number of "trials" depends on when (or whether) the first girl is born. The number of trials varies depending on what happens—the word *until* tells you that.

c. This experiment is not binomial because the flights are not independent. If the weather is bad, the chance of arriving on time for all flights is lower. Therefore, if one flight arrives late, then another flight is more likely to arrive late.

d. This is not a binomial experiment because the probability of success on each trial is not constant; the probability of success is lower on multiple-choice questions than on true/false questions. Therefore, criterion 3 is not met. However, if the test were subdivided into two sections (multiple-choice and true/false), then each separate section could be called a binomial experiment if we assumed the student was guessing.

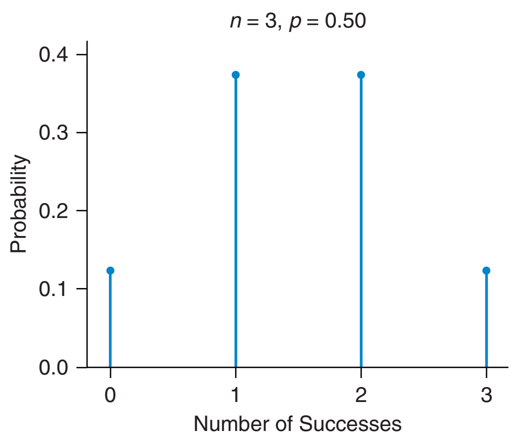
TRY THIS! Exercise 6.59

Visualizing the Binomial Distribution

All binomial models have the four characteristics listed above, but the list gives us flexibility in n and p. For example, if we had flipped the coin 6 times instead of 10, it would still be a binomial experiment. Also, if the probability of a success were 0.6 instead of 0.5, we would still have a binomial experiment. For different values of n and p, we have different binomial experiments, and the binomial distribution looks different in each case.

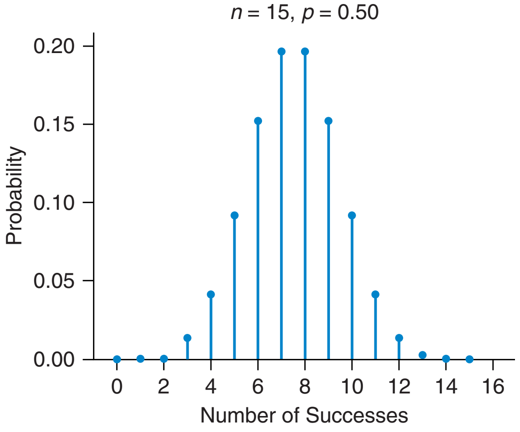
Figure 6.29 shows that the binomial distribution for $n = 3$ and $p = 0.5$ is symmetric. We can read from the graph that the probability of getting exactly 2 successes

▶ **FIGURE 6.29** Binomial distribution with $n = 3$, $p = 0.50$.



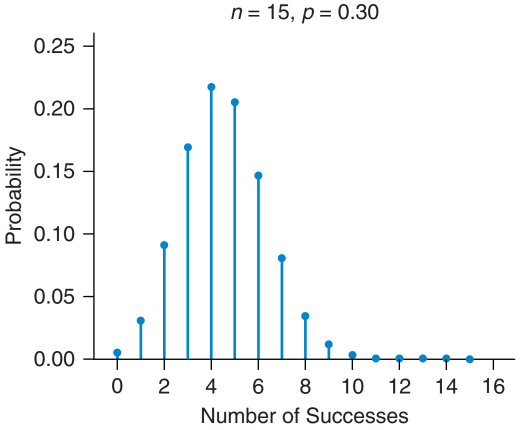
(2 heads in 3 flips of a coin) is almost 0.40, and the probability of getting no successes is the same as the probability of getting all successes.

If n is bigger but p remains fixed at 0.50, the distribution is still symmetric because the chance of a success is the same as the chance of a failure, as shown in Figure 6.30.



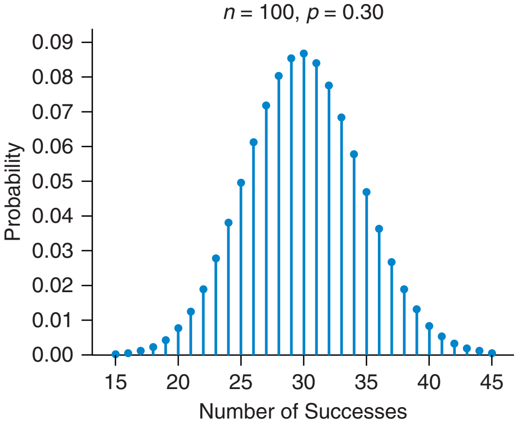
◀ **FIGURE 6.30** Binomial distribution with $n = 15$, $p = 0.50$.

If the probability of success is not 50%, the distribution might not be symmetric. Figure 6.31 shows the distribution for $p = 0.3$, which means we're less likely to get a large number of successes than a smaller number, so the probability "spikes" are taller for smaller numbers of successes. The plot is now right-skewed.



◀ **FIGURE 6.31** Binomial distribution with $n = 15$, $p = 0.30$.

However, even if the distribution is not symmetric, if we increase the number of trials, it becomes symmetric. The shape of the distribution depends on both n and p. If we keep $p = 0.3$ but increase n to 100, we get a more symmetric shape, as shown in Figure 6.32.



◀ **FIGURE 6.32** Binomial distribution with $n = 100$, $p = 0.30$. Note that we show x only for values between 15 and 45. The shape is symmetric, even though p is not 0.50.

Binomial distributions have the interesting property that if the number of trials is large enough, the distributions are symmetric.

KEY POINT The shape of a binomial distribution depends on both n and p. Binomial distributions are symmetric when $p = 0.5$, but they are also symmetric when n is large, even if p is close to 0 or 1.

Finding Binomial Probabilities

Because the binomial distribution depends only on the values of n and p, once you have identified an experiment as binomial and have identified the values of n and p, you can find any probability you wish. We use the notation $b(n, p, x)$ to represent the **binomial probability** of getting x successes in a binomial experiment with n trials and probability of success p. For example, imagine tossing a coin 10 times. If each side is equally likely, what is the probability of getting 4 heads? We represent this with $b(10, 0.50, 4)$.

The easiest and most accurate way to find binomial probabilities is to use technology. Statistical calculators and software have the binomial distribution built in and can easily calculate probabilities for you.

Tech

EXAMPLE 11 Stolen Bicycles

According to the website nationalbikeregistry.com, at the campus of UC Berkeley, only 3% of stolen bicycles are returned to owners.

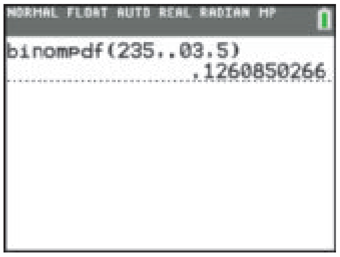
QUESTIONS Accepting for the moment that the four characteristics of a binomial experiment are satisfied, write the notation for the probability that exactly 5 bicycles will be returned if 235 are stolen in the course of a year. What is this probability?

SOLUTION The number of trials is 235, the probability of success is 0.03, and the number of successes is 5. Therefore, we can write the binomial probability as $b(235, 0.03, 5)$.

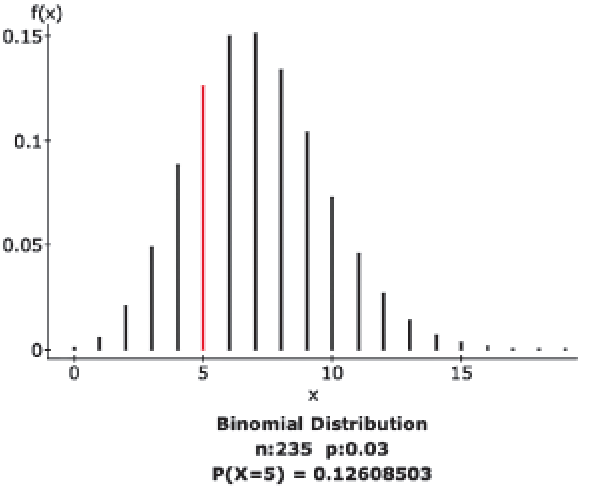
Figure 6.33 shows the command and output for finding this probability using a TI-84 (part a) and using StatCrunch (part b). The command binompdf stands for "binomial probability distribution function." The results show us that the probability that exactly 5 bicycles will be returned, assuming that this is in fact a binomial experiment, is about 12.6%.

▶ **FIGURE 6.33** Calculations for binomial trial using **(a)** TI-84 and **(b)** StatCrunch.

(a)



(b)



TRY THIS! Exercise 6.63

Although we applied the binomial model to the stolen bicycles example, we had to make the assumption that the trials are independent. A "trial" in this context consists of a bicycle being stolen, and a "success" occurs if the bike is returned. If one person stole several bikes on a single day from a single location, this assumption of independent trials would be wrong, because if police found one of the bikes, they would have a good chance of finding the others that the thief took. Sometimes we have no choice but to make certain assumptions to complete a problem, but we need to be sure to check our assumptions when we can, and we must be prepared to change our conclusion if our assumptions were wrong.

Another approach to finding binomial probabilities is to use a table. Published tables are available that list binomial probabilities for a variety of combinations of values for n and p. One such table is provided in Appendix A. This table lists binomial probabilities for values of n between 2 and 15 and for several different values of p.

EXAMPLE 12 Recidivism in Texas

The three-year recidivism rate of parolees in Texas is 30% (www.lbb.state.tx.us). In other words, 30% of released prisoners return to prison within three years of their release. Suppose a prison in Texas released 15 prisoners.

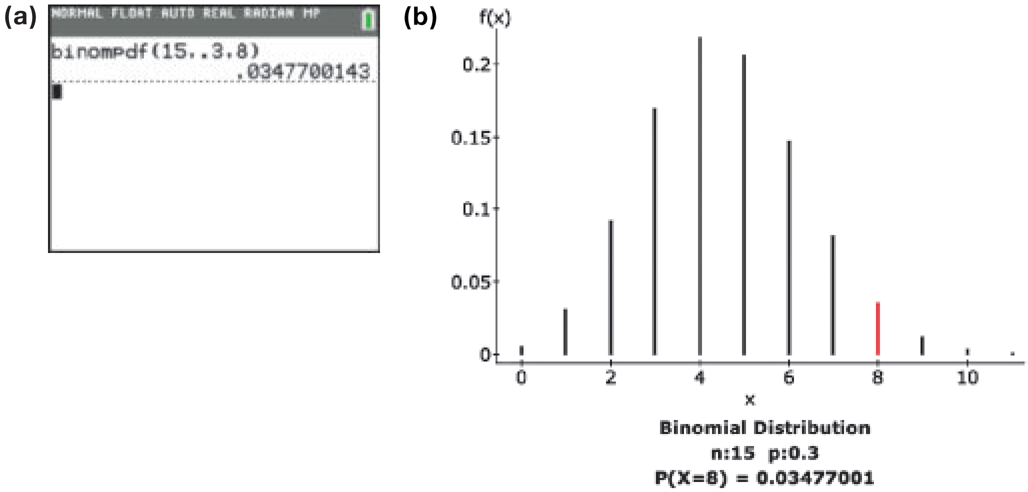
QUESTION Assuming that whether one prisoner returns to prison is independent of whether any of the others returns, use Table 6.8, which shows binomial probabilities for $n = 15$ and for various values of p, to find the probability that exactly 8 out of 15 will end up back in prison within three years.

SOLUTION Substituting the numbers, you can see that we are looking for $b(15, 0.30, 8)$. Referring to Table 6.8, you can see—by looking in the table in the row for $x = 8$, and the column for $p = 0.30$—that the probability that exactly 8 parolees will be back in prison within three years is 0.035, or about a 3.5% chance.

x	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9
6	.002	.043	.092	.147	.207	.153	.061	.012	.003	.001	.000
7	.000	.014	.039	.081	.177	.196	.118	.035	.013	.003	.000
8	.000	.003	.013	.035	.118	.196	.177	.081	.039	.014	.000
9	.000	.001	.003	.012	.061	.153	.207	.147	.092	.043	.002
10	.000	.000	.001	.003	.024	.092	.186	.206	.165	.103	.010
11	.000	.000	.000	.001	.007	.042	.127	.219	.225	.188	.043
12	.000	.000	.000	.000	.002	.014	.063	.170	.225	.250	.129
13	.000	.000	.000	.000	.000	.003	.022	.092	.156	.231	.267
14	.000	.000	.000	.000	.000	.000	.005	.031	.067	.132	.343
15	.000	.000	.000	.000	.000	.000	.000	.005	.013	.035	.206

▲ **TABLE 6.8** Binomial probabilities with a sample of 15 and x-values of 6 or higher.

Using a TI-84 or using StatCrunch, as shown in Figure 6.34 on the next page, we can see another way to get the same answer.



▶ **FIGURE 6.34** Output for calculating $b(15, 0.3, 8)$ using **(a)** TI-84 and **(b)** StatCrunch.

TRY THIS! Exercise 6.65

Finding (Slightly) More Complex Probabilities

EXAMPLE 13 ESP with 10 Trials

For a test of psychic abilities, researchers have asked the sender to draw 10 cards at random from a large deck of Zener cards (see Example 9). Assume that the cards are replaced in the deck after each use and the deck is shuffled. Recall that this deck contains equal numbers of 5 unique shapes. The receiver guesses which card the sender has drawn.

QUESTIONS

a. What is the probability of getting *exactly 5* correct answers (out of 10 trials) if the receiver is simply guessing (and has no psychic ability)?

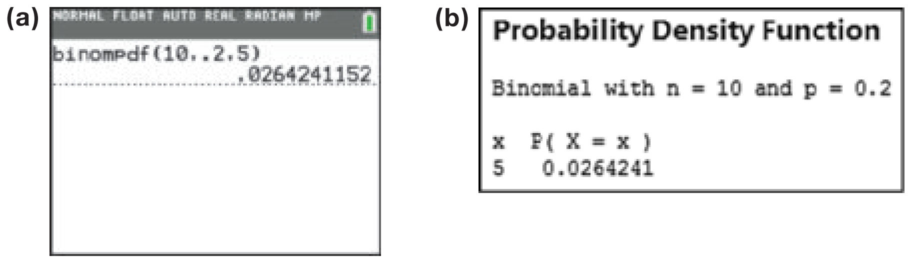
b. What is the probability that the receiver will get *5 or more* of the cards correct out of 10 trials?

c. What is the probability of getting *fewer than 5* correct in 10 trials with the ESP cards?

SOLUTIONS

a. In Example 9 we identified this as a binomial experiment. With that done, we must now identify n and p. The number of trials is 10, so $n = 10$. If the receiver is guessing, then the probability of a correct answer is $p = 1/5 = 0.20$. Therefore, we wish to find $b(10, 0.2, 5)$.

Figure 6.35a gives the TI-84 output, where you can see that the probability of getting 5 right out of 10 is only about 0.0264. Figure 6.35b shows Minitab output for the same question.

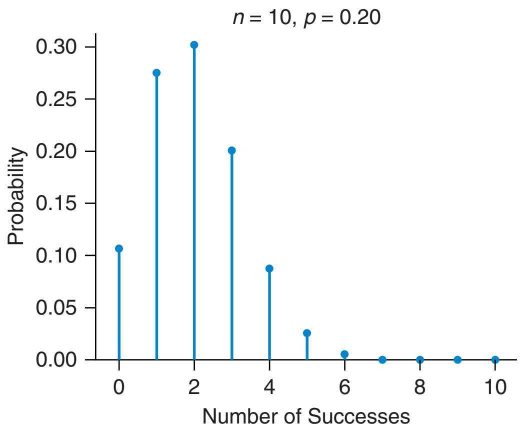


▶ **FIGURE 6.35** Technology output for $b(10, 0.2, 5)$. **(a)** Output from a TI-84. **(b)** Output from Minitab.

Figure 6.36a shows a graph of the pdf. The probability $b(10, 0.2, 5)$ is quite small. The graph shows that it is unusual to get exactly 5 correct when the receiver is guessing.

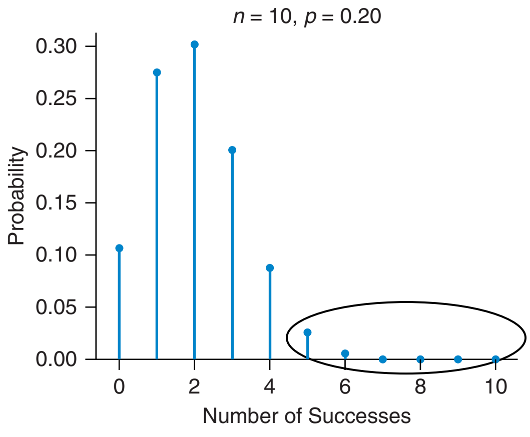
b. The phrase *5 or more* means we need the probability that the receiver gets 5 correct **or** 6 correct **or** 7 **or** 8 **or** 9 **or** 10. The outcomes 5 correct, 6 correct, and so on are mutually exclusive, because if you get exactly 5 correct, you cannot possibly also get exactly 6 correct. Therefore, we can find the probability of 5 or more correct by adding the individual probabilities together:

$$
\begin{aligned}
&b(10, 0.2, 5) + b(10, 0.2, 6) + b(10, 0.2, 7) + b(10, 0.2, 8) + b(10, 0.2, 9) + b(10, 0.2, 10) \\
&= 0.026 + 0.006 + 0.001 + 0.000 + 0.000 + 0.000 \\
&= 0.033
\end{aligned}
$$



◀ **FIGURE 6.36(a)** The probability distribution for the numbers of successes for 10 trials with the Zener deck, assuming guessing.

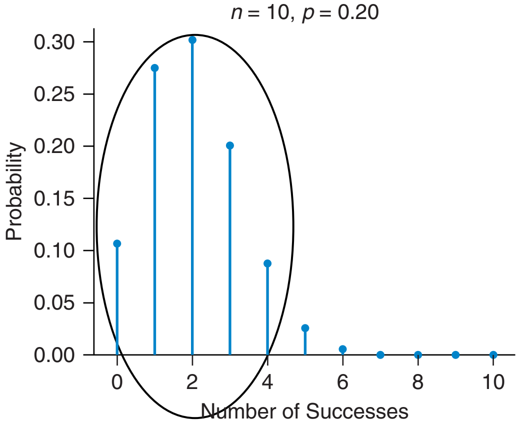
These probabilities are circled in Figure 6.36b.



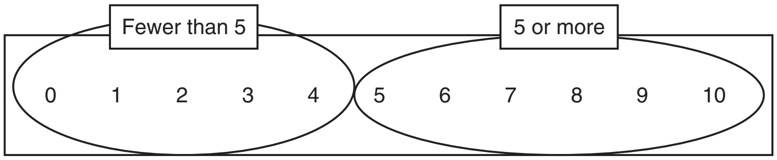
◀ **FIGURE 6.36(b)** Summing the probabilities represented by the circled bars gives us the probability of getting 5 or more correct.

c. The phrase *fewer than 5 correct* means 4, 3, 2, 1, or 0 correct. These probabilities are circled in Figure 6.36c on the next page.

The event that we get fewer than 5 correct is the complement of the event that we get 5 or more correct, as you can see in Figure 6.37 on the next page, which shows all possible numbers of successes with 10 trials. *Fewer than 5* is the same event as *4 or fewer* and is shown in the left oval in the figure.



▲ **FIGURE 6.36(c)** Summing the probabilities represented by the circled bars gives the probability of getting fewer than 5 correct.



▲ **FIGURE 6.37** The possible numbers of successes out of 10 trials with binomial data. Note that *fewer than 5* is the complement of *5 or more.*

Because we know the probability of 5 or more, we can find its complement by subtracting from 1:

$$1 - 0.033 = 0.967$$

CONCLUSIONS

a. The probability of exactly 5 correct is 0.026.

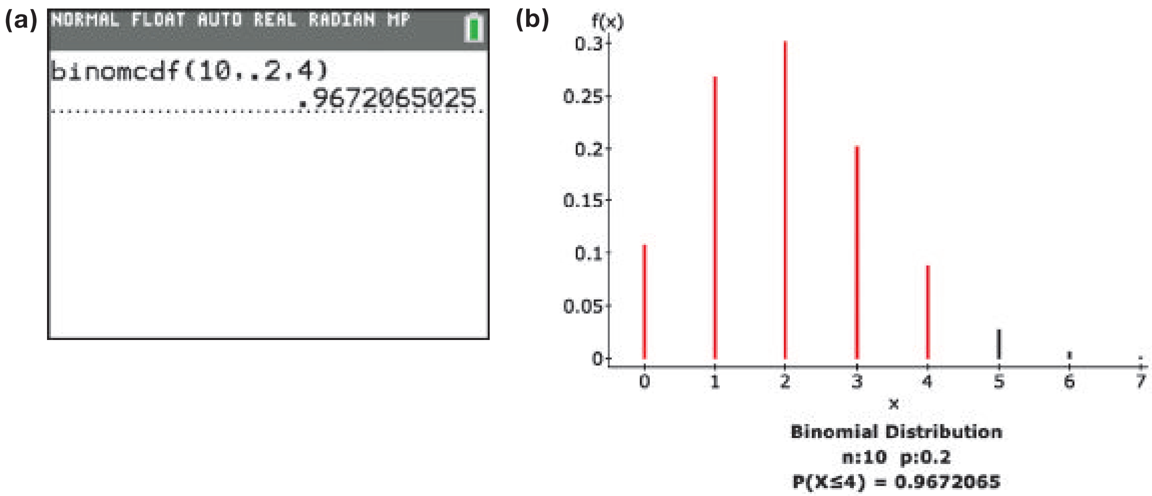
b. The probability of 5 or more correct is 0.033.

c. The probability of fewer than 5 (that is, of 4 or fewer) is 0.967.

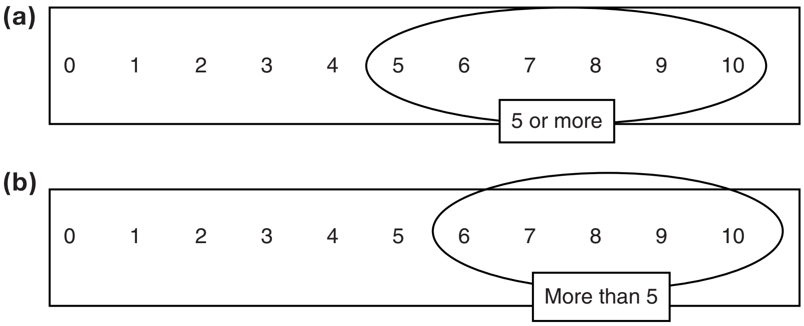
TRY THIS! Exercise 6.73

Most technology also offers you the option of finding binomial probabilities for *x or fewer*. In general, probabilities of *x or fewer* are called **cumulative probabilities**. Figure 6.38a shows the cumulative binomial probabilities provided by the TI-84; notice the “c” in binom**c**df. Figure 6.38b shows the same calculations using StatCrunch.

The probability of getting 5 or more correct is different from the probability of getting more than 5 correct. This very small change in the wording gives very different



▲ **FIGURE 6.38** Output to calculate binomial trials for $n = 10, p = 0.2$, and 4 or fewer successes using **(a)** TI-84 and **(b)** StatCrunch.



◀ **FIGURE 6.39** Interpretation of words for discrete counts with *n* of 10 trials. **(a)** Results for *5 or more*; **(b)** results for *more than 5*. Note that they are different.

results. Figure 6.39 shows that *5 or more* includes the outcomes 5, 6, 7, 8, 9, and 10, whereas *more than 5* does not include the outcome 5.

When finding Normal probabilities, we did not have to worry about such subtleties of language, because for a continuous numerical variable, the probability of getting 5 or more is the same as the probability of getting more than 5.

How Binomial Probabilities Are Calculated

To understand how technology (or the table) computes binomial probabilities, consider a Zener card deck. The deck has equal proportions of five different shapes. After each card is guessed, it is returned to the deck, and the deck is shuffled before the next card is drawn. We want to count how many correct guesses a potential psychic will make in four trials. In particular, we want to find the probability that he will get exactly three guesses right in four trials.

The first step is to list all of the possible outcomes after four attempts. You might think that there are only five outcomes: he gets 0 right, gets 1 right, gets 2 right, gets 3 right, and gets 4 right. And so you might think the probability of getting three guesses right is simply 1/5. But in fact, there are several different ways of getting three right. The binomial distribution takes into account each of these different ways for us to find the correct probability.

4 Right	3 Right	2 Right	1 Right	0 Right
RRRR	RRRW	RRWW	RWWW	WWWW
	RRWR	RWRW	WRWW	
	RWRR	RWWR	WWRW	
	WRRR	WWRR	WWWR	
		WRWR		
		WRRW		

Here are all of the outcomes, with R representing a "right" guess and W representing a "wrong" guess. We can see that there are four different ways in which the psychic might correctly guess three cards. The probability of getting any one of these four outcomes, for instance RRRW, is

$$P(RRRW) = 0.2 \times 0.2 \times 0.2 \times 0.8 = 0.0064$$

Because there are four different ways in which we might get this probability, the probability of getting three out of four correct is

$$P(3 \text{ right and 1 wrong, in any order}) = P(RRRW) + P(RRWR) + P(RWRR) + P(WRRR) = 4(0.0064) = 0.0256$$

Looking Back

And
The multiplication rule was Probability Rule 5c in Chapter 5 and applies only to independent events: P(A AND B) = P(A) P(B).

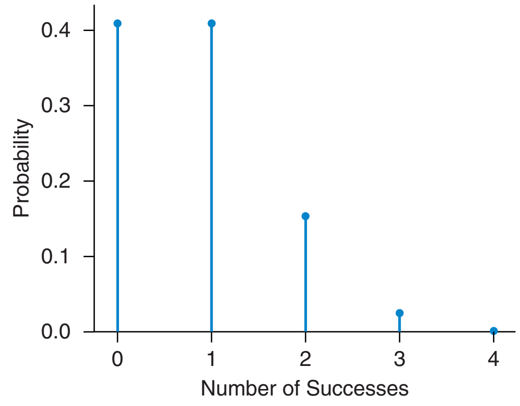
Number Right	Probability
4 right	0.0016
3 right	0.0256
2 right	0.1536
1 right	0.4096
0 right	0.4096

▲ **TABLE 6.9** Summary of the probabilities of all possible numbers of successes in four trials with the Zener cards.

In general, this is how binomial probabilities are calculated. For a given number of successes, we must first count the number of different ways in which that number of successes can occur. For each way, we find the probability of that particular sequence happening. We then add these probabilities together.

Table 6.9 shows the probability distribution. If you add the probabilities in Table 6.9, you will see that they add to 1, as they should, because this list includes all possible outcomes. Figure 6.40 shows a graph of the probability distribution. Note that the distribution is right-skewed because the probability of success is less than 0.50. Also note that the probability of getting four out of four right with the Zener cards is very small!

You might compare these probabilities with those in the binomial table. You will find that they agree if you round off the probabilities in Table 6.9 to three decimal places.

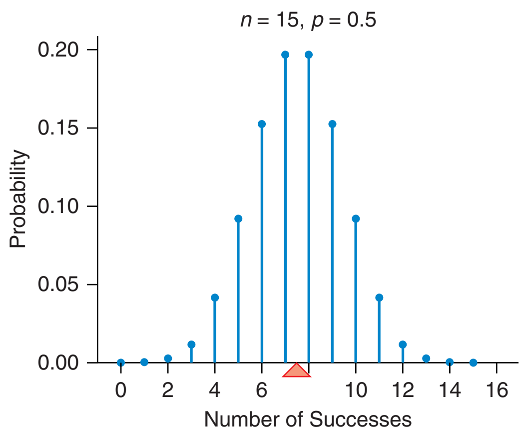


◀ **FIGURE 6.40** Probability distribution using the Zener cards with four trials.

The Shape of the Binomial Distribution: Center and Spread

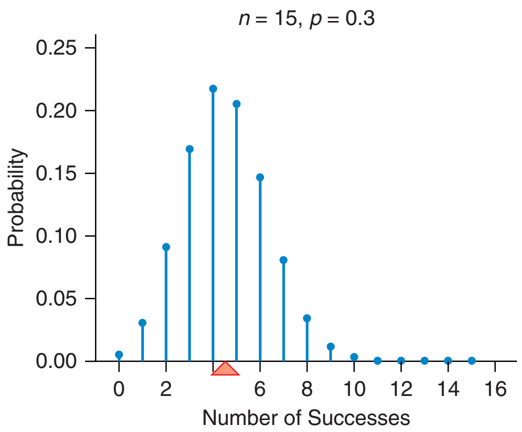
Unlike the Normal distribution, the mean and standard deviation of the binomial distribution can be easily calculated. Their interpretation is the same as with all distributions: The mean tells us where the distribution balances, and the standard deviation tells us how far values are, typically, from the mean.

For example, in Figure 6.41, the binomial distribution is symmetric, so the mean sits right in the center at 7.5 successes.



▶ **FIGURE 6.41** A binomial distribution with $n = 15$, $p = 0.50$. The mean is at 7.5 successes.

If the distribution is right-skewed, the mean will be just to the right of the peak, closer to the right tail, as shown in Figure 6.42. This is a binomial distribution with $n = 15$, $p = 0.3$, and the mean sits at the balancing point: 4.5 successes.



◀ **FIGURE 6.42** A binomial distribution with $n = 15$, $p = 0.3$. The mean is at the balancing point of 4.5 successes.

The mean, μ, of a binomial probability distribution can be found with a simple formula:

$$\textbf{Formula 6.1a: } \mu = np$$

In words, the mean number of successes in a binomial experiment is the number of trials times the probability of a success.

This formula should make intuitive sense. If you toss a coin 100 times, you would expect about half of the tosses to be heads. Thus the "typical" outcome is $(1/2) \times 100 = 50$, or $np = 100 \times 0.5$.

The standard deviation, σ, of a binomial probability distribution, which measures the spread, is less intuitive:

$$\textbf{Formula 6.1b: } \sigma = \sqrt{np(1 - p)}$$

For example, Figure 6.42 shows a binomial distribution with $n = 15$, $p = 0.30$. The mean is $15 \times 0.3 = 4.5$. The standard deviation is $\sqrt{15 \times 0.3 \times 0.7} = \sqrt{3.15} = 1.775$.

KEY POINT For a binomial experiment, the mean is

$$\mu = np$$

For a binomial experiment, the standard deviation is

$$\sigma = \sqrt{np(1 - p)}$$

Interpreting the Mean and Standard Deviation The mean of any probability distribution, including the binomial, is sometimes called the **expected value**. This name gives some intuitive understanding. If you were to actually carry out a binomial experiment, you would *expect* about μ successes. If I flip a coin 100 times, I *expect* about 50 heads. If, in an ESP study, 10 trials are made and the probability of success at each trial is 0.20, we *expect* about 2 successes due to chance $(10 \times 0.2 = 2)$.

Will I get exactly 50 heads? Will the ESP receiver get exactly 2 cards correct? Sometimes, yes. Usually, no. Although we expect μ successes, we usually get μ give or take some amount. That give-or-take amount is what is measured by σ.

In 100 tosses of a fair coin, we expect $\mu = 50$ heads, give or take $\sigma = \sqrt{100 \times 0.5 \times 0.5} = 5$ heads. We expect 50, but we will not be surprised if we get between 45 and 55 heads. In the Zener card experiment with 10 trials, we expect the receiver to guess about 2 cards correctly, but in practice we expect him or her to get 2 give or take 1.3, because

$$\sqrt{10 \times 0.2 \times 0.8)} = \sqrt{1.6} = 1.26$$

SNAPSHOT THE BINOMIAL DISTRIBUTION

WHAT IS IT? ▶ A distribution for some discrete variables.

WHAT DOES IT DO? ▶ Gives probabilities for the number of successes observed in a fixed number of trials in a binomial experiment.

HOW DOES IT DO IT? ▶ If the conditions of a binomial experiment are met, once you identify n (the number of trials), p (the probability of success), and x (the number of successes), it gives the probability.

HOW IS IT USED? ▶ The probabilities are generally provided in the form of a table or a formula, but if you need to calculate them, use a calculator or technology.

EXAMPLE 14 Basketball Free-Throw Shots

During one season, basketball player LeBron James had a free-throw success percentage of 79%. Assume the free-throw shots are independent; that is, success or failure on one shot does not affect the chance of success on another shot.

QUESTION If Lebron James has 600 free throws in an upcoming season, how many would you expect him to sink, give or take how many?

SOLUTION This is a binomial experiment. (Why?) The number of trials is $n = 600$, and the probability of success at each trial is $p = 0.79$. You would expect LeBron James to sink 79% of 600, or 474, free throws. Using Formula 6.1a for the mean yields

$$\mu = np = 600 \times (0.79) = 474$$

The give-or-take amount is measured by the standard deviation as given in Formula 6.1b:

$$\sigma = \sqrt{600 \times 0.79 \times 0.21} = 9.9770$$

You should expect him to hit about 474 free throws, give or take about 10.0.

$$474 + 10.0 = 484.0$$

$$474 - 10.0 = 464.0$$

CONCLUSION You would expect LeBron James to sink between 464 and 484 out of 600 free-throw shots.

TRY THIS! Exercise 6.77

Surveys: An Application of the Binomial Model

Perhaps the most common application of the binomial model is in survey sampling. Imagine a large population of people, say the 100 million or so registered voters in the United States. Some percentage of them, call it p, have a certain characteristic. If we choose ten people at random (with replacement), we can ask what the probability is that all of them, or three of them, or six of them, have that characteristic.

EXAMPLE 15 News Survey

The Pew Research Center says that 23% of people are "news integrators": people who get their news both from traditional media (television, radio, newspapers, magazines) and from the Internet (www.people-press.org). Suppose we take a random sample of 100 people.

QUESTIONS If, as is claimed, 23% of the population are news integrators, then how many people in our sample should we expect to be integrators? Give or take how many? Would you be surprised if 34 people in the sample turned out to be integrators?

SOLUTION Assuming that all four of the characteristics of a binomial distribution are satisfied, then we would expect 23% of our sample of 100 people to be integrators—that is, 23 people. The standard deviation is $\sqrt{100 \times 0.23 \times 0.77} = 4.2$. Thus we should expect 23 people, give or take about 4.2 people, to be integrators. This means that we shouldn't be surprised if we got as many as $23 + 4.2 = 27.2$, or about 27, people. However, 34 people is quite a bit more than 1 standard deviation above what we expect. In fact, it is almost 3 standard deviations away, so it would be a surprisingly large number of people.

TRY THIS! Exercise 6.79

The binomial model works well for surveys when people are selected *with* replacement (which means that once they are selected, they have a chance of being selected again), or when there are very many more people in the population than in the sample. If we are counting the people who have a certain characteristic, then the four characteristics of the binomial model are usually satisfied:

1. A fixed number of people are surveyed. In Example 15, $n = 100$.
2. The outcome is success (integrator) or failure (not an integrator).
3. The probability of a success is the same at each trial: $p = 0.23$.
4. The trials are independent. (This means that if one person is found to be an integrator, no one else in the sample will change their response.)

With surveys, we usually don't report the number of people who have an interesting characteristic; instead, we report the percentage of our sample. We would not report that 23 people in our sample of 100 were integrators; we would report that 23% of our sample were integrators. But the binomial distribution still applies, because we are simply converting our counts to percentages. You will learn more about surveys in Chapter 7.

CASE STUDY REVISITED

McDonald's claims that its ice cream cones weigh 3.18 ounces. However, one of the authors bought five cones and found that all five weighed more than that. Is this surprising?

Suppose we assume that the amount of ice cream dispensed follows a Normal distribution centered on 3.18 ounces. In other words, typically, a cone weighs 3.18 ounces, but sometimes a cone weighs a little more, and sometimes a little less. If this is the case, then, because the Normal distribution is symmetric and centered on its mean, the probability that a cone weighs more than 3.18 ounces is 0.50.

If the probability that a cone weighs more than 3.18 ounces is 0.50, what is the probability that five cones all weigh more? We can think of this as a binomial experiment if we assume the trials are independent. (This seems like a reasonable assumption.) We can then calculate $b(5, 0.5, 5)$ to find that the probability is 0.031.

If the typical McDonald's ice cream cone really weighs 3.18 ounces, then our outcome is fairly surprising: It happens only about 3% of the time. This means it is fairly rare and raises the possibility that this particular McDonald's might deliberately dispense more than 3.18 ounces.

EXPLORING STATISTICS
CLASS ACTIVITY

ESP with Coin Flipping

GOALS	MATERIALS
Apply the binomial model to a real situation.	One coin for each pair of students.

ACTIVITY

Choose a partner. One of you will play the role of the "sender" and the other will be the "receiver." The sender will flip a coin in such a way that the receiver can't see the outcome (heads or tails). The sender will then look at the coin and concentrate, trying to mentally send a thought about whether the coin landed heads or tails. The receiver writes down (quickly) the outcome he or she believes was sent. (Just write the first thing that comes to mind.) The sender should make a tally of the number of right answers the receiver achieved in 10 trials. Now switch roles and try it again. Each of you should be prepared to report how many you got correct in your 10 trials.

BEFORE THE ACTIVITY

1. With 10 trials, how many would you expect to guess correctly if there is no ESP? If there are 20 people trying this, do you expect all of them to have the same results?
2. Find the standard deviation—the "give-or-take value"—of the number of correct guesses in 10 trials. Assuming that the receiver does not have ESP, what's the smallest number of correct guesses you might expect? What's the largest? How does the standard deviation help you determine this?

AFTER THE ACTIVITY

1. Make a histogram of the class results. Where is the distribution centered? Is this what you expected?
2. Are all of the results within the range of results predicted above?
3. Are any of the results unusually good? Does that show that the person with unusually good results has ESP? Why or why not? Explain.

CHAPTER REVIEW

KEY TERMS

probability model, *274*
probability distribution, *274*
probability distribution function (pdf), *274*
discrete outcomes, *274*
continuous outcomes, *274*
probability density curve, *277*
Normal model: notation, $N(\mu, \sigma)$, *279*
Normal curve, *279*
Normal distribution, *279*
mean of a probability distribution, μ, *280*
standard deviation of a probability distribution, σ, *280*
standard Normal model notation, $N(0, 1)$, *285*
percentile, *288*
binomial probability model Notation, $b(n, p, x)$, *292*
n is the number of trials
p is the probability of success on one trial
x is the number of successes
cumulative probabilities, *300*
expected value, *303*

LEARNING OBJECTIVES

After reading this chapter and doing the assigned homework problems, you should

- Be able to distinguish between discrete and continuous-valued variables.
- Know when a Normal model is appropriate and be able to apply the model to find probabilities.
- Know when the binomial model is appropriate and be able to apply the model to find probabilities.

SUMMARY

Probability models try to capture the essential features of real-world experiments and phenomena that we want to study. In this chapter, we focused on two very useful models: the Normal model and the binomial model.

The Normal model is an example of a model of probabilities for continuous numerical variables. The Normal probability model is also called the Normal distribution and the Gaussian curve. It can be a useful model when a histogram of data collected for a variable is unimodal and symmetric. Probabilities are found by finding the area under the appropriate region of the Normal curve. These areas are best calculated using technology. If technology is not available, you can also convert measures to standard units and then use the table of areas for the standard Normal distribution, provided in Appendix A.

The binomial model is an example of a model of probabilities for discrete numerical outcomes. The binomial model applies to binomial experiments, which occur when we are interested in counting the number of times some event happens. These four characteristics must be met for the binomial model to be applied:

1. There must be a fixed number of trials, n.
2. Each trial has exactly two possible outcomes.
3. Each of the trials must have the same probability of "success." This probability is represented by the letter p.
4. The trials must be independent of one another.

You can find binomial probabilities with technology or sometimes with a table, such as the one in Appendix A.

It is important to distinguish between continuous and discrete numerical variables, because if the variable has discrete numerical outcomes, then the probability of getting, say, *5 or more* (5 OR 6 OR 7 OR . . .) is different from the probability of getting *more than 5* (6 OR 7 OR . . .). This is not the case for a continuous numerical variable.

Formulas

For converting to standardized scores: $z = \dfrac{x - \mu}{\sigma}$

x is a measurement.
μ is the mean of the probability distribution.
σ is the standard deviation of the probability distribution.

For the mean of a binomial model:

Formula 6.1a: $\mu = np$

μ is the binomial mean.
n is the number of trials.
p is the probability of success of one trial.

For the standard deviation of a binomial model:

Formula 6.1b: $\sigma = \sqrt{np(1 - p)}$

σ is the binomial standard deviation.
n is the number of trials.
p is the probability of success of one trial.

SOURCES

Men's cholesterol levels, blood pressures, and heights throughout this chapter: NHANES (www.cdc.gov/nchs/nhanes).

Women's heights and weights throughout this chapter: NHANES (www.cdc.gov/nchs/nhanes).

SECTION EXERCISES

SECTION 6.1

6.1–6.4 Directions Determine whether each of the following variables would best be modeled as continuous or discrete.

TRY **6.1 (Example 1)**

a. Length and breadth of a classroom

b. Number of students present in a class

6.2 a. Number of expeditions to Mount Everest

b. Height of a mountain in the Himalayas (in meters)

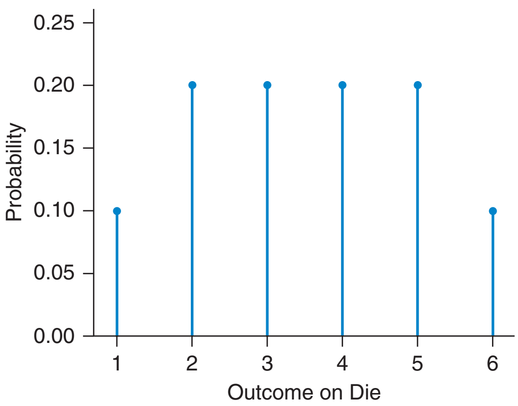
6.3 a. Weight of a person (kilograms)

b. Weight of a person (pounds)

6.4 a. Number of months in a calendar year

b. The time taken by earth to complete one revolution around the sun

TRY **6.5 Loaded Die (Example 2)** A magician has shaved an edge off one side of a six-sided die, and as a result, the die is no longer "fair." The figure shows a graph of the probability density function (pdf). Show the pdf in table format by listing all six possible outcomes and their probabilities.



6.6 Fair Die Toss a fair six-sided die. The probability density function (pdf) in table form is given. Make a graph of the pdf for the die.

Number of Spots	1	2	3	4	5	6
Probability	1/6	1/6	1/6	1/6	1/6	1/6

* **6.7 Distribution of Two Thumbtacks** When a certain type of thumbtack is flipped, the probability of its landing tip up (U) is 0.60 and the probability of its landing tip down (D) is 0.40.

Now suppose we flip two such thumbtacks: one red, one blue. Make a list of all the possible arrangements using U for up and D for down, listing the red one first; include both UD and DU. Find the probabilities of each possible outcome, and record the result in table form. Be sure the total of all the probabilities is 1.

6.8 Distribution of Two Dices When two dices are thrown, the probability of getting a multiple of 3 (M) is 0.33 and the probability of not getting a multiple of 3 (N) is 0.67.

Make a list of all possible arrangements for getting a multiple of 3, using M for multiples and N for numbers that are not. Find the probabilities of each arrangement, and record the results in table form. Be sure the total of all the probabilities is 1.

6.9 Two Thumbtacks

a. From your answers in Exercise 6.7, find the probability of getting 0 ups, 1 up, or 2 ups when flipping two thumbtacks, and report the distribution in a table.

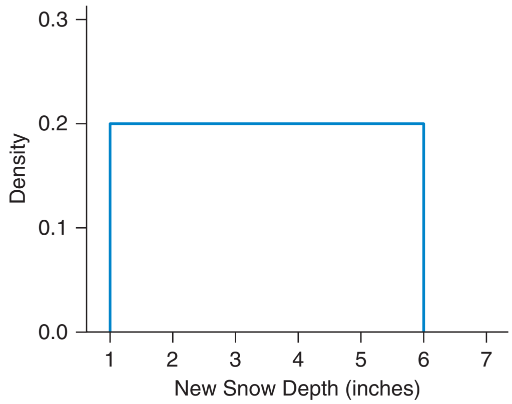
b. Make a probability distribution graph of this.

6.10 Two Dices

a. From your answers in Exercise 6.8, find the probability of getting no multiples, 1 multiple, or 2 multiples of 3, and report the distribution in a table.

b. Make a probability distribution graph of this.

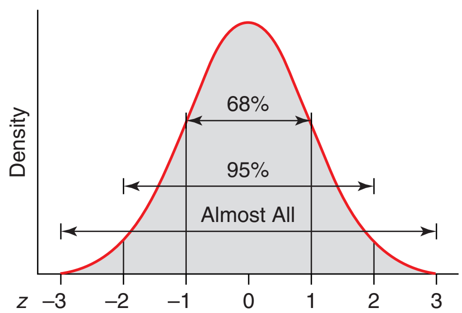
TRY **6.11 Snow Depth (Example 3)** Eric wants to go skiing tomorrow, but only if there are 3 inches or more of new snow. According to the weather report, any amount of new snow between 1 inch and 6 inches is equally likely. The probability density curve for tomorrow's new snow depth is shown. Find the probability that the new snow depth will be 3 inches or more tomorrow. Copy the graph, shade the appropriate area, and calculate its numerical value to find the probability. The total area is 1.



6.12 Snow Depth Refer to Exercise 6.11. What is the probability that the amount of new snow will be between 2 and 4 inches? Copy the graph from Exercise 6.11, shade the appropriate area, and report the numerical value of the probability.

SECTION 6.2

6.13 Applying the Empirical Rule with *z*-Scores The Empirical Rule applies rough approximations to probabilities for any unimodal, symmetric distribution. But for the Normal distribution we can be more precise. Use the figure and the fact that the Normal curve is symmetric to answer the questions. Do not use a Normal table or technology.



According to the Empirical Rule:

a. Roughly what percentage of z-scores are between -2 and 2?
   i. almost all ii. 95% iii. 68% iv. 50%

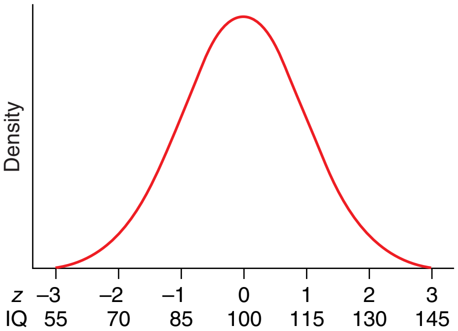
b. Roughly what percentage of z-scores are between -3 and 3?
   i. almost all ii. 95% iii. 68% iv. 50%

c. Roughly what percentage of z-scores are between -1 and 1.
   i. almost all ii. 95% iii. 68% iv. 50%

d. Roughly what percentage of z-scores are greater than 0?
   i. almost all ii. 95% iii. 68% iv. 50%

e. Roughly what percentage of z-scores are between 1 and 2?
   i. almost all ii. 13.5% iii. 50% iv. 2%

6.14 IQs Wechsler IQs are approximately Normally distributed with a mean of 100 and a standard deviation of 15. Use the probabilities shown in the figure in Exercise 6.13 to answer the following questions. Do *not* use the Normal table or technology. You may want to label the figure with Empirical Rule probabilities to help you think about this question.



a. Roughly what percentage of people have IQs more than 100?
   i. almost all ii. 95% iii. 68% iv. 50%

b. Roughly what percentage of people have IQs between 100 and 115?
   i. 34% ii. 17% iii. 2.5% iv. 50%

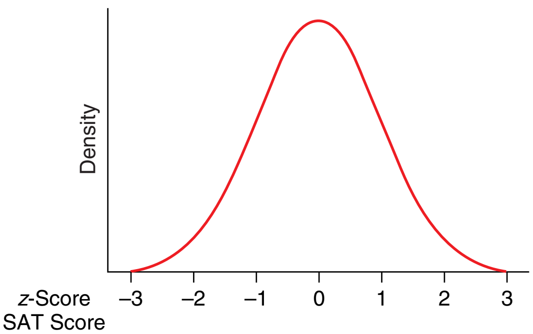
c. Roughly what percentage of people have IQs below 55?
   i. almost all ii. 50% iii. 34% iv. about 0%

d. Roughly what percentage of people have IQs between 70 and 130?
   i. almost all ii. 95% iii. 68% iv. 50%

e. Roughly what percentage of people have IQs above 130?
   i. 34% ii. 17% iii. 2.5% iv. 50%

f. Roughly what percentage people have IQs above 145?
   i. almost all ii. 50% iii. 34% iv. about 0%

6.15 SAT Scores Quantitative SAT scores are approximately Normally distributed with a mean of 500 and a standard deviation of 100. On the horizontal axis of the graph, indicate the SAT scores that correspond with the provided z-scores. (See the labeling in Exercise 6.14.) Answer the questions using *only* your knowledge of the Empirical Rule and symmetry.



a. Roughly what percentage of students earn quantitative SAT scores greater than 500?
   i. almost all ii. 75% iii. 50% iv. 25% v. about 0%

b. Roughly what percentage of students earn quantitative SAT scores between 400 and 600?
   i. almost all ii. 95% iii. 68% iv. 34% v. about 0%

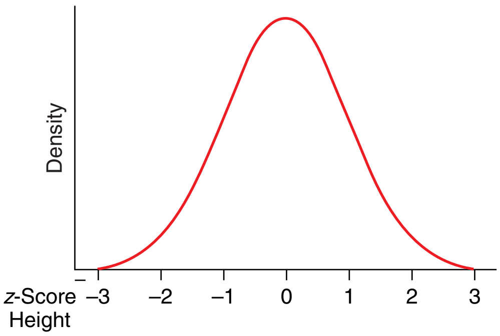
c. Roughly what percentage of students earn quantitative SAT scores greater than 800?
   i. almost all ii. 95% iii. 68% iv. 34% v. about 0%

d. Roughly what percentage of students earn quantitative SAT scores less than 200?
   i. almost all ii. 95% iii. 68% iv. 34% v. about 0%

e. Roughly what percentage of students earn quantitative SAT scores between 300 and 700?
   i. almost all ii. 95% iii. 68% iv. 34% v. 2.5%

f. Roughly what percentage of students earn quantitative SAT scores between 700 and 800?
   i. almost all ii. 95% iii. 68% iv. 34% v. 2.5%

6.16 Women's Heights Assume that college women's heights are approximately Normally distributed with a mean of 65 inches and a standard deviation of 2.5 inches. On the horizontal axis of the graph, indicate the heights that correspond to the z-scores provided. (See the labeling in Exercise 6.14.) Use only the Empirical Rule to choose your answers. Sixty inches is 5 feet, and 72 inches is 6 feet.



a. Roughly what percentage of women's heights are greater than 72.5 inches?
   i. almost all ii. 75% iii. 50% iv. 25% v. about 0%

b. Roughly what percentage of women's heights are between 60 and 70 inches?
   i. almost all ii. 95% iii. 68% iv. 34% v. about 0%

c. Roughly what percentage of women's heights are between 65 and 67.5 inches?
   i. almost all ii. 95% iii. 68% iv. 34% v. about 0%

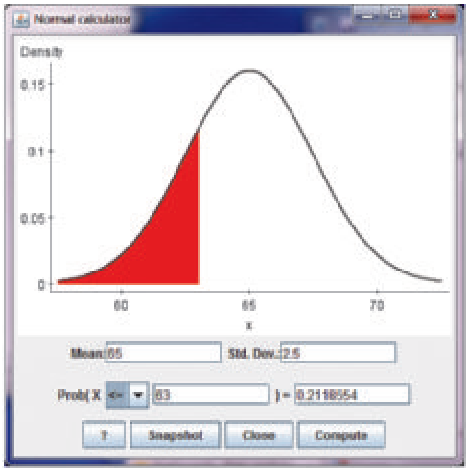
d. Roughly what percentage of women's heights are between 62.5 and 67.5 inches?
   i. almost all ii. 95% iii. 68% iv. 34% v. about 0%

e. Roughly what percentage of women's heights are less than 57.5 inches?
   i. almost all ii. 95% iii. 68% iv. 34% v. about 0%

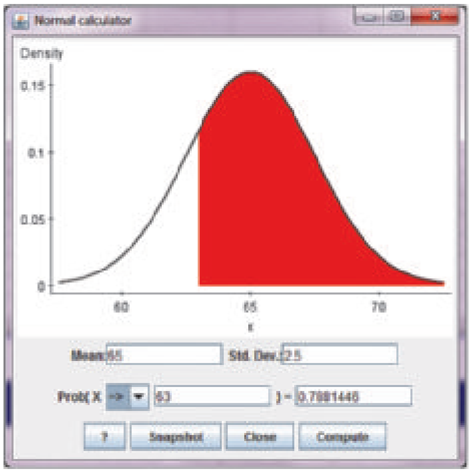
f. Roughly what percentage of women's heights are between 65 and 70 inches?
   i. almost all ii. 95% iii. 47.5% iv. 34% v. 2.5%

TRY **6.17 Women's Heights (Example 4)** College women have a mean height of 65 inches and a standard deviation of 2.5 inches. The distribution of heights for this group is Normal. Choose the correct StatCrunch output for finding the percentage of college women with heights of less than 63 inches, and report the correct percentage.

(A)

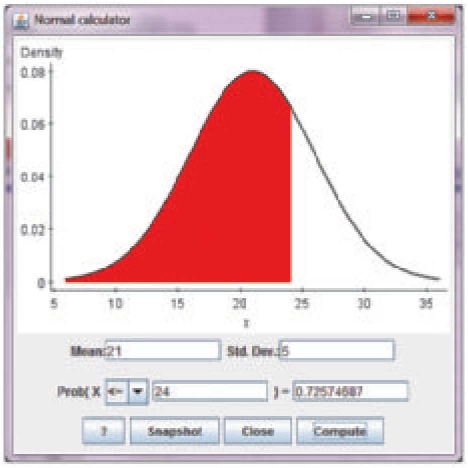


(B)



6.18 ACT Scores ACT scores are approximately Normally distributed with a mean of 21 and a standard deviation of 5, as shown in the figure. (ACT scores are test scores that some colleges use for determining admission.)

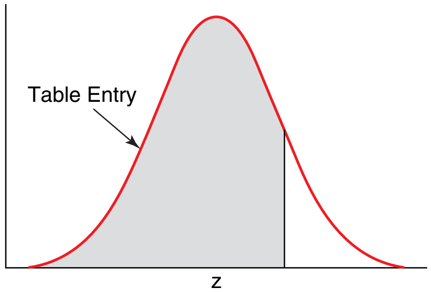
What is the probability that a randomly selected person scores *24 or more*?



6.19 Standard Normal Use the table or technology to answer each question. Include an appropriately labeled sketch of the Normal curve for each part. Shade the appropriate region.

a. Find the area in a standard Normal curve to the left of 1.02 by using the Normal table. (See the excerpt provided.) Note the shaded curve.

b. Find the area in a standard Normal curve to the right of 1.02. Remember that the total area under the curve is 1.



Format of the Normal Table: The area given is the area to the *left* of (less than) the given *z*-score.

z	.00	.01	.02	.03	.04
0.9	.8159	.8186	.8212	.8238	.8264
1.0	.8413	.8438	**.8461**	.8485	.8508
1.1	.8643	.8665	.8686	.8708	.8729

6.20 Standard Normal Use a table or technology to answer each question. Include an appropriately labeled sketch of the Normal curve for each part. Shade the appropriate region.

a. Find the area to the left of a z-score of -0.50.

b. Find the area to the right of a z-score of -0.50.

6.21 Standard Normal Use a table or technology to answer each question. Include an appropriately labeled sketch of the Normal curve for each part. Shade the appropriate region.

a. Find the probability that a z-score will be 1.76 or less.

b. Find the probability that a z-score will be 1.76 or more.

c. Find the probability that a z-score will be between -1.3 and -1.03.

6.22 Standard Normal Use a table or technology to answer each question. Include an appropriately labeled sketch of the Normal curve for each part. Shade the appropriate region.

a. Find the probability that a z-score will be -1.00 or less.

b. Find the probability that a z-score will be more than -1.00.

c. Find the probability that a z-score will be between 0.90 and 1.80.

6.23 Extreme Positive *z*-Scores For each question, find the area to the right of the given z-score in a standard Normal distribution. In this question, round your answers to the nearest 0.000. Include an appropriately labeled sketch of the $N(0, 1)$ curve.

a. $z = 4.00$

b. $z = 10.00$ (*Hint*: Should this tail proportion be larger or smaller than the answer to part a? Draw a picture and think about it.)

c. $z = 50.00$

d. If you had the *exact* probability for these tail proportions, which would be the largest and which would be the smallest?

e. Which is equal to the area in part b: the area below (to the left of) $z = -10.00$ or the area above (to the right of) $z = -10.00$?

6.24 Extreme Negative *z*-Scores For each question, find the area to the right of the given z-score in a standard Normal distribution. In *this* question, round your answers to the nearest 0.000. Include an appropriately labeled sketch of the $N(0, 1)$ curve.

a. $z = -4.00$

b. $z = -8.00$

c. $z = -30.00$

d. If you had the *exact* probability for these right proportions, which would be the largest and which would be the smallest?

e. Which is equal to the area in part b: the area below (to the left of) $z = 8.00$ or the area above (to the right of) $z = 8.00$?

TRY g **6.25 Females' SAT Scores (Example 5)** According to data from the College Board, the mean quantitative SAT score for female college-bound high school seniors is 500. Assume that SAT scores are approximately Normally distributed with a population standard deviation of 100. What percentage of female college-bound students had scores above 675? Please include a well-labeled Normal curve as part of your answer. *See page 317 for guidance*.

6.26 Males' SAT Scores According to data from the College Board, the mean quantitative SAT score for male college-bound high school seniors is 530. Assume that SAT scores are approximately Normally distributed with a population standard deviation of 100. If a male college-bound high school senior is selected at random, what is the probability that he will score higher than 675?

6.27 Stanford–Binet IQs Stanford–Binet IQ scores for children are approximately Normally distributed and have $\mu = 100$ and $\sigma = 15$. What is the probability that a randomly selected child will have an IQ below 115?

6.28 Stanford–Binet IQs Stanford–Binet IQs for children are approximately Normally distributed and have $\mu = 100$ and $\sigma = 15$. What is the probability that a randomly selected child will have an IQ of 115 or above?

TRY **6.29 Birth Length (Example 6)** According to National Vital Statistics, the average length of a newborn baby is 19.5 inches with a standard deviation of 0.9 inch. The distribution of lengths is approximately Normal. Use a table or technology for each question. Include an appropriately labeled and shaded Normal curve for each part. There should be three separate curves.

a. What is the probability that a baby will have a length of 20.4 inches or more?

b. What is the probability that a baby will have a length of 21.4 inches or more?

c. What is the probability that a baby will be between 18 and 21 inches in length?

6.30 White Blood Cells The distribution of white blood cell count per cubic millimeter of whole blood is approximately Normal with mean 7500 and standard deviation 1750 for healthy patients. Include an appropriately labeled and shaded Normal curve for each part. There should be three separate curves.

a. What is the probability that a randomly selected person will have a white blood cell count of between 7000 and 10,000?

b. What is the probability that a randomly selected person will have a white blood cell count of between 5000 and 12,000?

c. What is the probability that a randomly selected person will have a white blood cell count of more than 10,000?

6.31 Protein Intake: Men The Dietary Reference Intake of proteins is different for men and women. For both, the distribution is approximately Normal. For men, the middle 95% range from 54 to 58 grams per day, and for women, the middle 95% have protein intake recommendation between 44.8 and 47.4 grams per day.

a. What is the mean for the men? Explain your reasoning.

* b. Find the standard deviation for the men. Explain your reasoning.

* **6.32 Protein Intake: Women** Answer the previous question for the women.

6.33 Bilingual Minors A survey shows that in one year, the average number of bilingual minors in a school was 384. Assume that the standard deviation is 98 and the number of bilingual minors is Normally distributed. Include an appropriately labeled and shaded Normal curve for each part.

a. What percentage of schools have more than 384 bilingual minors?

b. What percentage of schools have 350 or less bilingual minors?

6.34 Bilingual Employees A survey shows that in one year, the average number of bilingual employees in an office was 12. Assume that the standard deviation is 3 and the number of bilingual employees is Normally distributed. Include an appropriately labeled and shaded Normal curve for each part.

a. What percentage of offices have more than 12 bilingual employees?

b. What percentage of offices have 10 or more bilingual employees?

6.35 Bilingual Army Officers A survey shows that in one year, the average number of bilingual officers in an army battalion was 196. Assume that the standard deviation is 22 and the number of bilingual officers in an army battalion is Normally distributed. Include an appropriately labeled and shaded Normal curve for each part.

a. What percentage of battalions have between 150 and 200 bilingual officers?

b. What percentage of battalions have between 200 and 250 bilingual officers?

6.36 Bilingual Doctors A survey shows that in one year, the average number of bilingual doctors in a hospital was 42. Assume that the standard deviation is 12 and the number of bilingual doctors in a hospital is normally distributed. Include an appropriately labeled and shaded Normal curve for each part.

a. What percentage of hospitals have between 40 and 45 bilingual doctors?

b. What percentage of hospitals have between 45 and 50 bilingual doctors?

6.37 New York City Weather New York City's mean minimum daily temperature in February is 27°F (http://www.ny.com). Suppose the standard deviation of the minimum temperature is 6°F and the distribution of minimum temperatures in February is approximately Normal. What percentage of days in February has minimum temperatures below freezing (32°F)?

* **6.38 Women's Heights** Assume for this question that college women's heights are approximately Normally distributed with a mean of 64.6 inches and a standard deviation of 2.6 inches. Draw a well-labeled Normal curve for each part.

a. Find the percentage of women who should have heights of 63.5 inches or less.

b. In a sample of 123 women, according to the probability obtained in part a, how many should have heights of 63.5 inches or less?

c. The table shows the frequencies of heights for a sample of women, collected by statistician Brian Joiner in his statistics class. Count the women who appear to have heights of 63 inches or less by looking at the table. They are in the oval.

d. Are the answers to parts b and c the same or different? Explain.

Height	Inches	Frequency
	59	2
5'	60	5
	61	7
5'2"	62	10
	63	16
5'4"	64	23
	65	19
5'6"	66	15
	67	9
5'8"	68	6
	69	6
5'10"	70	3
	71	1
6'	72	1

TRY **6.39 Probability or Measurement (Inverse)? (Example 7)** The Normal model $N(500,100)$ describes the distribution of critical reading SAT scores in the United States. Which of the following questions asks for a probability and which asks for a measurement (and is thus an inverse Normal question)?

a. What reading SAT score is at the 65th percentile?

b. What is the probability that a randomly selected person will score 550 or more?

6.40 Probability or Measurement (Inverse)? The Normal model $N(65, 2.5)$ describes the distribution of heights of college women (inches). Which of the following questions asks for a probability and which asks for a measurement (and is thus an inverse Normal question)?

a. What is the probability that a random college woman has a height of 68 inches or more?

b. To be in the Tall Club, a woman must have a height such that only 2% of women are taller. What is this height?

6.41 Inverse Normal, Standard In a standard Normal distribution, if the area to the left of a z-score is about 0.6986, what is the approximate z-score?

First locate, inside the table, the number closest to 0.6986. Then find the z-score by adding 0.5 and 0.02; refer to the table. Draw a sketch of the Normal curve, showing the area and z-score.

z	.00	.01	.02	.03	.04	.05
0.4	.6554	.6591	.6628	.6664	.6700	.6736
0.5	.6915	.6950	.6985	.7019	.7054	.7088
0.6	.7257	.7291	.7324	.7357	.7389	.7422

6.42 Inverse Normal, Standard In a standard Normal distribution, if the area to the left of a z-score is about 0.2000, what is the approximate z-score?

6.43 Inverse Normal, Standard Assume a standard Normal distribution. Draw a separate, well-labeled Normal curve for each part.

a. Find the z-score that gives a left area of 0.8577.

b. Find the z-score that gives a left area of 0.0146.

6.44 Inverse Normal, Standard Assume a standard Normal distribution. Draw a separate, well-labeled Normal curve for each part.

a. Find the z-score that gives a left area of 0.9774.

b. Find the z-score that gives a left area of 0.8225.

TRY g **6.45 Females' SAT Scores (Example 8)** According to the College Board, the mean quantitative SAT score for female college-bound high school seniors in one year was 500. SAT scores are approximately Normally distributed with a population standard deviation of 100. A scholarship committee wants to give awards to college-bound women who score at the 96th percentile or above on the SAT. What score does an applicant need? Include a well-labeled Normal curve as part of your answer. *See page 317 for guidance*.

6.46 Males' SAT Scores According to the College Board, the mean quantitative SAT score for male college-bound high school seniors in one year was 530. SAT scores are approximately Normally distributed with a population standard deviation of 100. What is the SAT score at the 96th percentile for male college-bound seniors?

6.47 Tall Club, Women Suppose there is a club for tall people that requires that women be at or above the 98th percentile in height. Assume that women's heights are distributed as $N(64, 2.5)$. Find what women's height is the minimum required for joining the club, rounding to the nearest inch. Draw a well-labeled sketch to support your answer.

6.48 Tall Club, Men Suppose there is a club for tall people that requires that men be at or above the 98th percentile in height. Assume that men's heights are distributed as $N(69, 3)$. Find what men's height is the minimum required for joining the club, rounding to the nearest inch. Draw a well-labeled sketch to support your answer.

6.49 Women's Heights Suppose college women's heights are approximately Normally distributed with a mean of 65 inches and a population standard deviation of 2.5 inches. What height is at the 20th percentile? Include an appropriately labeled sketch of the Normal curve to support your answer.

6.50 Men's Heights Suppose college men's heights are approximately Normally distributed with a mean of 70.0 inches and a population standard deviation of 3 inches. What height is at the 20th percentile? Include an appropriately labeled Normal curve to support your answer.

6.51 Inverse SATs Critical reading SAT scores are distributed as $N(500, 100)$.

a. Find the SAT score at the 75th percentile.

b. Find the SAT score at the 25th percentile.

c. Find the interquartile range for SAT scores.

d. Is the interquartile range larger or smaller than the standard deviation? Explain.

6.52 Inverse Women's Heights College women have heights with the following distribution (inches): $N(65, 2.5)$.

a. Find the height at the 75th percentile.

b. Find the height at the 25th percentile.

c. Find the interquartile range for heights.

d. Is the interquartile range larger or smaller than the standard deviation? Explain.

6.53 Girls' and Women's Heights According to the National Health Center, the heights of 6-year-old girls are Normally distributed with a mean of 45 inches and standard deviation of 2 inches.

a. In which percentile is a 6-year-old girl who is 46.5 inches tall?

b. If a 6-year-old girl who is 46.5 inches tall grows up to be a woman at the same percentile of height, what height will she be? Assume women are distributed as N (64, 2.5).

6.54 Boys' and Men's Heights According to the National Health Center, the heights of 5-year-old boys are Normally distributed with a mean of 43 inches and standard deviation of 1.5 inches.

a. In which percentile is a 5-year-old boy who is 46.5 inches tall?

b. If a 5-year-old boy who is 46.5 inches tall grows up to be a man at the same percentile of height, what height will he be? Assume adult men's heights (inches) are distributed as N (69, 3).

6.55 Cats' Birth Weights The average birth weight of domestic cats is about 3 ounces. Assume that the distribution of birth weights is Normal with a standard deviation of 0.4 ounce.

a. Find the birth weight of cats at the 90th percentile.

b. Find the birth weight of cats at the 10th percentile.

6.56 Elephants' Birth Weights The average birth weight of elephants is 230 pounds. Assume that the distribution of birth weights is Normal with a standard deviation of 50 pounds. Find the birth weight of elephants at the 95th percentile.

SECTION 6.3

TRY **6.57 Gender of Children (Example 9)** A married couple plans to have four children, and they are wondering how many boys they should expect to have. Assume none of the children will be twins or other multiple births. Also assume the probability that a child will be a boy is 0.50. Explain why this is a binomial experiment. Check all four required conditions.

6.58 Dice Roll A dice will be rolled four times, and the multiples of 3 that appear will be recorded. Explain why this is a binomial experiment. Check all four required conditions.

TRY **6.59 Dice Rolls (Example 10)** A group wants to find out whether dice rolls have a 1/3 chance of coming up with multiples of 3. The leader of the group asks all members to roll dices for 10 minutes and then report their results to him. Which condition or conditions for use of the binomial model is or are not met?

6.60 Twins In Exercise 6.57 you are told to assume that none of the children will be twins or other multiple births. Why? Which of the conditions required for a binomial experiment would be violated if there were twins?

6.61 Joint Bank Account Suppose the probability that a randomly selected person who has a joint bank account with a spouse will close it within 10 years is 0.1. Suppose we follow 20 such persons (40 account holders) for 10 years and record the number of people closing their accounts. Why is the binomial model inappropriate for finding the probability that at least 19 of these 40 account holders will close their accounts within 10 years? List all binomial conditions that are not met.

6.62 Joint Bank Account Suppose the probability that a randomly selected couple opens a joint bank account within 1 year of marriage is 0.3, and the probability that a randomly selected couple opens a joint bank account after a year of marriage is 0.4. Take a random sample of 15 people opening their accounts within 1 year and 15 people opening their accounts after 1 year. The sample chosen is such that either the husband or the wife is included in it. Why is the binomial model inappropriate for finding the probability that exactly 8 out of the 30 people in the sample will open joint bank accounts within 1 year? List all of the binomial conditions that are not met.

TRY **6.63 Identifying *n*, *p*, and *x* (Example 11)** For each situation, identify the sample size n, the probability of success p, and the number of successes x. When asked for the probability, state the answer in the form $b(n, p, x)$. There is no need to give the numerical value of the probability. Assume the conditions for a binomial experiment are satisfied.

National Bureau of Statistics, Republic of Maldives, reported that 60.3% of males in the age group of 45–49 can read English.

a. If 25 males in the age group of 40–49 are chosen at random, what is the probability that 14 of them can read English?

b. If 25 males in the age group of 40–49 are chosen at random, what is the probability that 17 of them cannot read English?

6.64 Identifying *n*, *p*, and *x* For each situation, identify the sample size n, the probability of success p, and the number of successes x. When asked for the probability, state the answer in the form $b(n, p, x)$. There is no need to give the numerical value of the probability. Assume the conditions for a binomial experiment are satisfied.

a. According to the Federal Highway Research Institute in Germany, 2 out of 3 persons in an accident get killed. In a random sample of 27 persons meeting with an accident, what is the probability that exactly 12 persons would have died?

b. Twenty-five percent of the persons killed in accidents are pedestrians. If we randomly select 27 persons who have died in an accident, what is the probability that 12 persons are pedestrians?

TRY **6.65 Stolen Bicycles (Example 12)** According to the *Sydney Morning Herald*, 40% of bicycles stolen in Holland are recovered. (In contrast, only 2% of bikes stolen in New York City are recovered.) Find the probability that, in a sample of 6 randomly selected cases of bicycles stolen in Holland, exactly 2 out of 6 bikes are recovered.

6.66 Florida Recidivism Rate The three-year recidivism rate of parolees in Florida is about 30%; that is, 30% of parolees end up back in prison within three years (http://www.floridaperforms.com). Assume that whether one parolee returns to prison is independent of whether any of the others returns.

a. Find the probability that exactly 6 out of 20 parolees will end up back in prison within three years,

b. Find the probability that 6 or fewer out of 20 parolees will end up back in prison within three years.

6.67 Harvard Admission The undergraduate admission rate at Harvard University is about 6%.

a. Assuming the admission rate is still 6%, in a sample of 100 applicants to Harvard, what is the probability that exactly 5 will be admitted? Assume that decisions to admit are independent.

b. What is the probability that exactly 95 out of 100 applicants will be *rejected*?

6.68 Cornell Admission The undergraduate admission rate at Cornell University is about 16%.

a. Assuming the admission rate is still 16%, in a sample of 100 applicants to Cornell, what is the probability that exactly 15 will be admitted?

b. What is the probability that exactly 85 out of 100 independent applicants will be *rejected*?

6.69 Wisconsin Graduation Wisconsin has the highest high school graduation rate of all states at 90%.

a. In a random sample of 10 Wisconsin high school students, what is the probability that 9 will graduate?

b. In a random sample of 10 Wisconsin high school students, what is the probability than 8 or fewer will graduate?

c. What is the probability that at least 9 high school students in our sample of 10 will graduate?

6.70 Colorado Graduation Colorado has a high school graduation rate of 75%.

a. In a random sample of 15 Colorado high school students, what is the probability that exactly 9 will graduate?

b. In a random sample of 15 Colorado high school students, what is the probability that 8 or fewer will graduate?

c. What is the probability that at least 9 high school students in our sample of 15 will graduate?

6.71 Florida Homicide Clearance The homicide clearance rate in Florida is 60%. A crime is cleared when an arrest is made, a crime is charged, and the case is referred to a court.

a. What is the probability that exactly 6 out of 10 independent homicides are cleared?

b. Without doing a calculation, state whether the probability that 6 or more out of 10 homicides are cleared will be larger or smaller than the answer to part a? Why?

c. What is the probability that 6 or more out of 10 independent homicides are cleared?

6.72 Virginia Homicide Clearance The homicide clearance rate in Virginia is 74%. A crime is cleared when an arrest is made, a crime is charged, and the case is referred to a court.

a. What is the probability that 7 or fewer out of 10 independent homicides are cleared?

b. What is the probability that 7 or more out of 10 independent homicides are cleared?

c. Are the answers to parts a and part b complementary? Why or why not?

TRY **6.73 DWI Convictions (Example 13)** In New Mexico, about 70% of drivers who are arrested for driving while intoxicated (DWI) are convicted (http://www.drunkdrivingduilawblog.com).

a. If 15 independently selected drivers were arrested for DWI, how many of them would you expect to be convicted?

b. What is the probability that exactly 11 out of 15 independent selected drivers are convicted?

c. What is the probability that 11 or fewer are convicted?

6.74 Internet Access A 2013 Gallup poll indicated that about 80% of U.S. households had access to a high-speed Internet connection. Assume this rate has not changed.

a. Suppose 100 households were randomly selected from the United States. How many of the households would you expect to have access to a high-speed Internet connection?

b. If 10 households are selected randomly, what is the probability that exactly 6 have high-speed access?

c. If 10 households are selected randomly, what is the probability that 6 or fewer have high-speed access?

6.75 Drunk Walking You may have heard that drunk driving is dangerous, but what about drunk walking? According to federal information (reported in the *Ventura County Star* on August 6, 2013), 50% of the pedestrians killed in the United States had a blood-alcohol level of 0.08% or higher. Assume that two randomly selected pedestrians who were killed are studied.

a. If the pedestrian was drunk (had a blood-alcohol level of 0.08% or higher), we will record a D, and if the pedestrian was not drunk, we will record an N. List all possible sequences of D and N.

b. For each sequence, find the probability that it will occur, by assuming independence.

c. What is the probability that neither of the two pedestrians was drunk?

d. What is the probability that exactly one out of two independent pedestrians was drunk?

e. What is the probability that both were drunk?

6.76 Fish Caught in Spain According to the Eurostat Statistics Database, Spain accounts for 19% of the total fishes caught in the European Union (EU). Assume that we randomly sample two fishes caught in the EU.

a. If a fish is caught in Spain, record Y; if not, record N. List all possible sequences of Y and N.

b. For each sequence, find by hand the probability that it will occur, assuming each outcome is independent.

c. What is the probability that neither of the two randomly selected fishes have been caught in Spain?

d. What is the probability that exactly one out of the two fishes has been caught in Spain?

e. What is the probability that both have been caught in Spain?

TRY **6.77 Die Roll (Example 14)** A fair die is rolled 60 times.

a. What is the expected number of times that an odd number will turn up?

b. Find the standard deviation for the outcome to be an odd number.

c. How many times should you expect odd numbers to turn up, give or take how many times? Based on these numbers, give the range of the number of times odd numbers can turn up.

6.78 Drivers Aged 60–65 According to GMAC Insurance, 20% of drivers aged 60–65 fail the written drivers' test. This is the lowest failure rate of any age group. (Source: http://www.gmacinsurance.com/SafeDriving/PressRelease.asp) If 200 people aged 60–65 independently take the exam, how many would you expect to pass? Give or take how many?

TRY **6.79 Road Accidents** According to the Ministry of Transport in the Republic of South Africa, 80% of road accidents in Johannesburg involve males in the age group of 19–34. Suppose that there are 200 road accidents in a day.

a. What is the number of accidents involving males between the ages of 19 and 34?

b. What is the standard deviation for the number of accidents involving males between the ages of 19 and 34?

c. After a great many days, according to the Empirical Rule, on about 95% of these days, the number of accidents involving males between the ages of 19 and 34 will be as low as _______ and as high as _______. (*Hint:* Find two standard deviations below and two standard deviations above the mean.)

d. If you found that on one day, 158 out of 200 accidents involved males between the ages of 19 and 34, would you consider this to be a very high number?

6.80 Road Accidents in Small Towns In smaller cities like Port Elizabeth in South Africa, the rate of road accidents caused by males between the ages of 19 and 34 is 64%, which is much lower than the rate of accidents caused by males between the ages of 19 and 34 in Johannesburg. Suppose that there are 200 road accidents in a day.

a. What is the number of accidents caused by males between the ages of 19 and 34?

b. What is the standard deviation for the number of accidents caused by males between the ages of 19 and 34?

c. After a great many days, according to the Empirical Rule, on about 95% of these days, the number of accidents caused by males between the ages of 19 and 34 will be as low as _______ and as high as _______.

d. If you found that on one day, 158 out of 200 accidents were caused by males between the ages of 19 and 34, would you consider this to be a very high number?

CHAPTER REVIEW EXERCISES

6.81 Birth Length A study of U.S. births published on the website *Medscape from WebMD* reported that the average birth length of babies was 20.5 inches and the standard deviation was about 0.90 inch. Assume the distribution is approximately Normal. Find the percentage of babies with birth lengths of 22 inches or less.

6.82 Birth Length A study of U.S. births published on the website *Medscape from WebMD* reported that the average birth length of babies was 20.5 inches and the standard deviation was about 0.90 inch. Assume the distribution is approximately Normal. Find the percentage of babies who have lengths of 19 inches or less at birth.

6.83 Males' Body Temperatures A study of human body temperatures using healthy men showed a mean of 98.1°F and a standard deviation of 0.70°F. Assume the temperatures are approximately Normally distributed.

a. Find the percentage of healthy men with temperatures below 98.6°F (that temperature was considered typical for many decades).

b. What temperature does a healthy man have if his temperature is at the 76th percentile?

6.84 Females' Body Temperatures A study of human body temperatures using healthy women showed a mean of 98.4°F and a standard deviation of about 0.70°F. Assume the temperatures are approximately Normally distributed.

a. Find the percentage of healthy women with temperatures below 98.6°F (this temperature was considered typical for many decades).

b. What temperature does a healthy woman have if her temperature is at the 76th percentile?

6.85 Cremation Rates in Nevada

Cremation rates have been increasing. In Nevada the cremation rate is 70%. Suppose that we take a random sample of 400 deaths in Nevada.

a. How many of these decedents would you expect to be cremated?

b. What is the standard deviation for the number to be cremated?

c. How many would you expect not to be cremated?

d. What is the standard deviation for the number not to be cremated?

* e. Explain the relationship between the answers to parts b and d.

* **6.86 Self-employment Rates in India** Self-employment rates have been increasing in the rural areas of India. In urban areas, this rate is as low as 41.1%. Suppose that we take a random sample of 800 fresh graduates in urban areas and they independently make the decision of being self-employed.

a. How many of the 800 graduates would you expect to be self-employed?

b. What is the standard deviation for the number of self-employed?

c. According to the Empirical Rule, in 95% of all samples of 800 persons, the number of self-employed would vary between what two values?

d. In a random sample of 800, how many would you expect not to choose self-employment?

e. If 500 were self-employed and 300 were not self-employed, would 500 be a surprisingly large number of self-employed?

* **6.87 Low Birth Weights, Normal *and* Binomial** Babies weighing 5.5 pounds or less at birth are said to have low birth weights, which can be dangerous. Full-term birth weights for single babies (not twins or triplets or other multiple births) are Normally distributed with a mean of 7.5 pounds and a standard deviation of 1.1 pounds.

a. For one randomly selected full-term single-birth baby, what is the probability that the birth weight is 5.5 pounds or less?

b. For two randomly selected full-term, single-birth babies, what is the probability that both have birth weights of 5.5 pounds or less?

c. For 200 random full-term single births, what is the approximate probability that 7 or fewer have low birth weights?

d. If 200 independent full-term single-birth babies are born at a hospital, how many would you expect to have birth weights of 5.5 pounds or less? Round to the nearest whole number.

e. What is the standard deviation for the number of babies out of 200 who weigh 5.5 pounds or less? Retain two decimal digits for use in part f.

f. Report the birth weight for full-term single babies (with 200 births) for two standard deviations below the mean and for two standard deviations above the mean. Round both numbers to the nearest whole number.

g. If there were 45 low-birth-weight full-term babies out of 200, would you be surprised?

* **6.88 Quantitative SAT Scores, Normal *and* Binomial**

The distribution of the math portion of SAT scores has a mean of 500 and a standard deviation of 100, and the scores are approximately Normally distributed.

a. What is the probability that one randomly selected person will have an SAT score of 550 or more?

b. What is the probability that four randomly selected people will all have SAT scores of 550 or more?

c. For 800 randomly selected people, what is the probability that 250 or more will have scores of 550 or more?

d. For 800 randomly selected people, on average how many should have scores of 550 or more? Round to the nearest whole number.

e. Find the standard deviation for part d. Round to the nearest whole number.

f. Report the range of people out of 800 who should have scores of 550 or more from two standard deviations below the mean to two standard deviations above the mean. Use your rounded answers to part d and e.

g. If 400 out of 800 randomly selected people had scores of 550 or more, would you be surprised? Explain.

6.89 Babies' Birth Length, Inverse Babies in the United States have a mean birth length of 20.5 inches with a standard deviation of 0.90 inch. The shape of the distribution of birth lengths is approximately Normal.

a. How long is a baby born at the 20th percentile?

b. How long is a baby born at the 50th percentile?

c. How does your answer to part b compare to the mean birth length? Why should you have expected this?

6.90 Birth Length and *z*-Scores, Inverse Babies in the United States have a mean birth length of 20.5 inches with a standard deviation of 0.90 inch. The shape of the distribution of birth lengths is approximately Normal.

a. Find the birth length at the 2.5th percentile.

b. Find the birth length at the 97.5th percentile.

c. Find the *z*-score for the length at the 2.5th percentile.

d. Find the *z*-score for the length at the 97.5th percentile.

GUIDED EXERCISES

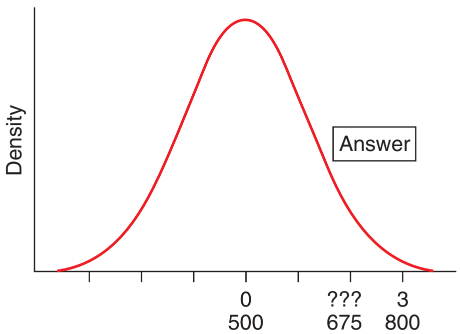
g **6.25 Females' SAT Scores** According to data from the College Board, the mean quantitative SAT score for female college-bound high school seniors is 500. SAT scores are approximately Normally distributed with a population standard deviation of 100.

QUESTION What percentage of the female college-bound high school seniors had scores above 675? Answer this question by following the numbered steps.

Step 1 ▶ Find the *z*-score
To find the *z*-score for 675, subtract the mean and divide by the standard deviation. Report the *z*-score.

Step 2 ▶ Explain the location of 500
Refer to the Normal curve. Explain why the SAT score of 500 is right below the *z*-score of 0. The tick marks on the axis mark the location of *z*-scores that are integers from −3 to 3.

Step 3 ▶ Label with SAT scores
Carefully sketch a copy of the curve. Pencil in the SAT scores of 200, 300, 400, 600, and 700 in the correct places.



Step 4 ▶ Add the line, *z*-score, and shading
Draw a vertical line through the curve at the location of 675. Just above the 675 (indicated on the graph with "???") put in the corresponding *z*-score. We want to find what percentage of students had scores *above* 675. Therefore, shade the area to the *right* of this boundary, because numbers to the right are larger.

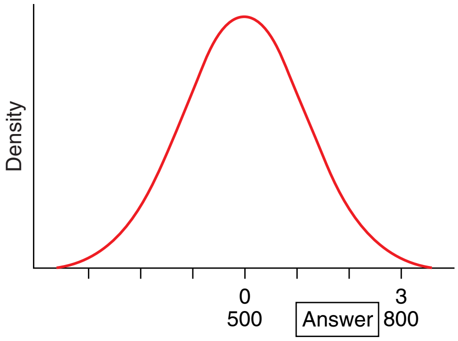
Step 5 ▶ Use the table for the left area
Use the following excerpt from the Normal table to find and report the area to the left of the *z*-score that was obtained from an SAT score of 675. This is the area of the unshaded region.

Step 6 ▶ Answer
Because you want the area to the right of the *z*-score, you will have to subtract the area you obtained in step 5 from 1. This is the area of the shaded region. Put it where the box labeled "Answer" is. Check to see that the number makes sense. For example, if the shading is less than half the area, the answer should not be more than 0.5000.

Step 7 ▶ Sentence
Finally, write a sentence stating what you found.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	**.9599**	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706

g **6.45 Females' SAT scores** According to the College Board, the mean quantitative SAT score for female college-bound high school seniors in one year was 500. SAT scores are approximately Normally distributed with a population standard deviation of 100. A scholarship committee wants to give awards to college-bound women who score at the 96th percentile or above on the SAT. What score does an applicant need to receive an award? Include a well-labeled Normal curve as part of your answer.



QUESTION What is the SAT score at the 96th percentile? Answer this question by following the numbered steps.

Step 1 ▶ Think about it
Will the SAT test score be above the mean or below it? Explain.

Step 2 ▶ Label *z*-scores
Label the curve with integer *z*-scores. The tick marks represent the position of integer *z*-scores from −3 to 3.

Step 3 ▶ Use the table
The 96th percentile has 96% of the area to the *left* because it is higher than 96% of the scores. The table above gives the areas to the *left* of *z*-scores. Therefore, we look for 0.9600 in the interior part of the table.

Use the excerpt of the Normal table given on page 317 for Exercise 6.25 to locate the area *closest* to 0.9600.

Then report the *z*-score for that area.

Step 4 ▶ Add the *z*-score, line, and shading to the sketch
Add that *z*-score to the sketch, and draw a vertical line above it through the curve. Shade the left side because the area to the left is what is given.

Step 5 ▶ Find the SAT score
Find the SAT score that corresponds to the *z*-score. The score should be *z* standard deviations above the mean, so

$$x = \mu + z\sigma$$

Step 6 ▶ Add the SAT score to the sketch
Add the SAT score on the sketch where it says "Answer."

Step 7 ▶ Write a sentence
Finally, write a sentence stating what you found.

TechTips

For All Technology

All technologies will use the two examples that follow.

EXAMPLE A: NORMAL ► Wechsler IQs have a mean of 100 and standard deviation of 15 and are Normally distributed.

a. Find the probability that a randomly chosen person will have an IQ between 85 and 115.

b. Find the probability that a randomly chosen person will have an IQ that is 115 or less.

c. Find the Wechsler IQ at the 75th percentile.

Note: If you want to use technology to find areas from standard units (z-scores), use a mean of 0 and a standard deviation of 1.

EXAMPLE B: BINOMIAL ► Imagine that you are flipping a fair coin (one that comes up heads 50% of the time in the long run).

a. Find the probability of getting 28 or fewer heads in 50 flips of a fair coin.

b. Find the probability of getting exactly 28 heads in 50 flips of a fair coin.

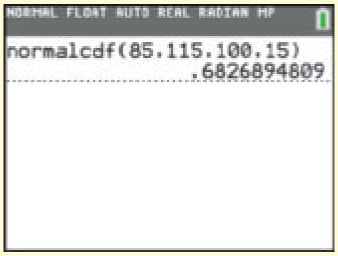
TI-84

NORMAL

a. Between Two Values

1. Press **2ND DISTR** (located below the four arrows on the keypad).
2. Select **2:normalcdf** and press **ENTER**.
3. Enter **lower**: **85**, **upper: 115**, **μ: 100**, **σ: 15**. For **Paste**, press **ENTER**. Then press **ENTER** again.

Your screen should look like Figure 6A, which shows that the probability that a randomly selected person will have a Wechsler IQ between 85 and 115 is equal to 0.6827.



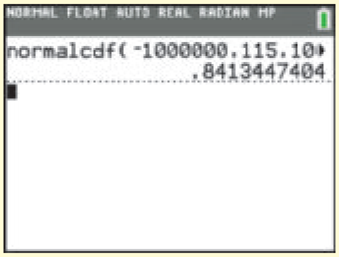
▲ **FIGURE 6A** TI-84 normalcdf (c stands for "cumulative")

b. Some Value or Less

1. Press **2ND DISTR**.
2. Select **2:normalcdf** and press **ENTER**.
3. Enter: **−1000000, 115, 100, 15,** press **ENTER** and press **ENTER** again.

Caution: The negative number button **(–)** is to the left of the **ENTER** button and is not the same as the minus button that is above the plus button.

The probability that a person's IQ is 115 *or less* has an *indeterminate* lower (left) boundary, for which you may use negative 1000000 or any extreme value that is clearly out of the range of data. Figure 6B shows the probability that a randomly selected person will have an IQ of 115 or less.



▲ **Figure 6B** TI-84 normalcdf with indeterminate left boundary

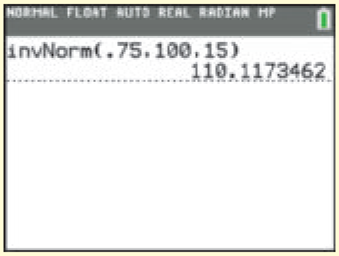
(If you have an indeterminate upper, or right boundary, then to find the probability that the person's IQ is 85 or more, for example, use a upper, or right boundary (such as 1000000) that is clearly above all the data.)

c. Inverse Normal

If you want a measurement (such as an IQ) from a proportion or percentile:

1. Press **2ND DISTR**.
2. Select **3:invNorm** and press **ENTER**.
3. Enter (left) **area: .75**, **μ: 100**, **σ: 15**. For **Paste**, press **ENTER**. Then **ENTER** again.

Figure 6C shows the Wechsler IQ at the 75th percentile, which is 110. Note that the 75th percentile is entered as **.75**.



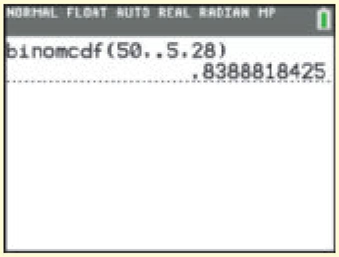
▲ **Figure 6C** TI-84 Inverse Normal

BINOMIAL

a. Cumulative (or Fewer)

1. Press **2ND DISTR.**
2. Select **B:binomcdf** (you will have to scroll down to see it) and press **ENTER**. (On a TI-83, it is **A:binomcdf**.)
3. Enter **trials: 50, p: .5, x value: 28**. For **Paste**, press **ENTER**. Then press **ENTER** again.

The answer will be the probability for *x or fewer*. Figure 6D shows the probability of 28 *or fewer* heads out of 50 flips of a fair coin. (You could find the probability of 29 *or more* heads by subtracting your answer from 1.)

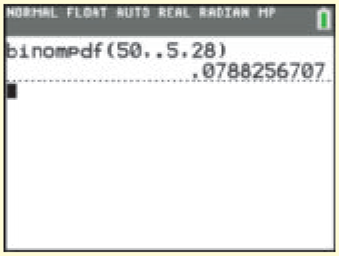


▲ **Figure 6D** TI-84 binomcdf (cumulative)

b. Individual (Exact)

1. Press **2ND DISTR**.
2. Select **A:binompdf** and press **ENTER**. (On a TI-83, it is **0:binompdf**.)
3. Enter **trials: 50, p: .5, x value: 28**. For **Paste,** press **ENTER**. Then press **ENTER** again.

Figure 6E shows the probability of *exactly* 28 heads out of 50 flips of a fair coin.



▲ **Figure 6E** TI-84 binompdf (individual)

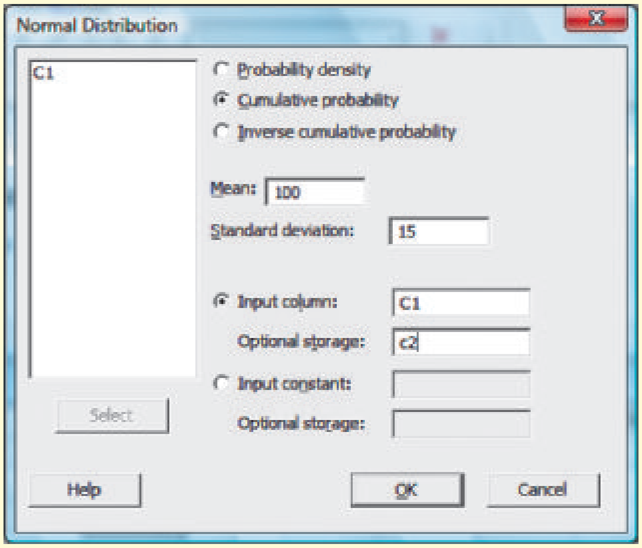
MINITAB

Normal

a. Between Two Values

1. Enter the upper boundary, **115**, in the top cell of an empty column; here we use column C1, row 1. Enter the lower boundary, **85**, in the cell below; here column C1, row 2.
2. **Calc** > **Probability Distributions** > **Normal**.
3. See Figure 6F: Choose **Cumulative probability**. Enter: **Mean**, 100**; Standard deviation**, 15**; Input column**, C1; **Optional storage,** C2**.**
4. Click **OK**.
5. Subtract the lower probability from the larger shown in column C2.

   $0.8413 - 0.1587 = 0.6836$ is the probability that a Wechsler IQ is between 85 and 115.



▲ **Figure 6F** Minitab Normal

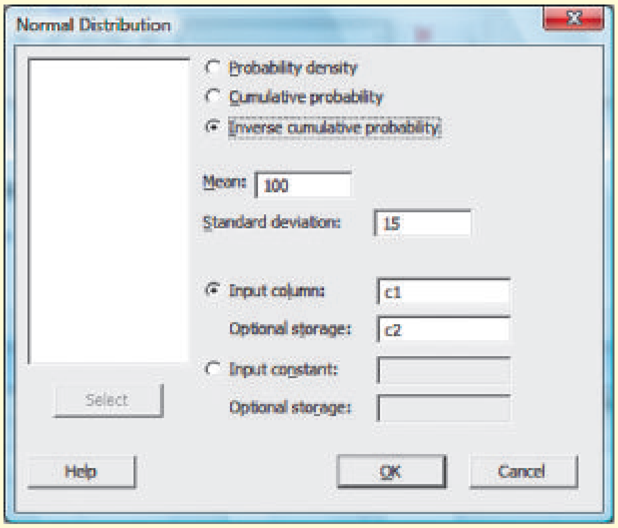
b. Some Value or Less

1. The probability of an IQ of 115 *or less*, 0.8413, is shown in column C2, row 1. (In other words, do as in part a above, except do *not* enter the lower boundary, 85.)

c. Inverse Normal

If you want a measurement (such as an IQ or height) from a proportion or percentile:

1. Enter the decimal form of the left proportion (**.75** for the 75th percentile) into a cell in an empty column in the spreadsheet; here we used column C1, row 1.
2. **Calc** > **Probability Distributions** > **Normal**.
3. See Figure 6G: Choose **Inverse cumulative probability**. Enter: **Mean**, **100**; **Standard deviation**, **15**; **Input column**, **c1**; and **Optional storage**, **c2** (or an empty column).



▲ **Figure 6G** Minitab Inverse Normal

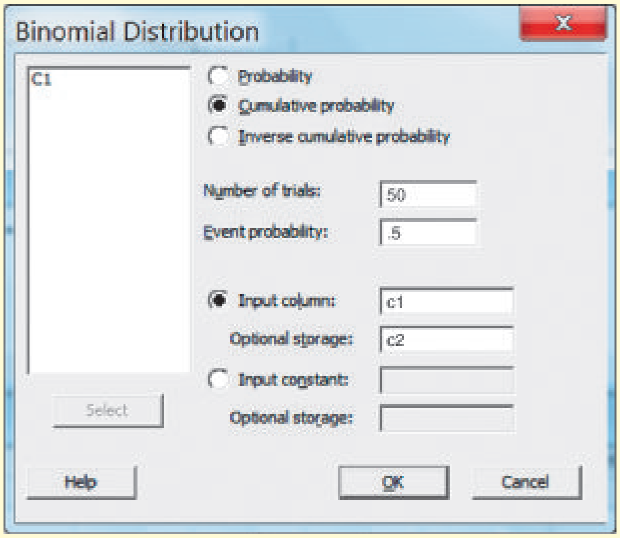
4. Click **OK**.

You will get **110**, which is the Wechsler IQ at the 75th percentile.

Binomial

a. Cumulative (or Fewer)

1. Enter the upper bound for the number of successes in an empty column; here we used column C1, row 1. Enter **28** to get the probability of 28 or fewer heads.
2. **Calc** > **Probability Distributions** > **Binomial**.
3. See Figure 6H. Choose **Cumulative probability**.



▲ **Figure 6H** Minitab Binomial

Enter: **Number of trials**, **50; Event probability, .5; Input column, c1; Optional storage, c2** (or an empty column).

4. Click **OK**.

Your answer will be 0.8389 for the probability of 28 *or fewer* heads.

b. Individual (Exact)

1. Enter the number of successes at the top of column 1, **28** for 28 heads.
2. **Calc** > **Probability Distributions** > **Binomial**.
3. Choose **Probability** (at the top of Figure 6H) instead of **Cumulative Probability** and enter: **Number of trials**, **50; Event probability, .5; Input column**, **c1; Optional storage, c2** (or an empty column).
4. Click **OK.**

Your answer will be 0.0788 for the probability of *exactly* 28 heads.

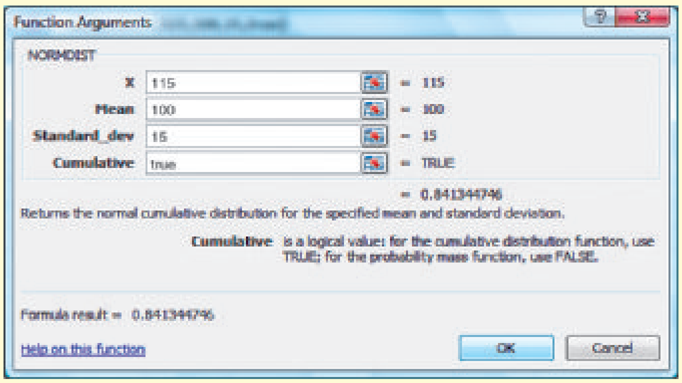
EXCEL

Normal

Unlike the TI-84, Excel makes it easier to find the probability that a random person has an IQ of 115 or less than to find the probability that a random person has an IQ between 85 and 115. This is why, for Excel, part b appears before part a.

b. Some Value or Less

1. Click ***fx*** (and **select a category All**).
2. Choose **NORM.DIST**.
3. See Figure 6I. Enter: **X**, **115; Mean, 100; Standard_dev, 15; Cumulative, true** (for 115 *or less*). The answer is shown as 0.8413. Click **OK** to make it show up in the active cell on the spreadsheet.



▲ **Figure 6I** Excel Normal

a. Between Two Values

If you want the probability of an IQ between 85 and 115:

1. First, follow the instructions given for part b. *Do not change the active cell in the spreadsheet.*
2. You will see **=NORMDIST(115,100,15,TRUE)** in the ***fx*** box. Click in this box, to the right of **TRUE)** and put in a minus sign.
3. Now repeat the steps for part b, starting by clicking ***fx***, except enter 85 instead of 115 for X. The answer, **0.682689**, will be shown in the active cell.

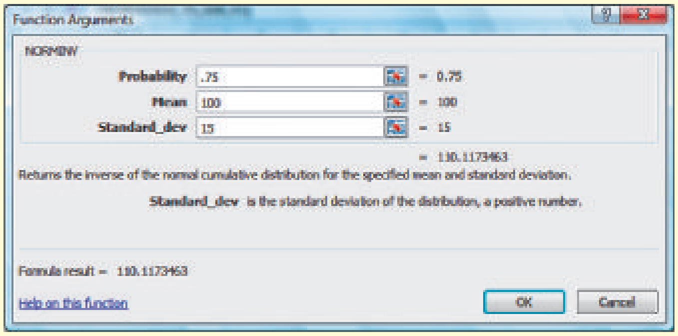
   (Alternatively, just repeat steps 123 for part b, using 85 instead of 115. Subtract the smaller probability value from the larger (0.8413 − 0.1587 = 0.6826).)

c. Inverse Normal

If you want a measurement (such as an IQ or height) from a proportion or percentile:

1. Click f_x.
2. Choose **NORM.INV** and click **OK**.
3. See Figure 6J. Enter: **Probability**, **.75** (for the 75th percentile); **Mean, 100: Standard_dev**, **15**. You may read the answer off the screen or click **OK** to see it in the active cell in the spreadsheet.

   The IQ at the 75th percentile is 110.



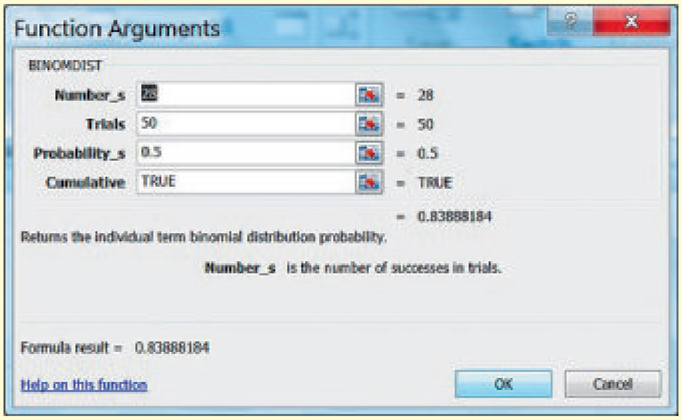
▲ **Figure 6J** Excel Inverse Normal

Binomial

a. Cumulative (or Fewer)

1. Click f_x.
2. Choose **BINOM.DIST** and click **OK**.
3. See Figure 6K. Enter: **Number_s, 28**; **Trials, 50**; **Probability_s**, **.5**; and **Cumulative**, **true** (for the probability of 28 *or fewer*).

The answer (**0.8389**) shows up in the dialogue box and in the active cell when you click **OK**.



▲ **Figure 6K** Excel Binomial

b. Individual (Exact)

1. Click f_x.
2. Choose **BINOM.DIST** and click **OK**.
3. Use the numbers in Figure 6K, but enter **False** in the Cumulative box. This will give you the probability of getting *exactly* 28 heads in 50 tosses of a fair coin.

The answer (**0.0788**) shows up in the dialogue box and in the active cell when you click **OK**.

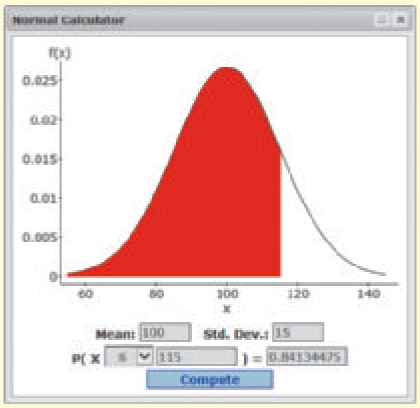
STATCRUNCH

Normal

Unlike the TI-84, StatCrunch makes it easier to find the probability for 115 or less than to find the probability between 85 and 115. This is why part b is done before part a.

b. Some Value or Less

1. **Stat** > **Calculators** > **Normal**
2. See Figure 6L. To find the probability of having a Wechsler IQ of 115 or less, Enter: **Mean**, **100; Std Dev**, **15**. Make sure that the arrow to the right of **P(X** points left (for less than). Enter the **115** in the box above **Compute**.
3. Click **Compute** to see the answer, **0.8413**.



▲ **Figure 6L** StatCrunch Normal

a. Between Two Values

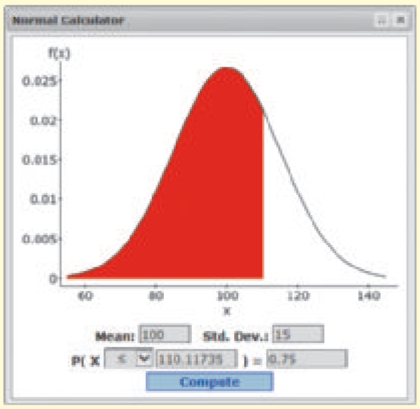
To find the probability of having a Wechsler IQ between 85 and 115, use steps 1, 2, and 3 again, but use **85** instead of **115** in the box above **Snapshot**. When you find that probability, subtract it from the probability found in Figure 6L.

$$0.8413 - 0.1587 = 0.6826$$

c. Inverse Normal

If you want a measurement (such as an IQ or height) from a proportion or percentile:

1. **Stat** > **Calculators** > **Normal**
2. See Figure 6M. To find the Wechsler IQ at the 75th percentile, enter: Mean, **100**; **Std. Dev.**, **15**. Make sure that the arrow to the right of **P(X** points to the left, and enter **0.75** in the box to the right of the = sign.



▲ **Figure 6M** StatCrunch Inverse Normal

3. Click **Compute** and the answer (**110**) is shown above **Compute**.

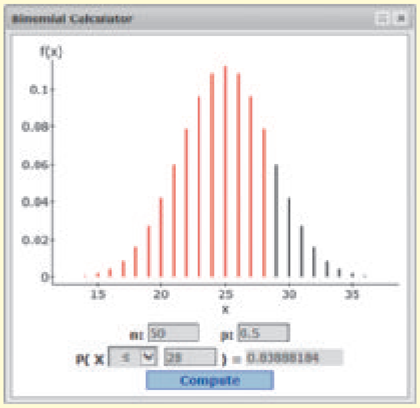
Binomial

a. Cumulative (or Fewer)

1. **Stat** > **Calculators** > **Binomial**
2. See Figure 6N. To find the probability of 28 or fewer heads in 50 tosses of a fair coin, enter: **n**, **50;** and **p, 0.5**. The arrow after **P(X** should point left (for *less than*). Enter **28** in the box above **Compute**.
3. Click **Compute** to see the answer (**0.8389**).

b. Individual (Exact)

1. **Stat** > **Calculators** > **Binomial**
2. To find the probability of exactly 28 heads in 50 tosses of a fair coin, use a screen similar to Figure 6N, but to the right of **P(X** choose the equals sign. You will get **0.0788**.



▲ **Figure 6N** StartCrunch Binomial

7 Survey Sampling and Inference

THEME

If survey subjects are chosen randomly, then we can use their answers to infer how the entire population would answer. We can also quantify how far off our estimate is likely to be.

Somewhere in your town or city, possibly at this very moment, people are participating in a survey. Perhaps they are filling out a customer satisfaction card at a restaurant. Maybe their television is automatically transmitting information about which show is being watched so that marketers can estimate how many people are viewing their ads. They may even be text messaging in response to a television survey. Most of you will receive at least one phone call from a survey company that will ask whether you are satisfied with local government services or plan to vote for one candidate over another. The information gathered by these surveys is used to piece together, bit by bit, a picture of the larger world.

You've reached a pivotal point in the text. In this chapter, the data summary techniques you learned in Chapters 2 and 3, the probability you learned about in Chapter 5, and the Normal distribution, which you studied in Chapter 6, are all combined to enable us to generalize what we learn about a small sample to a larger group. Politicians rely on surveys of 1000 voters not because they care how those 1000 individuals will vote. Surveys are important to politicians only if they help them learn about *all* potential voters. In this and later chapters, we study ways to understand and measure just how reliable this projection from sample to the larger world is.

Whenever we draw a conclusion about a large group based on observations of some parts of that group, we are making an inference. Inferential reasoning lies at the foundation of science but is far from foolproof. As the following case study illustrates, when we make an inference, we can never be absolutely certain of our conclusions. But applying the methods introduced in this chapter ensures that if we collect data carefully, we can at least measure how certain or uncertain we are.

CASE STUDY

Spring Break Fever: Just What the Doctors Ordered?

In 2006, the American Medical Association (AMA) issued a press release ("Sex and intoxication among women more common on spring break according to AMA poll") in which it concluded, among other things, that "eighty-three percent of the [female, college-attending] respondents agreed spring break trips involve more or heavier drinking than occurs on college campuses and 74 percent said spring break trips result in increased sexual activity." This survey made big news, particularly since the authors of the study claimed these percentages reflected the opinions not only of the 644 women who responded to the survey but of all women who participated in spring break.

The AMA's website claimed the results were based on "a nationwide random sample of 644 women who . . . currently attend college. . . . The survey has a margin of error of +/−4 percentage points at the 95 percent level of confidence." It all sounds very scientific, doesn't it?

However, some survey specialists were suspicious. After Cliff Zukin, a specialist who was president of the American Association for Public Opinion Research, corresponded with the AMA, it changed its website posting to say the results were based not on a random sample, but instead on "a nationwide sample of 644 women . . .

who are part of an *online survey panel* . . . [emphasis added].” “Margin of error” is no longer mentioned.

Disagreements over how to interpret these results show just how difficult inference is. In this chapter you’ll see why the method used to collect data is so important to inference, and how we use probability, under the correct conditions, to calculate a margin of error to quantify our uncertainty. At the end of the chapter, you’ll see why the AMA changed its report.

SECTION 7.1

Learning about the World through Surveys

Surveys are probably the most often encountered application of statistics. Most news shows, newspapers, and magazines report on surveys or polls several times a week—and during a major election, several times a day. We can learn quite a bit through a survey if the survey is done correctly.

Survey Terminology

A **population** is a group of objects or people we wish to study. Usually, this group is large—say, the group of all U.S. citizens, or all U.S. citizens between the ages of 13 and 18, or all senior citizens. However, it might be smaller, such as all phone calls made on your cell phone in January. We wish to know the value of a **parameter**, a numerical value that characterizes some aspect of this *population*. For example, political pollsters want to know what percentage of people say they will vote in the next election. Drunk-driving opponents want to know what percentage of all teenagers with driver’s licenses have drunk alcohol while driving. Designers of passenger airplanes want to know the mean length of passengers’ legs so that they can put the rows of seats as close together as possible without causing discomfort.

In this text we focus on two frequently used parameters: the mean of a population and the population proportion. This chapter deals with population proportions.

If the population is relatively small, we can find the exact value of the parameter by conducting a census. A **census** is a survey in which every member of the population is measured. For example, if you wish to know the percentage of people in your classroom who are left-handed, you can perform a census. The classroom is the population, and the parameter is the percentage of left-handers. We sometimes try to take a census with a large population (such as the U.S. Census), but such undertakings are too expensive for nongovernmental organizations and are filled with complications caused by trying to track down and count people who may not want to be found. (For example, the U.S. Census tends to undercount poor, urban-dwelling residents, as well as undocumented immigrants.)

In fact, most populations we find interesting are too large for a census. For this reason, we instead observe a smaller sample. A **sample** is a collection of people or objects taken from the population of interest.

Once a sample is collected, we measure the characteristic we’re interested in. A **statistic** is a numerical characteristic of a sample of data. We use statistics to estimate parameters. For instance, we might be interested in knowing what proportion of all registered voters will vote in the next national election. The proportion of all registered voters who will vote in the next election is our *parameter*. Our method to estimate this parameter is to survey a small sample. The proportion of the sample who say they will vote in the next election is a *statistic*.

Statistics are sometimes called **estimators**, and the numbers that result are called **estimates**. For example, our *estimator* is the proportion of people in a sample who say they will vote in the next election. When we conduct this survey, we find, perhaps, that 0.75 of the sample say they will vote. This number, 0.75, is our *estimate*.

KEY POINT A statistic is a number that is based on data and used to estimate the value of a characteristic of the population. Thus it is sometimes called an estimator.

Statistical inference is the art and science of drawing conclusions about a population on the basis of observing only a small subset of that population. Statistical inference always involves uncertainty, so an important component of this science is measuring our uncertainty.

EXAMPLE 1 Pew Poll: Age and the Internet

In February 2014 (about the time of Valentine's Day), the Pew Research Center surveyed 1428 adults in the United States who were married or in a committed partnership. The survey found that 25% of cell phone owners felt that their spouse or partner was distracted by her or his cell phone when they were together.

QUESTIONS Identify the population and the sample. What is the parameter of interest? What is the statistic?

SOLUTION The population that the Pew Research Center wanted to study consists of all American adults who were married or in a committed partnership and owned a cell phone. The sample, which was taken from the population consists of 1428 such people. The parameter of interest is the percentage of all adults in the United States who were married or in a committed partnership and felt that their spouse or partner was distracted by her or his cell phone when they were together. The statistic, which is the percentage of the sample who felt this way, is 25%.

TRY THIS! Exercise 7.1

An important difference between statistics and parameters is that statistics are knowable. Any time we collect data, we can find the value of a statistic. In Example 1, we know that 25% of those surveyed felt that their partner was distracted by the cell phone. In contrast, a parameter is typically unknown. We do not know for certain the percentage of *all* people who felt this way about their partners. The only way to find out would be to ask everyone, and we have neither the time nor the money to do this. Table 7.1 compares the known and the unknown in this situation.

Unknown	Known
Population All cell phone owners in a committed relationship	*Sample* A small number of cell phone owners in a committed relationship
Parameter Percentage of all cell-phone owners in a committed relationship who felt that their partner was distracted when they were together	*Statistic* Percentage of the sample who felt their partner was distracted when they were together

◀ **TABLE 7.1** Some examples of unknown quantities we might wish to estimate, and their knowable counterparts.

Statisticians have developed notation for keeping track of parameters and statistics. In general, Greek characters are used to represent population parameters. For example, μ (mu, pronounced "mew," like the beginning of *mu*sic) represents the mean of a population. Also, σ (sigma) represents the standard deviation of a population. Statistics (estimates based on a sample) are represented by English letters: $\bar{x}$ (pronounced "x-bar") is the mean of a sample, and s is the standard deviation of a sample, for instance.

One frequently encountered exception is the use of the letter p to represent the proportion of a population and $\hat{p}$ (pronounced "p-hat") to indicate the proportion of a sample. Table 7.2 summarizes this notation. You've seen most of these symbols before, but this table organizes them in a new way that is important for statistical inference.

TABLE 7.2 Notation for some commonly used statistics and parameters.

Statistics (based on data)		Parameters (typically unknown)	
Sample mean	$\bar{x}$ (x-bar)	Population mean	μ (mu)
Sample standard deviation	s	Population standard deviation	σ (sigma)
Sample variance	s^2	Population variance	σ^2
Sample proportion	$\hat{p}$ (p-hat)	Population proportion	p

What Could Possibly Go Wrong? The Problem of Bias

Unfortunately, it is far easier to conduct a bad survey than to conduct a good survey. One of the many ways in which we can reach a wrong conclusion is to use a survey method that is biased.

A method is **biased** if it has a tendency to produce an untrue value. Bias can enter a survey in three ways. The first is through **sampling bias**, which results from taking a sample that is not representative of the population. A second way is **measurement bias**, which comes from asking questions that do not produce a true answer. For example, if we ask people their income, they are likely to inflate the value. In this case, we will get a positive (or "upward") bias: Our estimate will tend to be too high. Measurement bias occurs when measurements tend to record values larger (or smaller) than the true value.

The third way occurs because some statistics are naturally biased. For example, if you use the statistic $10\bar{x}$ to estimate the mean, you'll typically get estimates that are ten times too big. Therefore, even when no measuring or sampling bias is present, you must also take care to use an estimator that is not biased.

> **! Caution**
>
> **Bias**
> Statistical bias is different from the everyday use of the term *bias*. You might perhaps say a friend is biased if she has a strong opinion that affects her judgment. In statistics, bias is a way of measuring the performance of a method over many different applications.

Measurement Bias In February 2010, the *Albany Times Union* newspaper reported on two recent surveys to determine the opinions of New York State residents on taxing soda (Crowley 2010). The Quinnipiac University Polling Institute asked, "There is a proposal for an 'obesity tax' or a 'fat tax' on non-diet sugary soft drinks. Do you support or oppose such a measure?" Forty percent of respondents said they supported the tax. Another firm, Kiley and Company, asked, "Please tell me whether you feel the state should take that step in order to help balance the budget, should seriously consider it, should consider it only as a last resort, or should definitely not consider taking that step: 'Imposing a new 18 percent tax on sodas and other soft drinks containing sugar, which would also reduce childhood obesity.'" Fifty-eight percent supported the tax when asked this question. One or both of these surveys have measurement bias.

A famous example occurred in 1993, when, on the basis of the results of a Roper Organization poll, many U.S. newspapers published headlines similar to this one from the *New York Times*: "1 in 5 in New Survey Express Some Doubt About the Holocaust" (April 20, 1993). Almost a year later, the *New York Times* reported that this alarmingly high percentage of alleged Holocaust doubters could be due to measurement error. The actual question respondents were asked contained a double negative: "Does it seem possible, or does it seem impossible to you, that the Nazi extermination of the Jews never happened?" When Gallup repeated the poll but did not use a double negative, only 9% expressed doubts (*New York Times* 1994).

Sampling Bias Writing good survey questions to reduce measurement bias is an art and a science. This text, however, is more concerned with sampling bias, which occurs when the estimation method uses a sample that is not representative of the population. (By "not representative" we mean that the sample is fundamentally different from the population.)

Have you ever heard of Alfred Landon? Unless you're a political science student, you probably haven't. In 1936, Landon was the Republican candidate for U.S. president, running against Franklin Delano Roosevelt. The *Literary Digest*, a popular news magazine, conducted a survey with over 10 million respondents and predicted that Landon would easily win the election with 57% of the vote. The fact that you probably haven't heard of Landon suggests that he didn't win, and in fact, he lost big, setting a record at the time for the fewest electoral votes ever received by a major-party candidate. What went wrong? The *Literary Digest* had a biased sample. The journal relied largely on polling its own readers, and its readers were more well-to-do than the general public and more likely to vote for a Republican. The reputation of the *Literary Digest* was so damaged that two years later it disappeared and was absorbed into *Time* magazine.

The U.S. presidential elections of 2004 and 2008 both had candidates who claimed to have captured the youth vote, and both times, candidates claimed the polls were biased. The reason given was that the surveys used to estimate candidate support relied on landline phones, and many young voters don't own landlines, relying instead on their cell phones. Reminiscent of the 1936 *Literary Digest* poll, these surveys were potentially biased because their sample systematically excluded an important part of the population: those who did not use landlines (Cornish 2007).

In fact, the Pew Foundation conducted a study after the 2010 congressional elections. This study found that polls that excluded cell phones had a sampling bias in favor of Republican candidates.

Today, the most commonly encountered biased surveys are probably Internet polls. These can be found on many news organization websites. ("Tea Party Influence in Washington, D.C. is (a) on the rise (b) on the decline (c) unchanged?" www.foxnews.com, February 2014.). Internet polls suffer from what is sometimes called **response bias**. People tend to respond to such surveys only if they have strong feelings about the results; otherwise, why bother? This implies that the sample of respondents is not necessarily representative of the population. Even if the population in this case is, for example, all readers of the Foxnews.com website, the survey may not accurately reflect their views, because the voluntary nature of the survey means the sample will probably be biased. This bias might be even worse if we took the population to be all U.S. residents. Readers of Internet websites may very well not be representative of all U.S. residents, and readers of particular websites such as Fox or CNN might be even less so.

To warn readers of this fact, most Internet polls have a disclaimer: "This is not a scientific poll." What does this mean? It means we should not trust the information reported to tell us anything about anyone other than the people who responded to the poll. (And remember, we can't even trust the counts on an Internet poll, because sometimes nothing prevents people from voting many times.)

KEY POINT

When reading about a survey, it is important to know

1. what percentage of people who were asked to participate actually did so
2. whether the researchers chose people to participate in the survey or people themselves chose to participate.

If a large percentage of those chosen to participate refused to answer questions, or if people themselves chose whether to participate, the conclusions of a survey are suspect.

Because of response bias, you should always question what type of people were included in a survey. But the other side of this coin is that you should also question what type of people were left out. Was the survey conducted at a time of day that meant that working people were less likely to participate? Were only landline phones used, thereby excluding people who had only cell phones? Was the question that was asked potentially embarrassing, so that people might have refused to answer? All of these circumstances can bias survey results.

Simple Random Sampling Saves the Day

How do we collect a sample that has as little **bias** as possible and is representative of the population? Only one way works: to take a random sample.

As we explained in Chapter 5, statisticians have a precise definition of *random.* A random sample does not mean that we stand on a street corner and stop whomever we like to ask them to participate in our survey. (Statisticians call this a **convenience sample**, for obvious reasons.) A random sample must be taken in such a way that every person in our population is equally likely to be chosen.

A true random sample is difficult to achieve. (And that's a big understatement!) Pollsters have invented many clever ways of pulling this off, often with great success. One basic method that's easy to understand but somewhat difficult to put into practice is **simple random sampling (SRS)**.

In SRS, we draw subjects from the population at random and without replacement. **Without replacement** means that once a subject is selected for a sample, that subject cannot be selected again. This is like dealing cards from a deck. Once a card is dealt for a hand, no one else can get the same card. A result of this method is that every sample of the same fixed size is equally likely to be chosen. As a result, we can produce unbiased estimations of the population parameters of interest and can measure the precision of our estimator.

In theory, we can take an SRS by assigning a number to each and every member of the population. We then use a random number table or other random number generator to select our sample, ignoring numbers that appear twice.

! Caution

Random
If a sample is not random, there's really nothing we can learn about the population. We can't measure the survey's precision, and we can't know how large the bias might be.

Details

Simple random sampling is not the only valid method for statistical inference. Statisticians collect representative samples using other methods, as well (for example, sampling **with replacement**). What these methods all have in common is that they take samples randomly.

EXAMPLE 2 Taking a Simple Random Sample

Alberto, Justin, Michael, Audrey, Brandy, and Nicole are in a class.

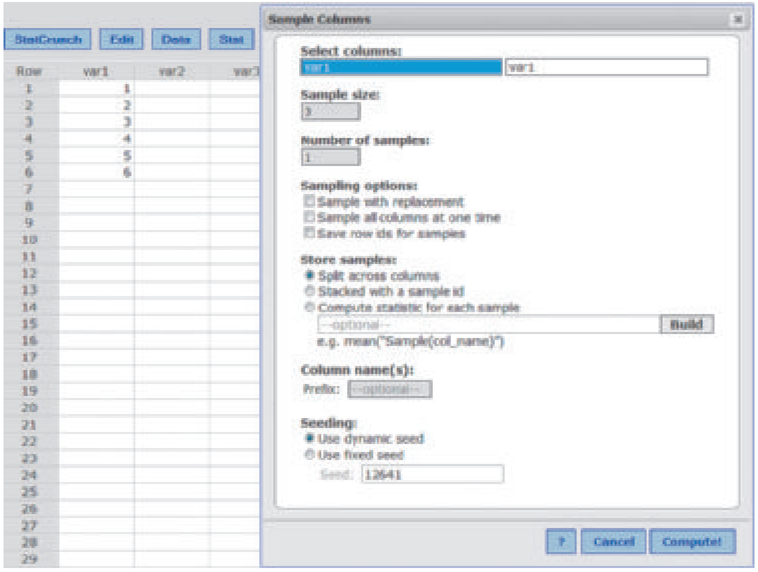
QUESTION Select an SRS of three names from these six names.

SOLUTION First assign each person a number, as shown:

Alberto	1
Justin	2
Michael	3
Audrey	4
Brandy	5
Nicole	6

Tech

Next, select three of these numbers without replacement. Figure 7.1 shows how this is done in StatCrunch, and almost all statistical technologies let you do this quite easily.



◀ **FIGURE 7.1** StatCrunch will randomly select, without replacement, three numbers from the six shown in the var1 column.

Using technology, we got these three numbers: 1, 2, and 6. These correspond to Alberto, Justin, and Nicole.

If technology is not available, a random number table, such as the one provided in Appendix A, can be used. Here are two lines from such a table:

7 7 5 9 8	2 9 5 1 1	9 8 1 4 9	6 3 9 9 1
3 1 9 4 2	0 4 6 8 4	6 9 3 6 9	5 0 8 1 4

You can start at any row or column you please. Here, we choose to start at the upper left (shown in bold face). Next, read off digits from left to right, skipping digits that are not in our population. Because no one has the number 7, skip this number, twice. The first person selected is number 5: Brandy. Then skip 9 and 8 and select number 2: Justin. Skip 9 and 5 (because you already selected Brandy) and select number 1: Alberto.

CONCLUSION Using technology, we got a sample consisting of Alberto, Justin, and Nicole. Using the random number table, we got a different sample: Brandy, Justin, and Alberto.

TRY THIS! Exercise 7.11

EXAMPLE 3 Survey on Sexual Harassment

A newspaper at a large college wants to determine whether sexual harassment is a problem on campus. The paper takes a simple random sample of 1000 students and asks each person whether he or she has been a victim of sexual harassment on campus. About 35% of those surveyed refuse to answer. Of those who do answer, 2% say they have been victims of sexual harassment.

QUESTION Give a reason why we should be cautious about using the 2% value as an estimate for the population percentage of those who have been victims of sexual harassment.

CONCLUSION There is a large percentage of students who did not respond. Those who did not respond might be different from those who did, and if their answers had been included, the results could have been quite different. When those surveyed refuse to respond, it can create a biased sample.

TRY THIS! Exercise 7.15

There are always some people who refuse to participate in a survey, but a good researcher will do everything possible to keep the percentage of nonresponders as small as possible, to reduce this source of bias.

SECTION 7.2

Measuring the Quality of a Survey

A frequent complaint about surveys is that a survey based on 1000 people can't possibly tell us what the entire country is thinking. This complaint raises interesting questions: How do we judge whether our estimators are working? What separates a good estimation method from a bad?

It's difficult, if not impossible, to judge whether any particular survey is good or bad. Sometimes we can find obvious sources of bias, but often we don't know whether a survey has failed unless we later learn the true parameter value. (This sometimes occurs in elections, when we learn that a survey must have had bias because it severely missed predicting the actual outcome.) Instead, statisticians evaluate the *method* used to estimate a parameter, not the outcome of a particular survey.

KEY POINT Statisticians evaluate the method used for a survey, not the outcome of a single survey.

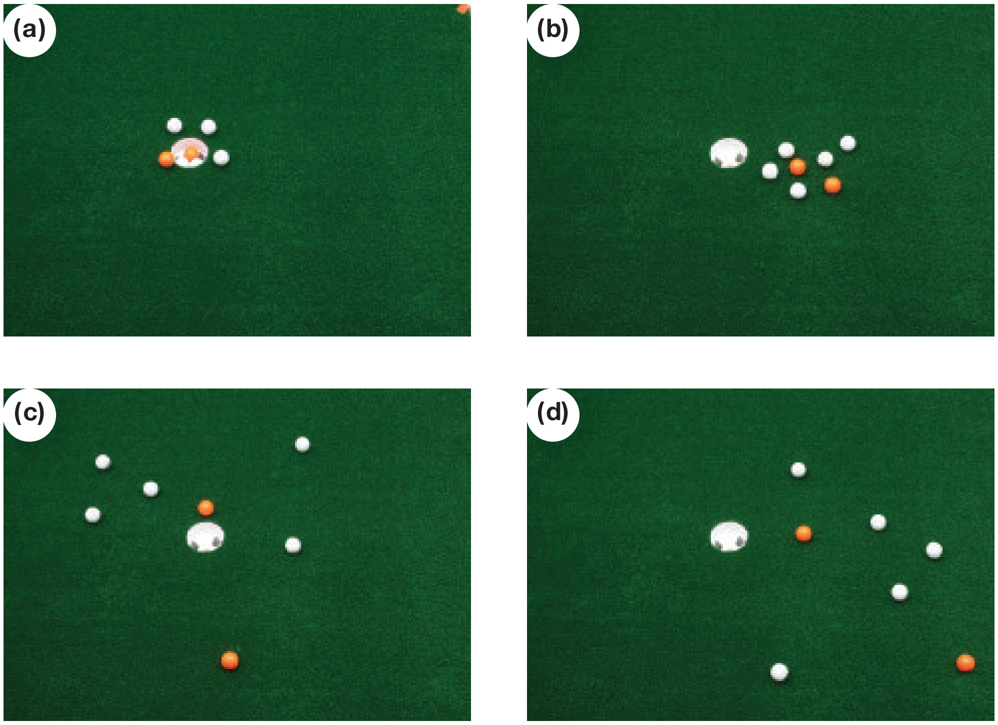
Before we talk about how to judge surveys, imagine the following scenario: We are not taking just one survey of 1000 randomly selected people. We are sending out an army of pollsters. Each pollster surveys a random sample of 1000 people, and they all use the same method for collecting the sample. Each pollster asks the same question and produces an estimate of the proportion of people in the population who would answer yes to the question. When the pollsters return home, we get to see not just a single estimate (as happens in real life) but a great many estimates. Because each estimate is based on a separate random collection of people, each one will differ slightly. We expect some of these estimates to be closer to the mark than others just because of random variation. What we really want to know is how the group did as a whole. For this reason, we talk about evaluating *estimation methods*, not estimates.

An estimation method is a lot like a golfer. To be a good golfer, we need to get the golf ball in the cup. A good golfer is both *accurate* (tends to hit the ball near the cup) and *precise* (even when she misses, she doesn't miss by very much.)

It is possible to be precise and yet be inaccurate, as shown in Figure 7.2b. Also, it is possible to aim in the right direction (be accurate) but be imprecise, as shown in Figure 7.2c. (Naturally, some of us are bad at both, as shown in Figure 7.2d.) But the best golfers can both aim in the right direction and manage to be very consistent, which Figure 7.2a shows us.

! Caution

Estimator and Estimates We often use the word **estimator** to mean the same thing as "estimation method." An **estimate**, on the other hand, is a number produced by our estimation method.



◀ **FIGURE 7.2** **(a)** Shots from a golfer with good aim and precision; the balls are tightly clustered and centered around the cup. **(b)** Shots from a golfer with good precision but poor aim; the balls are close together but centered to the right of the cup. **(c)** Shots from a golfer with good aim—the balls are centered around the cup—but bad precision. **(d)** The worst-case scenario: bad precision *and* bad aim.

Think of the cup as the population parameter, and think of each golf ball as an estimate, a value of $\hat{p}$, that results from a different survey. We want an estimation method that aims in the right direction. Such a method will, on average, get the correct value of the population parameter. We also need a precise method so that if we repeated the survey, we would arrive at nearly the same estimate.

The aim of our method, which the *accuracy,* is measured in terms of the *bias.* The *precision* is measured by a number called the *standard error.* Discussion of simulation studies in the next sections will help clarify how accuracy and precision are measured. These simulation studies show how bias and standard error are used to quantify the uncertainty in our inference.

Using Simulations to Understand the Behavior of Estimators

The three simulations that follow will help measure how well the sample proportion works as an estimator of the population proportion.

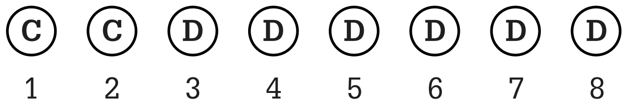
In the first simulation, imagine doing a survey of 4 people in a very small population with only 8 people. You'll see that the estimator of the population proportion is accurate (no bias) but, because of the small sample size, not terribly precise.

In the second simulation, the first simulation is repeated, using a larger population and sample. The estimator is still unbiased, and you will see a perhaps surprising change in precision. Finally, the third simulation will reveal that using a much larger sample size makes the result even more precise.

To learn how our estimation method behaves, we're going to create a very unusual, unrealistic situation: We're going to create a world in which we know the truth. In this world, there are two types of people: those who like dogs and those who like cats. No one likes both. Exactly 25% of the population are Cat People, and 75% are Dog People. We're going to take a random sample of people from this world and see what proportion of our sample are Cat People. Then we'll do it again. And again. From this repetition, we'll see some interesting patterns emerge.

Simulation 1: Statistics Vary from Sample to Sample

To get started, let's create a very small world. This world has 8 people named 1, 2, 3, 4, 5, 6, 7, and 8. People 1 and 2 are Cat People.



▶ **FIGURE 7.3** The entire population of our simulated world; 25% are Cat People.

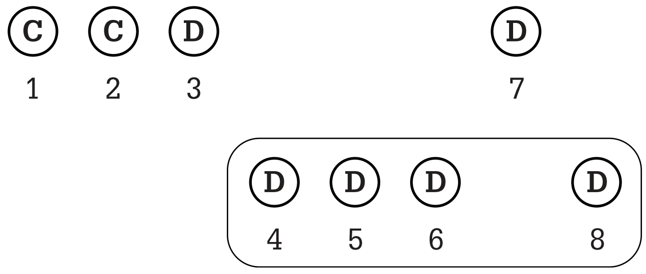
From this population, we use the random number table to generate four random numbers between 1 and 8. When a person's number is chosen, he or she steps out of the population and into our sample.

Before we tell who was selected, think for a moment about what you expect to happen. What proportion of our sample will be Cat People? Is it possible for 0% of the sample to be Cat People? For 100%?

Below is our random sample. Note that we sampled without replacement, as in a real survey. We don't want the same person to be in our sample twice.

6	8	4	5
D	D	D	D

None of those selected are Cat People, as Figure 7.4 indicates. The proportion of Cat People in our sample is 0%. We call this the *sample proportion* because it comes from the sample, not the population.



▶ **FIGURE 7.4** The first sample, which has 0% Cat People.

Let's take another random sample. It is possible that we will again get 0%, but it is also possible that we will get a different percentage.

7	2	6	3
D	C	D	D

This time, our sample proportion is 25%.

One more time:

2	8	6	5
C	D	D	D

Again, our sample proportion is 25%.

Table 7.3 shows what has happened so far. Even though we have done only three repetitions, we can make some interesting observations.

Repetition	Population Parameter	Sample Statistics
1	$p = 25\%$ Cat People	$\hat{p} = 0\%$ Cat People
2	$p = 25\%$ Cat People	$\hat{p} = 25\%$ Cat People
3	$p = 25\%$ Cat People	$\hat{p} = 25\%$ Cat People

◀ **TABLE 7.3** The results of three repetitions of our simulation.

First, notice that the population proportion, p, never changes. It can't, because in our made-up world, the population always has the same 8 people, and the same 2 are Cat People. However, the sample proportion, $\hat{p}$, can be different in each sample. In fact, $\hat{p}$ is random, because it depends on a random sample.

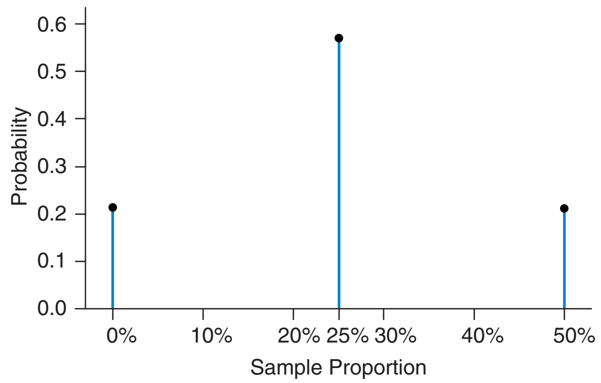
KEY POINT No matter how many different samples we take, the value of p (the population proportion) is always the same, but the value of $\hat{p}$ changes from sample to sample.

This simulation is, in fact, a random experiment and $\hat{p}$ is our outcome. Because it is random, $\hat{p}$ has a probability distribution. The probability distribution of $\hat{p}$ has a special name: **sampling distribution**. This term reminds us that $\hat{p}$ is not just any random outcome; it is a statistic we use to estimate a population parameter.

Because our world has only 8 people in it and we are taking samples of 4 people, we can write down all of the possible outcomes. There are only 70. By doing this, we can see exactly how often $\hat{p}$ will be 0%, how often 25%, and how often 50%. (Notice that it can never be more than 50%.) These probabilities are listed in Table 7.4, which presents the sampling distribution for $\hat{p}$. Figure 7.5 visually represents this sampling distribution.

Value of $\hat{p}$	Probability of Seeing That Value
0%	0.21429
25%	0.57143
50%	0.21429

▲ **TABLE 7.4** The sampling distribution for $\hat{p}$, based on our random sample.



◀ **FIGURE 7.5** Graphical representation of Table 7.4, the sampling distribution for $\hat{p}$ when p is 0.25.

From Table 7.4 and Figure 7.5, we learn several things:

1. Our estimator, $\hat{p}$, is not always the same as our parameter, p. Sometimes $\hat{p}$ turns out to be 0%, sometimes it is 50%, and sometimes it hits the target value of 25%.
2. The mean of this distribution is 25%—the same value as p.
3. Even though $\hat{p}$ is not always the "true" value, p, we are never more than 25 percentage points away from the true value.

Why are these observations important? Let's consider each one separately.

The first observation reminds us that statistics based on random samples are random. Thus we cannot know ahead of time, with certainty, exactly what estimates our survey will produce.

The second observation tells us that our estimator has no bias—that, *on average*, it is the same as the parameter value. **Bias** is measured as the distance between the mean value of the estimator (the center of the sampling distribution) and the population parameter. In this case, the center of the sampling distribution and the population parameter are both 0.25, so the distance is 0. In other words, there is no bias.

The third observation is about precision. We know that our estimator is, on average, the same as the parameter, but the sampling distribution tells us how far away, typically, the estimator might stray from average. **Precision** is reflected in the spread of the sampling distribution and is measured by using the standard deviation of the sampling distribution. In this simulation, the standard deviation is 0.16366, or roughly 16%. The standard deviation of a sampling distribution has a special name: the **standard error (SE)**.

The standard error measures how much our estimator typically varies from sample to sample. Thus, in the above example, if we survey 4 people, we usually get 25% Cat People, but this typically varies by plus or minus 16.4% (16.4 percentage points). Looking at the graph in Figure 7.5, we might think that the variability is typically plus or minus 25 percentage points, but we must remember that the standard deviation measures how spread out observations are from the average value. Many observations are identical to the average value, so the typical, or "standard," deviation from average is only 16.4 percentage points.

KEY POINT Bias is measured using the center of the sampling distribution: It is the distance between the center and the population value.

Precision is measured using the standard deviation of the sampling distribution, which is called the standard error. When the standard error is small, we say the estimator is precise.

SNAPSHOT SAMPLING DISTRIBUTION

WHAT IS IT? ► A special name for the probability distribution of a statistic.

WHAT DOES IT DO? ► Gives us probabilities for a statistic.

WHAT IS IT USED FOR? ► It tells us how often we can expect to see particular values of our estimator, and it also gives us important characteristics of the estimator, such as bias and precision.

HOW IS IT USED? ► It is used for making inferences about a population.

Simulation 2: The Size of the Population Does Not Affect Precision The first simulation was very simple, because our made-up world had only 8 people. In our first simulation, the bias was 0, which is good; this means we have an accurate estimator. However, the precision was fairly poor (we had a large standard error). How can we improve precision? To understand, we need a slightly more realistic simulation.

This time, we'll use the same world but make it somewhat bigger. Let's assume we have 1000 people and 25% are Cat People ($p = 0.25$). (In other words, there are 250 Cat People.) We take a random sample of 10 people and find the sample proportion, $\hat{p}$, of Cat People.

Because we've already seen how this is done, we're going to skip a few steps and show the results. This time the different outcomes are too numerous to list, so instead we just do a simulation:

1. Take a random sample, without replacement, of 10 people.
2. Calculate $\hat{p}$: the proportion of Cat People in our sample.
3. Repeat steps 1 and 2 a total of 10,000 times. Each time, calculate $\hat{p}$ and record its value.

Here are our predictions:

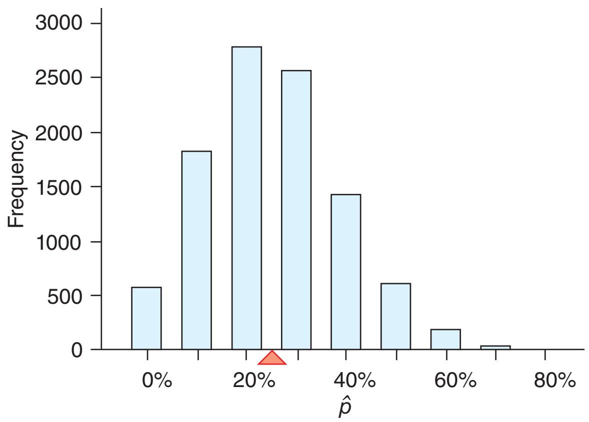
1. We predict that $\hat{p}$ will not be the same value every time because it is based on a random sample, so the value of $\hat{p}$ will vary randomly.
2. We predict that the mean outcome, the typical value for $\hat{p}$, will be 25%—the same as the population parameter—because our estimator is unbiased.
3. Precision: This one is left to you. Do you think the result will be more precise or less precise than in the last simulation? In the last simulation, only 4 people were sampled, and the variation, as measured by the standard error, was about 16%. This time more people (10) are being sampled, but the population is much larger (1000). Will the standard error be larger (less precise) or smaller (more precise) than 16%?

Tech

Details

Simulations and Technology
Don't take our word for it. You can probably carry out this simulation using technology. See the TechTips to learn how to do this using StatCrunch.

After carrying out the 10,000 simulations, we make a graph of our 10,000 $\hat{p}$'s. Figure 7.6 shows a histogram of these. Figure 7.6 is an approximation of the sampling distribution; it is not the actual sampling distribution, because the histogram is based on a simulation. Still, with 10,000 replications, it is a very good approximation of the actual sampling distribution.



◀ **FIGURE 7.6** Simulation results for $\hat{p}$. This histogram is a simulation of the sampling distribution. The true value of p is 25%. Each sample is based on 10 people, and we repeated the simulation 10,000 times.

The center of the estimated distribution is at 0.2501, which indicates that essentially no bias exists, because the population parameter is 0.25.

We can estimate the standard error by finding the standard deviation of our simulated $\hat{p}$'s. This turns out to be about 13.56%.

The value of the standard error tells us that if we were to take another sample of 10 people, we would expect to get about 25% Cat People, give or take 13.6 percentage points.

From Figure 7.6 we learn important information:

1. The bias of $\hat{p}$ is still 0, even though we used a larger population and a larger sample.
2. The variation of $\hat{p}$ is less; this estimator is more precise, even though the population is larger. In general, as long as the population is large relative to the sample size, the precision has *nothing* to do with the size of the *population*, but only with the size of the *sample*.

Many people are surprised to learn that precision is not affected by population size. How can the level of precision for a survey in a town of 10,000 people be the same as for one in a country of 210 *million* people?

Figure 7.7 provides an analogy. The bowls of soup represent two populations: a big one (a country, perhaps) and a small one (a city). Our goal is to taste each soup (take a sample from the population) to judge whether we like it. If both bowls are well stirred, the size of the bowl doesn't matter—using the same-size spoon, we can get the same amount of taste from either bowl.

▶ **FIGURE 7.7** The bowls of soup represent two populations, and the sample size is represented by the spoons. The precision of an estimate depends only on the size of the sample, not on the size of the population.

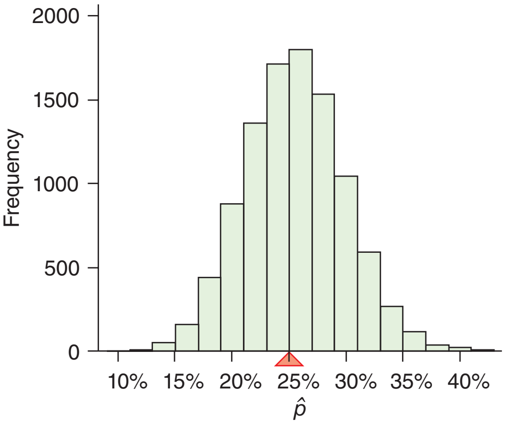
KEY POINT The precision of an estimator does not depend on the size of the population; it depends only on the sample size. An estimator based on a sample size of 10 is just as precise in a population of 1000 people as in a population of a million.

Simulation 3: Large Samples Produce More Precise Estimators

How do the simulation and bias change if we increase the sample size? We'll do another simulation with the same population (1000 people and 25% Cat People), but this time, instead of sampling 10 people, we'll sample 100.

Figure 7.8 shows the result. Note that the center of this estimated sampling distribution is still at 25%. Also, our estimation method remains unbiased. However, the shape looks pretty different. First, because many more outcomes are possible for $\hat{p}$, this histogram looks as though it belongs more to a continuous-valued random outcome than to a discrete value. Second, it is much more symmetric than Figure 7.6. You will see in Section 7.3 that the shape of the sampling distribution of $\hat{p}$ depends on the size of the random sample.

An important point to note is that this estimator is much more precise because it uses a larger sample size. By sampling more people, we get more information, so we can end up with a more precise estimate. The estimated standard error, which is simply the standard deviation of the data shown in Figure 7.8, is now 4.2 percentage points.



▶ **FIGURE 7.8** Simulated sampling distribution of a sample proportion of Cat People, based on a random sample of 100 people. The simulation was repeated 10,000 times.

Table 7.5 shows a summary of the three simulations.

Simulation	Population Size	Sample Size	Mean	Standard Error
1	8	4	25%	16.4%
2	1000	10	25%	13.5%
3	1000	100	25%	4.2%

Increasing Sample Size ↓ · Increasing Precision ↓

◀ **TABLE 7.5** Increasing sample size results in increasing precision (measured as decreasing standard error).

Here is what we learned from Figure 7.8, which is based on sample sizes of 100 "people":

1. The estimator $\hat{p}$ is unbiased for all sample sizes (as long as we take random samples).
2. The precision improves as the sample size gets larger.
3. The shape of the sampling distribution is more symmetric for larger sample sizes.

KEY POINT Surveys based on larger sample sizes have smaller standard error (SE) and therefore better precision. Increasing the sample size improves precision.

Finding the Bias and the Standard Error

We've shown how to estimate bias and precision by running a simulation. But we can also do this mathematically, without running a simulation. Bias and standard error are easy to find for a sample proportion under certain conditions.

The bias of $\hat{p}$ is 0, and the standard error is

$$\textbf{Formula 7.1a:}\quad SE = \sqrt{\frac{p(1-p)}{n}}$$

if the following two conditions are met:

Condition 1. The sample must be randomly selected from the population of interest, either with or without replacement. The population parameter to be estimated is the proportion of people (or objects) with some characteristic. This proportion is denoted as p.

Condition 2. If the sampling is without replacement, the population needs to be much larger than the sample size; at least 10 times bigger is a good rule of thumb.

EXAMPLE 4 Pet World

Suppose that in Pet World, the population is 1000 people and 25% of the population are Cat People. Cat People love cats but hate dogs. We are planning a survey in which we take a random sample of 100 people, without replacement. We calculate the proportion of people in our sample who are Cat People.

QUESTION What value should we expect for our sample proportion? What's the standard error? How do we interpret these values?

SOLUTION The sample proportion is unbiased, so we expect it to be the same as the population proportion: 25%.

The standard error is $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.25 \times 0.75}{100}} = \sqrt{\frac{0.1875}{100}}$

$= \sqrt{0.001875} = 0.04330$, or about 4.3%

This formula is appropriate because the population size is big with respect to the sample size. The population size is 1000, and the sample size is 100; $100 \times 10 = 1000$, so the population is ten times larger than the sample size.

CONCLUSION We interpret the values to mean that if we were to take a survey of 100 people from Pet World, we would expect about 25% of them to be Cat People, give or take about 4.3%. The "give or take" means that if you were to draw a sample of 100 and I were to draw a sample of 100, our sample proportions would typically differ from the expected 25% by about 4.3 percentage points.

TRY THIS! Exercise 7.25

Real Life: We Get Only One Chance

In simulations, we could repeat the survey many times to understand what might happen. In real life, we get just one chance. We take a sample, calculate $\hat{p}$, and then have to live with it.

It is important to realize that bias and precision are both measures of what would happen if we could repeat our survey many times. Bias indicates the typical outcome of surveys repeated again and again. If the bias is 0, we will typically get the right value. If the bias is 0.10, then our estimate will characteristically be 10 percentage points too high. Precision measures how much our estimator will vary from the typical value if we do the survey again. To put it slightly differently, if someone else does the survey, precision helps determine how different her or his estimate could be from ours.

How small must the standard error be for a "good" survey? The answer varies, but the basic rule is that the precision should be small enough to be useful. A typical election poll has a sample of roughly 1000 registered voters and a standard error of about 1.5 percentage points. If the candidates are many percentage points apart, this is good precision. However, if they are neck and neck, this might not be good enough. In Section 7.4, we will discuss how to make decisions about whether the standard error is small enough.

In real life, we don't know the true value of the population proportion, p. This means we can't calculate the standard error. However, we can come pretty close by using the sample proportion. If p is unknown, then

Formula 7.1b: $SE_{est} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, where SE_{est} is the estimated standard error

is a useful approximation to the true standard error.

SECTION 7.3

The Central Limit Theorem for Sample Proportions

Remember that a probability tells us how often an event happens if we repeat an experiment an infinite number of times. For instance, the sampling distribution of $\hat{p}$ gives the probabilities of where our sample proportions will fall; that is, it tells us how often we would see particular values of $\hat{p}$ if we could repeat our survey infinitely many times. In the simulation, we repeated our fake survey 10,000 times. Ten thousand is a lot, but it's a far cry from infinity.

In the three simulations in Section 7.2, we saw that the shape of the sampling distribution (or our estimated version, based on simulations) changed as the sample size increased (compare Figures 7.5, 7.6, and 7.8). If we used an even larger sample size than 100 (the sample size for the last simulation), what shape would the sampling distribution have? As it turns out, we don't need a simulation to tell us. For this statistic, and for some others, a mathematical theorem called the **Central Limit Theorem (CLT)** gives us a very good approximation of the sampling distribution without our needing to do simulations.

The Central Limit Theorem is helpful because sampling distributions are important. They are important because they, along with the bias and standard error, enable us to measure the quality of our estimation methods. Sampling distributions give us the probability that an estimate falls a specified distance from the population value. For example, we don't want to know simply that 18% of our customers are likely to buy new cell phones in the next year. We also want to know the probability that the true percentage might be higher than some particular value, say, 25%.

Meet the Central Limit Theorem for Sample Proportions

The Central Limit Theorem has several versions. The one that applies to estimating proportions in a population tells us that if some basic conditions are met, then the sampling distribution of the sample proportion is close to the Normal distribution.

More precisely, when estimating a population proportion, p, we must have the same conditions that were used in finding bias and precision, and one new condition as well:

Condition 1. *Random and Independent.* The sample is collected randomly from the population, and observations are independent of each other. The sample can be collected either with or without replacement.

Condition 2. *Large Sample.* The sample size, n, is large enough that the sample expects at least 10 successes (yes's) and 10 failures (no's).

Condition 3. *Big Population.* If the sample is collected without replacement, then the population size must be much (at least ten times) bigger than the sample size.

The sampling distribution for $\hat{p}$ is then approximately Normal, with mean p (the population proportion) and standard deviation the same as the standard error, as given in Formula 7.1a:

$$SE = \sqrt{\frac{p(1-p)}{n}}$$

KEY POINT The Central Limit Theorem for Sample Proportions tells us that if we take a random sample from a population, and if the sample size is large and the population size much larger than the sample size, then the sampling distribution of $\hat{p}$ is approximately

$$N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

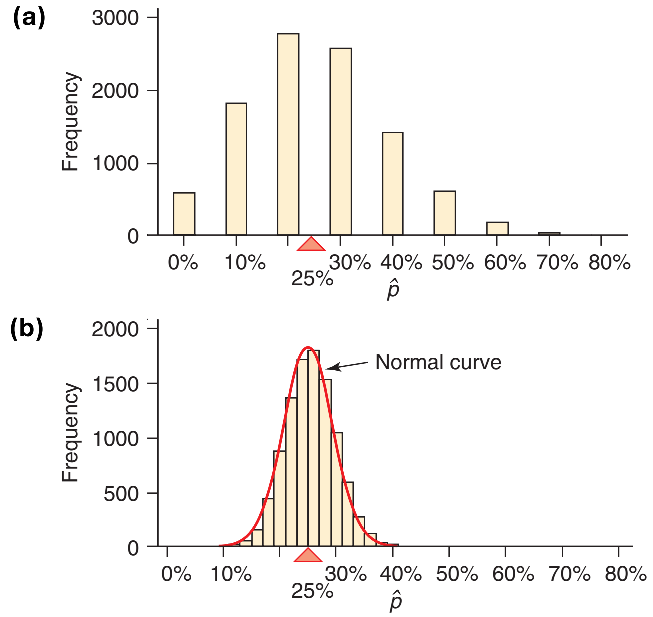
If you don't know the value of p, then you can substitute the value of $\hat{p}$ to calculate the estimated standard error.

Looking Back

Normal Notation
Recall that the notation N(mean, standard deviation) designates a particular Normal distribution.

Figure 7.9 illustrates the CLT for proportions. Figure 7.9a is based on simulations in which the sample size was just 10 people, which is too small for the CLT to apply. In this case, the simulated sampling distribution does not look Normal; it is right-skewed and has large gaps between values. Figure 7.9b is based on simulations of samples of 100 observations. Because the true population proportion is $p = 0.25$, a sample size of 100 is large enough for the CLT to apply, and our simulated sampling distribution looks very close to the Normal model. Figure 7.9b is actually a repeat of Figure 7.8 with the Normal curve superimposed. Now that the graphs' horizontal axes are on the same scale, we can see that the sample size of 100 gives better precision than the sample size of 10—the distribution is narrower.

▶ **FIGURE 7.9** **(a)** Revision of Figure 7.6, a histogram of 10,000 sample proportions, each based on $n = 10$ with a population percentage p equal to 25% **(b)** Revision of Figure 7.8, a histogram of 10,000 sample proportions, each based on $n = 100$ with a population percentage p equal to 25%.



The Normal curve shown in Figure 7.9b has a mean of 0.25 because $p = 0.25$, and it has a standard deviation (also called the standard error) of 0.0433 because

$$\sqrt{\frac{0.25 \times 0.75}{100}} = 0.0433$$

Before illustrating how to use the CLT, we show how to check conditions to see whether the CLT applies.

Checking Conditions for the Central Limit Theorem

The first condition requires that the sample be collected randomly and that observations be independent of each other. There is no way to check this just by looking at the data; you have to trust the researcher's report on how the data were collected, or, if you are the researcher, you must take care to use sound random sampling methods.

The second condition dictates that the sample size must be large enough. This we *can* check by looking at the data. The CLT says that the sample size needs to be sufficiently large to get at least 10 successes and 10 failures in our sample. If the probability of a success is p, then we would expect about np successes and $n(1 - p)$ failures. One problem, though, is that we usually don't know the value of p. In this case, we instead check that

$$n\hat{p} \geq 10 \quad \textbf{and} \quad n(1 - \hat{p}) \geq 10$$

For example, if our sample has 100 people and we are estimating the proportion of females in the population, and if our sample has 49% females, then we need to verify that both $100(0.49) \geq 10$ and $100(0.51) \geq 10$.

The third condition applies only to random samples done without replacement. In this case, the population must be at least 10 times bigger than the sample. In symbols, if N is the number of people in the population and n is the number in the sample, then

$$N \geq 10n$$

If this condition is not met, and the sample was collected without replacement, then the actual standard error will be a little smaller than what our formula says it should be.

In most real-life applications, the population size is much larger than the sample size. Over 300 million people live in the United States, so the typical survey of 1000 to 3000 easily meets this condition.

You can see how these conditions are used in the examples that follow.

KEY POINT The Central Limit Theorem for proportions requires (1) a random sample with independent observations; (2) a large sample; and (3) if SRS is used, a population with at least 10 times as many members as are in the sample.

Using the Central Limit Theorem

The following examples use the CLT to find the probability that the sample proportion will be near (or far from) the population value.

EXAMPLE 5 Pet World Revisited

Let's return to Pet World. The population is 1000 people, and the proportion of Cat People is 25%. We'll take a random sample of 100 people.

QUESTION What is the approximate probability that the proportion in our sample will be bigger than 29%? Begin by checking conditions for the CLT.

SOLUTION First we check conditions to see whether the Central Limit Theorem can be applied. The sample size is large enough because $np = 100(0.25) = 25$ is greater than 10, and $n(1 - p) = 100(0.75) = 75$, which is also greater than 10. Also, the population size is 10 times larger than the sample size, because $1000 = 10(100)$. Thus $N = 10(n)$; the population is just large enough. We are told that the sample was collected randomly.

According to the CLT, the sampling distribution will be approximately Normal. The mean is the same as the population proportion: $p = 0.25$. The standard deviation is the same as the standard error from Formula 7.1a:

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.25 \times 0.75}{100}} = \sqrt{\frac{0.1875}{100}} = \sqrt{0.001875} = 0.0433$$

We can use technology to find the probability of getting a value larger than 0.29 in a $N(0.25, 0.0433)$ distribution. Or we can standardize.

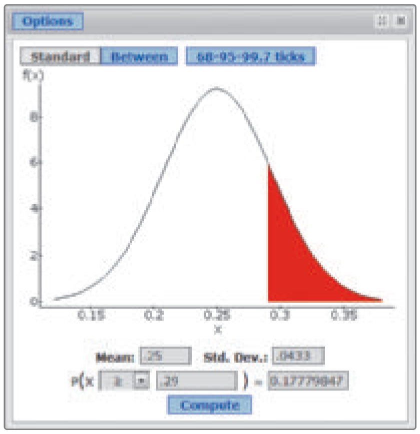
In standard units, 0.29 is

$$z = \frac{0.29 - 0.25}{0.0433} = 0.924 \text{ standard unit}$$

In a $N(0,1)$ distribution, the probability of getting a number bigger than 0.924 is, from Table A in the appendix, about 0.18, or 18%. Figure 7.10 on the next page shows the results using technology.

▶ **FIGURE 7.10** Output from StatCrunch. There is about an 18% chance that $\hat{p}$ will be more than 4 percentage points above 25%.

Tech



CONCLUSION With a sample size of 100, there is about an 18% chance that $\hat{p}$ will be more than 4 percentage points above 25%.

TRY THIS! Exercise 7.35

SNAPSHOT THE SAMPLE PROPORTION: $\hat{p}$ (p-HAT)

WHAT IS IT? ▶ The proportion of people or objects in a sample that have a particular characteristic in which we are interested.

WHAT IT IS USED FOR? ▶ To estimate the proportion of people or objects in a population that have that characteristic.

WHY DO WE USE IT? ▶ If the sample is drawn at random from the population, then the sample proportion is unbiased and has standard error $\sqrt{\dfrac{p(1-p)}{n}}$.

HOW IS IT USED? ▶ If, in addition to everything above, the sample size is fairly large, then we can use the Normal distribution to find probabilities concerning the sample proportion.

EXAMPLE 6 Presidential Election Survey

In a hotly contested U.S. election, two candidates for president, a Democrat and a Republican, are running neck and neck; each candidate has 50% of the vote. Suppose a random sample of 1000 voters are asked whether they will vote for the Republican candidate.

QUESTIONS What percentage of the sample should be expected to express support for the Republican? What is the standard error for this sample proportion? Does the Central Limit Theorem apply? If so, what is the approximate probability that the sample proportion will fall within two standard errors of the population value of $p = 0.50$?

SOLUTION Because we have collected a random sample, the sample proportion has no bias (assuming there are no problems collecting the sample). Therefore, we expect that 50% of our sample supports the Republican candidate.

Because the sample size, $n = 1000$, is small relative to the population (which is over 100 million), we can calculate the standard error with

$$SE = \sqrt{\frac{(0.50)(0.50)}{1000}} = 0.0158$$

We can interpret this to mean that we expect our sample proportion to be 50%, give or take 1.58 percentage points.

Because the sample size is fairly large (the expected numbers for successes and failures are both equal to $np = 1000 \times 0.50 = 500$, which is larger than 10), the CLT tells us we can use the Normal distribution—in particular, $N(0.50, 0.0158)$.

We are asked to find the probability that the sample proportion will fall within two standard errors of 0.50. In other words, that it will fall somewhere between

$$0.50 - 2SE$$

and

$$0.50 + 2SE$$

Because this is a Normal distribution, we know the probability will be very close to 95% (according to the Empirical Rule). But let's calculate the result anyway.

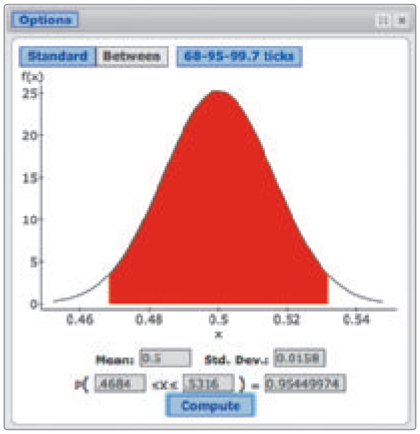
$$0.50 - 2SE = 0.50 - 2(0.0158) = 0.50 - 0.0316 = 0.4684$$

$$0.50 + 2SE = 0.50 + 0.0316 = 0.5316$$

That is, we want to find the area between 0.4684 and 0.5316 in a $N(0.5, 0.0158)$ distribution. Figure 7.11 shows the result using technology, which tells us this probability is 0.9545.

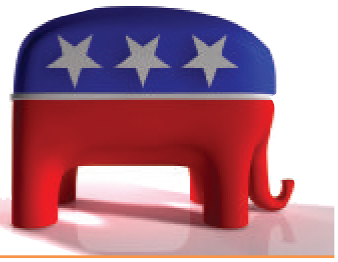
Looking Back

Empirical Rule
Recall that the Empirical Rule says that roughly 68% of observations should be within one standard deviation of the mean, about 95% within two standard deviations of the mean, and nearly all within three standard deviations of the mean. In this context, the standard error is the standard deviation for the sampling distribution.



◀ **FIGURE 7.11** The probability that a sample proportion based on a random sample of 1000 people taken from a population in which $p = 0.50$ has about a 95% chance of falling within two standard errors of 0.50.

CONCLUSION If each candidate truly has 50% of the vote, then we'd expect our sample proportion to be about 0.50 (or 50%). There is about a 95% chance that the sample proportion falls within two standard errors of 50%.



TRY THIS! Exercise 7.37

The conclusion from Example 6 is useful because it implies that, in general, we can predict where $\hat{p}$ will fall, relative to p. It indicates that $\hat{p}$ is very likely to fall within two standard errors of the true value, as long as the sample size is large enough. If, in addition, we have a small standard error, we know that $\hat{p}$ is quite likely to fall close to p.

KEY POINT If the conditions of a survey sample satisfy those required by the CLT, then the probability that a sample proportion will fall within two standard errors of the population value is 95%.

EXAMPLE 7 Morse and the Proportion of E's

Samuel Morse (1791–1872), the inventor of Morse code, claimed that the letter used most frequently in the English language was E and that the proportion of E's was 0.12. Morse code translates each letter of the alphabet into a combination of "dots" and "dashes," and it was used by telegraph operators, before the days of radio or telephones, to transmit messages around the world. It was important that the most frequently used letters be the easiest for the telegraph operator to type. In Morse code, the letter E is simply "dot."

To check whether Morse was correct about the proportion of E's, we took a simple random sample with replacement from a modern-day book. Our sample consisted of 876 letters, and we found 118 E's, so $\hat{p} = 0.1347$.

QUESTION Assume that the true proportion of E's in the population is, as Morse claimed, 0.12. Find the probability that, if we were to take another random sample of 876 letters, the sample proportion would be greater than or equal to 0.1347. As a first step, check that the Central Limit Theorem can be applied in this case.

SOLUTION To check whether we can apply the Central Limit Theorem, we need to make sure the sample size is large enough. Because $p = 0.12$, we check

$$np = 876(0.12) = 105.12, \text{ which is larger than 10}$$

and

$$n(1 - p) = 876(0.88) = 770.88, \text{ which is also larger than 10.}$$

The book contains far more than 8760 letters, so the population size is much larger than the sample size.

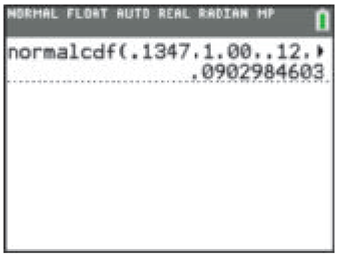
We can therefore use the Normal model for the distribution of sample proportions. The mean of this distribution is

$$p = 0.12$$

The standard error is

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.12(0.880)}{876}} = 0.010979$$

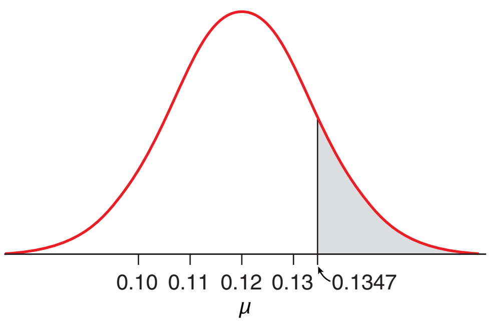
$$z = \frac{\hat{p} - p}{SE} = \frac{0.1347 - 0.12}{0.010979} = \frac{0.0147}{0.010979} = 1.339$$



▲ **FIGURE 7.12** TI-84 output

We therefore need to determine the probability of getting a z-score of 1.339 or larger. We can find this with the Normal table; it is the area to the right of a z-score of 1.34. We can also use technology (Figure 7.12) to find the area to the right of 0.1347 in a $N(0.12, 0.012)$ distribution. This probability is represented by the shaded area in Figure 7.13.

▶ **FIGURE 7.13** The shaded area represents the probability of finding a sample proportion of 0.1347 or larger from a population with a proportion of 0.12.



CONCLUSION If the sample is 876 letters, the probability of getting a sample proportion of 0.1347 or larger, when the true proportion of E's in the population is 0.12, is about 9%.

TRY THIS! Exercise 7.39

SECTION 7.4

Estimating the Population Proportion with Confidence Intervals

An example of a real survey illustrates this situation. The Pew Research Center took a random sample of 446 registered Democrats in the United States in 2013. In this sample, 57% of the 446 people agreed with the statement that the news media spent too much time on unimportant stories. (Pew also asked the same question of Republicans and Independents.) However, this percentage just tells us about our sample. What percentage of the population—that is, what percentage of *all* Democrats in the United States—agree with this statement? How much larger or smaller than 57% might the percentage who agree be? Can we conclude that a majority (more than 50%) of Americans share this belief?

We don't know p, the population parameter. We do know $\hat{p}$ for this sample; it is equal to 57%. Here's what else we know from the preceding sections:

1. Our estimator is unbiased, so even though our estimate of 57% may not be exactly equal to the population parameter, it's probably just a little higher or just a little lower.
2. The standard error can be estimated as

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.57 \times 0.43}{446}} = 0.023, \text{ or about } 2.3\%$$

   This tells us that the population proportion might not be very close to the value we saw, since a standard of error of 2.3 percentage points indicates a relatively imprecise estimator.
3. Because the sample size is large, we also realize that the probability distribution of $\hat{p}$ is pretty close to being Normally distributed and is centered around the true population parameter value. Thus, there's about a 68% chance that $\hat{p}$ is closer than one standard error away from the population proportion, and a 95% chance that it is closer than two standard errors away. (See Example 6.) Also, there is almost a 100% chance (99.7%, actually) that the sample proportion is closer than three standard errors from the population proportion. Thus we can feel very confident that the proportion of the population who agree with this statement is within three standard errors of 0.57. Three standard errors is 3(2.3%) = 6.9%, so we can be almost certain that the value of the population parameter is within 6.9 percentage points of 57%.

In other words, we can be highly confident that the population parameter is between these two numbers:

$$\begin{array}{ccc} 57\% - 6.9\% & \text{to} & 57\% + 6.9\%, \text{ or} \\ 50.1\% & \text{to} & 63.9\% \end{array}$$

We have just calculated a **confidence interval**. Confidence intervals are often reported as the estimate plus or minus some amount:

$$57\% \text{ plus or minus } 6.9\%, \text{ or } 57\% \pm 6.9\%.$$

The "some amount," in this case the 6.9 percentage points, is called the **margin of error**. The margin of error tells how far from the population value our estimate can be.

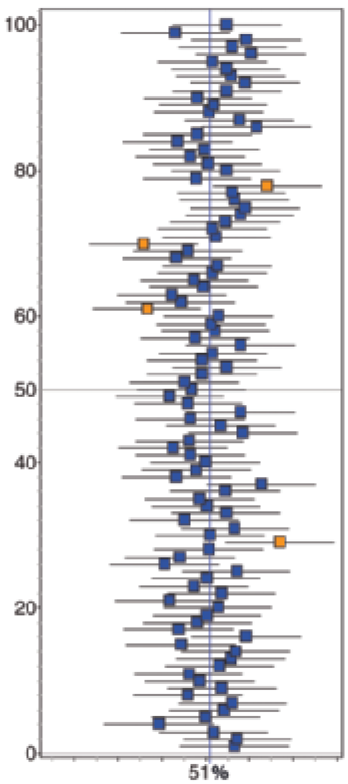
A confidence interval provides two pieces of information: (1) a range of plausible values for our population parameter (50.1% to 63.9%), and (2) a **confidence level**, which expresses (no surprise here) our level of confidence in this interval. Our high confidence level of 99.7% assures us we can be very confident that a majority of Democrats agree that the news media spend too much time on unimportant stories, because the smallest plausible level of agreement in the population is 50.1%, which is (just) bigger than a majority.

An analogy can help explain confidence intervals. Imagine a city park. In this park sit a mother and her daughter, a toddler. The mother sits in the same place every day, on a bench along a walkway, while her daughter wanders here and there. Most of the time, the child stays very close to her mother, as you would expect. In fact, our studies have revealed that 68% of the days we've looked, she is within 1 yard of her mother. Sometimes she strays a little bit farther, but on 95% of the days she is still within 2 yards of her mother. Only rarely does she move much farther; she is almost always within 3 yards of her mother.

One day the unimaginable happens, and the mother and the park bench become invisible. Fortunately, the child remains visible. The problem is to figure out where the mother is sitting.

Where is the mother? On 68% of the days, the child is within 1 yard of the mother, so at these times the mother must be within 1 yard of the child. If we think the mother is within 1 yard of the child on most days—that is, 68% of the days we observe—we will be right. But this also means we will be wrong on 32% of our visits. We could be more confident of being correct if we instead guessed that the mother is within 2 yards of the child. Then we would be wrong on only 5% of the days.

In this analogy, the mother is the population proportion. Like the mother, the population proportion never moves from its spot and never changes values. And just as we cannot see the invisible mother, we don't know where the parameter sits. The toddler is like our sample proportion, $\hat{p}$; we *do* know its value, and we know that it hangs out near the population proportion and moves around from sample to sample. Thus, even though we can't know exactly what the true population proportion is, we can infer that it is near the sample proportion.



▲ **FIGURE 7.14** Results from 100 simulations in which we draw a random sample and then find and display a confidence interval with a 95% confidence level. The orange squares indicate "bad" intervals.

Setting the Confidence Level

The confidence level tells us how often the estimation method is successful. Our method is to take a random sample and calculate a confidence interval to estimate the population proportion. If the method has a 100% confidence level, that method always works. If the method has a 10% confidence level, it works in 10% of surveys. We say the method works if the interval captures the true value of the population parameter. In this case, the interval works if the true population proportion is inside the interval.

Think of the confidence level as the capture rate; it tells us how often a confidence interval based on a random sample will capture the population proportion. Keep in mind that the population proportion, like the mother on the park bench, does not move—it is always the same. However, the confidence interval does change with every random sample collected. Thus, the confidence level measures the success rate of the *method*, not of any one particular interval.

KEY POINT The confidence level measures the capture rate for our method of finding confidence intervals.

Figure 7.14 demonstrates what we mean by a 95% confidence level. Let's suppose that in the United States, 51% of all voters favor stricter laws with respect to buying and selling guns. We simulate taking a random sample of 1000 people. We calculate the percentage of the sample who favor stricter laws, and then we find the confidence interval that gives us a 95% confidence level. We do this again and keep repeating. Figure 7.14 shows 100 simulations.

Each blue point and each orange point represent a sample percentage. Note that the points are centered around the population percentage of 51%. The horizontal lines represent the confidence interval: the sample percentage plus or minus the margin of error. The margin of error was chosen so that the confidence level is 95%. Notice that most of the lines cross the vertical line at 51%. These are successful confidence intervals that capture the population value of 51% (in blue). However, a few sample percentages miss the mark; these are indicated by orange points. In 100 trials, our method failed 4 times and was successful 96 times. In other words, it worked in 96% of these trials. When we use a 95% confidence level, our method works in 95% of all surveys we conduct.

We can change the confidence level by changing the margin of error. The greater the margin of error, the higher our confidence level. For example, we can be 100% confident that the true percentage of Americans who favor stricter gun laws is between 0% and 100%. We're 100% confident in this interval because it can never be wrong. Of course, it will also never be useful. We really don't need to spend money on a survey to learn that the answer lies between 0% and 100%, do we?

It would be more helpful—more precise—to have a smaller margin of error than "plus or minus 50 percentage points." However, if the margin of error is too small, then we are more likely to be wrong. Think of the margin of error as a tennis racket. The bigger the racket, the more confident you are of hitting the ball. Choosing an interval that ranges from 0% to 100% is like using a racket that fills the entire court—you will definitely hit the ball, but not because you are a good tennis player. If the racket is too small, you are less confident of hitting the ball, so you don't want it too small. Somewhere between too big and too small is just right.

Selecting a Margin of Error

We select a margin of error that will produce the desired confidence level. For instance, how can we choose a margin of error with a confidence level of 95%? We already know that if we take a large enough random sample and find the sample proportion, then the CLT tells us that 95% of the time, the sample proportion is within two standard errors of the population proportion. This is what we learned from Example 6. It stands to reason, then, that if we choose a margin of error that is two standard errors, then we'll cover the population proportion in 95% of our samples.

This means that

$$\hat{p} \pm 2SE$$

is a confidence interval with a 95% confidence level. More succinctly, we call this a 95% confidence interval.

Using the same logic, we understand that the interval

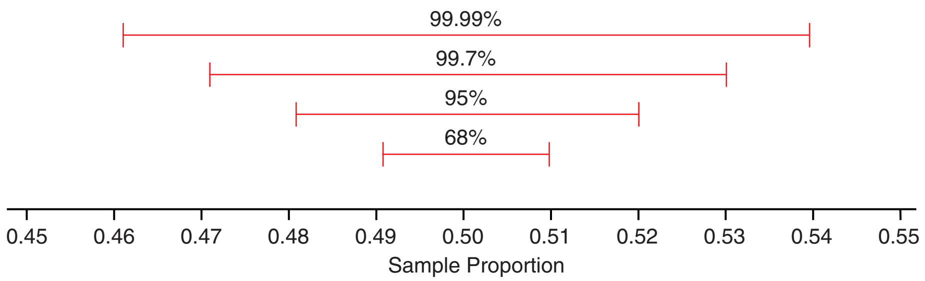
$$\hat{p} \pm 1SE$$

is a 68% confidence interval and that

$$\hat{p} \pm 3SE$$

is a 99.7% confidence interval.

Figure 7.15 shows four different margins of error for a sample in which $\hat{p} = 50\%$.



◀ **FIGURE 7.15** Four confidence intervals with confidence levels ranging from 99.99% (plus or minus 4 standard errors—top) to 68% (plus or minus 1 standard error). Notice how the interval gets wider with increasing confidence level.

This figure illustrates one reason why a 95% confidence interval is so desirable. If we increase the margin of error from 2 standard errors to 3, we gain only a small amount of confidence; the level goes from 95% to 99.7%. However, if we decrease from 2 standard errors to 1, we lose a lot of confidence; the level falls from 95% to 68%. Thus, the choice of 2 standard errors is very economical.

The margin of error has this structure:

$$\text{Margin of error} = z^{*}SE$$

where z^{*} is a number that tells how many standard errors to include in the margin of error. If $z^{*} = 1$, the confidence level is 68%. If $z^{*} = 2$, the confidence level is 95%. Table 7.6 summarizes the margin of error for four commonly used confidence levels.

▶ **TABLE 7.6** We can set the confidence level to the value we wish by choosing the appropriate margin of error.

Confidence Level	Margin of Error Is . . .
99%	2.58 standard errors
95%	1.96 (about 2) standard errors
90%	1.645 standard errors
80%	1.28 standard errors

Reality Check: Finding a Confidence Interval When *p* Is Not Known

As we have seen, a confidence interval for a population proportion has this structure:

$$\hat{p} \pm m$$

where m is the margin of error. Substituting for the margin of error, we can also write

$$\hat{p} \pm z^{*}SE$$

Finding the standard error requires us to know the value of p:

$$SE = \sqrt{\frac{p(1-p)}{n}}$$

However, in real life, we don't know p. So instead, we substitute our sample proportion and use Formula 7.1b for the estimated standard error:

$$SE_{\text{est}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

The result is a confidence interval with a confidence level close to, but not exactly equal to, the correct level. This tends to be close enough for most practical purposes.

In real life, then, Formula 7.2 is the method we use to find approximate confidence intervals for a population proportion.

Formula 7.2: $\hat{p} \pm m$, where $m = z^{*}SE_{\text{est}}$ and $SE_{\text{est}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

where:

m is the margin of error
$\hat{p}$ is the sample proportion of successes, or the proportion of people in the sample with the characteristic we are interested in
n is the sample size
z^{*} is a multiplier that is chosen to achieve the desired confidence level (Table 7.6)
SE_{est} is the estimated standard error

EXAMPLE 8 Ghostly Polls

Is it possible that more than 25% of Americans believe they have seen a ghost? A Pew Poll conducted in 2013 surveyed a random sample of 2003 adult Americans, and 18% of them said that they had seen a ghost.

QUESTION Estimate the standard error. Find an approximate 95% confidence interval for the percentage of all Americans who believe they have seen a ghost. Is it plausible to conclude that 25% or more Americans believe they have seen a ghost?

SOLUTION We first make sure the conditions of the Central Limit Theorem apply. We are told that the Pew Poll took a random sample. We must assume their observations were independent. We don't know whether the pollsters sampled with or without replacement, but because the population is very large—easily 10 times larger than the sample size—we don't need to worry about the replacement issue. (This confirms that conditions 1 and 3 apply.)

Next, we need to check that the sample size is large enough for us to use the CLT. We do not know p, the proportion of all Americans who say they've seen a ghost. We know only $\hat{p}$ (equal to 0.18), which the Pew Poll found on the basis of its sample. This means that our sample has at least 10 successes (people who believe they have seen a ghost) because $2003(0.18) = 360.5$, which is much larger than 10. Also, we know we have at least 10 failures (people who don't believe they have seen a ghost), because $2003(1 - 0.18)$ is even bigger than 360.5.

At this point, we can go directly to technology, such as StatCrunch or Minitab, or we can continue to compute using a calculator. Figure 7.16 shows the output from StatCrunch. If we use a calculator, the next step is to estimate the standard error. Using Formula 7.1b yields

$$SE_{\text{est}} = \sqrt{\frac{0.18(1 - 0.18)}{2003}} = 0.00858$$

We now use this result together with Formula 7.2 (using 2, rather than the slightly more accurate value 1.96, for our multiplier) to find the interval.

$$\hat{p} \pm 2SE_{\text{est}}$$
$$0.18 \pm 2(0.00858)$$
$$0.18 \pm 0.01716$$

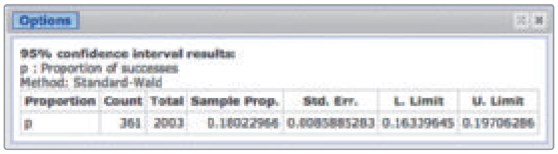
or, if you prefer,

$$18\% \pm 1.7\%$$

Expressing this as an interval, we get

$$18\% - 1.7\% = 16.3\%$$
$$18\% + 1.7\% = 19.7\%$$

The 95% confidence interval is 16.3% to 19.7%.



Options

95% confidence interval results:
p : Proportion of successes
Method: Standard-Wald

Proportion	Count	Total	Sample Prop.	Std. Err.	L. Limit	U. Limit
p	361	2003	0.18022966	0.0085885283	0.16339645	0.19706286

◀ **FIGURE 7.16** StatCrunch output of a 95% confidence interval for the proportion of all Americans who would say that they have seen a ghost.

CONCLUSION The confidence interval tells us which values are plausible for the population percentage. We have to conclude that it is not plausible that more than one-fourth (25%) of Americans say that they have seen a ghost, because the interval does not include 25%. The largest plausible value is 19.7%, which is very close to 1 in 5 Americans, but not 1 in 4.

TRY THIS! Exercise 7.51

Interpreting Confidence Intervals

A confidence interval for a sample proportion gives a set of values that are plausible for the population proportion. If a value is not in the confidence interval, we conclude that it is implausible. It's not impossible that the population value is outside the interval, but it would be pretty surprising.

Suppose a candidate for political office conducts a poll and finds that a 95% confidence interval for the proportion of voters who will vote for him is 42% to 48%. He would be wise to conclude that he does *not* have 50% of the population voting for him. The reason is that the value 50% is not in the confidence interval, so it is implausible to believe that the population value is 50%.

There are many common misinterpretations of confidence intervals that you must avoid. The most common mistake that students (and, indeed, many others) make is trying to turn confidence intervals into some sort of probability problem. For example, if asked to interpret a 95% confidence interval of 45.9% to 53.1%, many people would mistakenly say, "This means there is a 95% chance that the population proportion is between 45.9% and 53.1%."

What's wrong with this statement? Remember that probabilities are long-run frequencies. This sentence claims that if we were to repeat this survey many times, then in 95% of the surveys the true population percentages would be a number between 45.9% and 53.1%. This claim is wrong, because the true population percentage doesn't change. Either it is *always* between 45.9% and 53.1% or it is *never* between these two values. It can't be between these two numbers 95% of the time and somewhere else the rest of the time. In our story about the invisible mother, the mother, who represented the population proportion, *always* sat at the same place. Similarly, the population proportion (or percentage) is always the same value.

Another analogy will help make this clear. Suppose there is a skateboard factory. Say 95% of the skateboards produced by this factory are perfect, but 5% have no wheels. Once you buy a skateboard from this factory, you can't say that there is a 95% chance that it is a good board. Either it has wheels or it does not have wheels. It is *not* true that the board has wheels 95% of the time and, mysteriously, no wheels the other 5% of the time. A confidence interval is like one of these skateboards. Either it contains the true parameter (has wheels) or it does not. The "95% confidence" refers to the "factory" that "manufactures" confidence intervals: 95% of its products are good, and 5% are bad.

Our confidence is in the process, not in the product.

KEY POINT Our confidence is in the process that produces confidence intervals, not in any particular interval. It is incorrect to say that a particular confidence interval has a 95% (or any other percent) chance of including the true population parameter. Instead, we say that the *process* that produces intervals captures the true population parameter with a 95% probability.

EXAMPLE 9 Underwater Mortgages

A mortgage is "underwater" if the amount owed is greater than the value of the property that is mortgaged. A 2013 Rasmussen Poll of 715 homeowners in the United States found that 62% of them believed their homes were *not* underwater—the highest percentage recorded since 2009. Rasmussen reports that "The margin of sampling error is plus or minus 4 percentage points with a 95% level of confidence."

QUESTION State the confidence interval in interval form. How would you interpret this confidence interval? What does "95%" mean?

CONCLUSION The margin of error, we are told, is 4 percentage points. In interval form, then, the 95% confidence interval is

$$62\% - 4\% \quad \text{to} \quad 62\% + 4\%, \text{ or}$$
$$58\% \quad \text{to} \quad 66\%$$

We interpret this to mean that we are 95% confident that the true proportion of all U.S. homeowners who believe their homes are not underwater is between 58% and 66%. The 95% indicates that if we were to conduct not just this survey, but many, then 95% of them would result in confidence intervals that include the true population proportion.

TRY THIS! Exercise 7.57

Example 10 demonstrates the use of confidence intervals to make decisions about population proportions.

EXAMPLE 10 Morse and E's

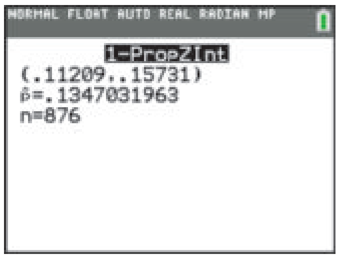
Recall from Example 7 that Morse believed the proportion of E's in the English language was 0.12 and that our sample showed 118 E's out of 876 randomly chosen letters from a modern-day book.

QUESTION Find a 95% confidence interval for the proportion of E's in the book. Is the proportion of E's in the book consistent with Morse's 0.12? Assume the conditions that allow us to interpret the confidence interval are satisfied. (The conditions were checked in Example 7.)

SOLUTION The best approach is to use technology. Figure 7.17 shows TI-84 output that gives a 95% confidence interval as

$$(0.112, 0.157) \quad \text{or} \quad (11.2\%, 15.7\%)$$

Tech



▲ **FIGURE 7.17** TI-84 output for a confidence interval for the proportion of E's.

If you do not have access to statistical technology, then the first step is to find the sample proportion of E's: 118/876, or 0.1347.

The estimated standard error is

$$SE_{\text{est}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.1347(0.8653)}{876}} = \sqrt{0.00013305} = 0.0115349$$

Because we want a 95% confidence level, our margin of error is plus or minus 1.96 standard errors:

$$\text{Margin of error} = 1.96SE_{\text{est}} = 1.96(0.0115349) = 0.022608$$

The interval boundaries are

$$\hat{p} \pm 1.96SE_{\text{est}} = 0.1347 \pm 0.0226$$

Upper end of interval: $0.1347 + 0.0226 = 0.1573$
Lower end of interval: $0.1347 - 0.0226 = 0.1121$

This confirms the result we got through technology: A 95% confidence interval is (0.1121 to 0.1573). Note that this interval *does* include the value 0.12.

CONCLUSION We are 95% confident that the proportion of E's in the modern book is between 0.112 and 0.157. This interval captures 0.12. Thus it is plausible that the population proportion of E's in the book is 0.12, as Morse suggested.

TRY THIS! Exercise 7.61

SECTION 7.5

Comparing Two Population Proportions with Confidence

People change their minds for a number of reasons. And policy makers, particularly in a democracy, like to keep updated on what people are thinking. Embryonic stem cell research is one area that has generated controversy over time. Embryonic stem cell research shows great promise in treating a number of serious diseases, but it violates many people's moral beliefs because it involves using cells derived from human embryos.

In 2002, a Pew Poll based on a random sample of 1500 people suggested that 43% of the American public approved of stem cell research. In 2009, a new poll of a different sample of 1500 people found that 58% approved.

Did American opinion really change? Perhaps. But it is also possible that these two sample proportions are different because the samples used different people. The people were randomly selected, but we know that random samples can vary. Quite possibly, the sample proportions differed just by chance. Although the sample proportions are different, the *population* proportions might be the same.

What's the Difference?

What's the difference between Coke and Pepsi? When asked a question like this, you probably think of qualitative characteristics: flavor, color, bubbliness. But when we ask about the difference between two *numbers*, we mean "How far apart are the two numbers?"

The answer to the question "How far apart are two numbers?" is found by subtracting. How far apart are the sample percentages 58% and 43%?

$$58\% - 43\% = 15\%$$

The two sample percentages are 15 percentage points apart.

Much of our analysis in comparing two samples is based on subtraction. In this section, and in Section 8.4, our comparison of two population proportions will be based on the statistic

$$\hat{p}_1 - \hat{p}_2$$

This statistic will be used to estimate the difference between two population proportions:

$$p_1 - p_2$$

Looking Back

Statistics

A statistic, as you learned in Section 7.1, is a number that is based on data and used to estimate a population parameter.

It might seem strange to you that we would have to go to so much trouble to determine whether two numbers are different from each other. After all, can't we just look and *see* that 0.43 does not equal 0.58?

This issue is subtle but important. Even when two proportions are equal in the *population,* their *sample* proportions can be different. This difference is caused by the fact that we see only *samples* of the populations and not the entire populations. This means that even if 23% of all men and 21% of all women believe that embryonic stem cell research is wrong, a random sample of men and women might have different percentages of believers, perhaps 22% and 28%.

KEY POINT Even if the proportions are equal for two populations, the sample proportions drawn from these populations are usually different.

Confidence intervals are one method for determining whether different sample proportions reflect "real" differences in the populations. The basic approach is this:

> First, we find a confidence interval, at the significance level we think best, for the difference in proportions $p_1 - p_2$.

Next, we check to see whether that interval includes 0. If it does, then this suggests that the two population proportions might be the same. Why? Because if $p_1 - p_2 = 0$, then $p_1 = p_2$ and the proportions are the same.

If the confidence interval does not contain 0, we also learn interesting things. As you will soon see, the confidence interval tells us how much greater one of the proportions might be than the other.

Example 11 Do Men's and Women's Views Differ?

In a random sample of roughly equal numbers of U.S. men and women, the Pew Foundation found, in 2013, that 23% of the men in their sample believed that research using embryonic stem cells is "morally wrong." For the women, 21% believed it is morally wrong.

QUESTION Can we conclude, on the basis of these sample percentages only, that in the United States, a greater percentage of men than of women believe that embryonic stem cell research is morally wrong? Explain.

SOLUTION No, we cannot conclude this. Although a greater proportion of *the sample* of men than of *the sample* of women believe stem cell research is morally wrong, in the *population* of all men and women these proportions might be the same, might not be the same, or might even be reversed. We need a confidence interval to answer this question.

TRY THIS! Exercise 7.65a

Confidence Intervals for Two Population Proportions

The confidence interval for two proportions has the same structure as the confidence interval for one proportion, as it was presented in Formula 7.2:

$$\text{Statistic} \pm z^* \times SE_{\text{est}}$$

The statistic for two proportions is different; it is now $\hat{p}_1 - \hat{p}_2$. And so the standard error is also different:

$$SE_{\text{est}} = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

The value for z^* is chosen to get the desired confidence level, exactly as we did for a one-proportion confidence interval. For a 95% confidence level, for example, use $z^* = 1.96$. The samples can be different sizes, so n_1 represents the size of the sample drawn from population 1, and n_2 represents the number of people or objects in the sample drawn from population 2.

Putting these together, we find that the confidence interval for the difference of two proportions is

$$\textbf{Formula 7.3: } \hat{p}_1 - \hat{p}_2 \pm z^*\sqrt{\frac{\hat{p}_1(1 - \hat{p}_2)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

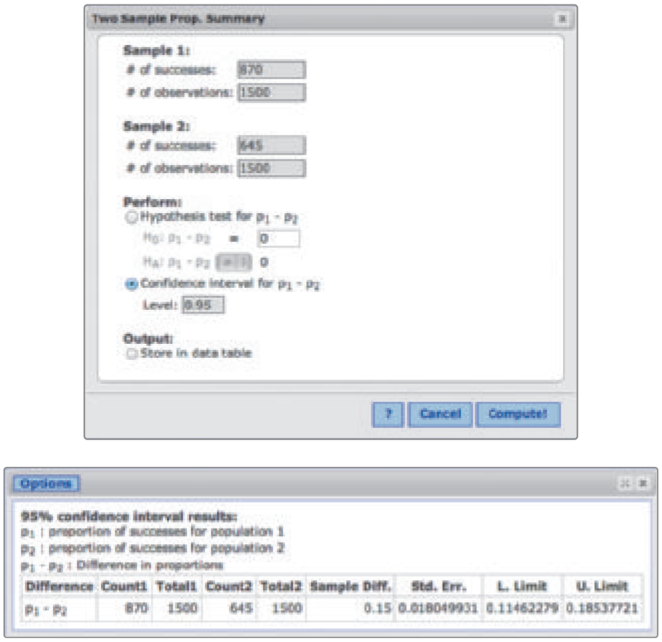
Naturally, we recommend that you use technology to do this calculation whenever possible.

Figure 7.18 shows the calculations used in StatCrunch to compare the results from the Pew Poll on views of people in the United States on embryonic stem cell research in 2002 with the results from a similar poll in 2009. We used the 2009 population as population 1 and the 2002 population as population 2. The 95% confidence interval is 0.11 to 0.19.

Looking Back

***z*-Scores**
You've already seen subtraction used to compare numbers. The numerator of the z-score (the observed value minus the sample mean) is used to tell us the distance between a number and the mean of the sample.

We'll discuss how to interpret this interval later.

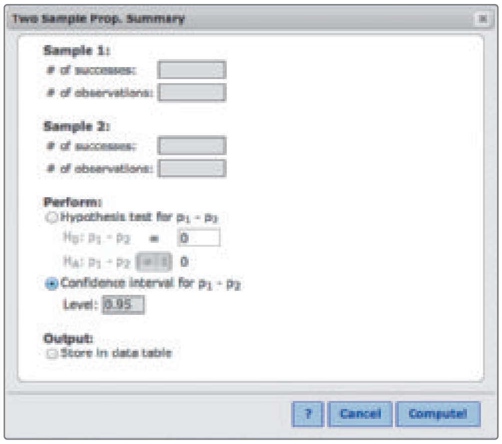


▶ **FIGURE 7.18** StatCrunch calculations for finding a 95% confidence interval for the difference between the proportions of those supporting stem cell research in 2009, p_1, and in 2002, p_2. After rounding, the lower limit of the interval is 0.115, and the upper limit is 0.185. The confidence level is 95% only if conditions are met.

EXAMPLE 12 Estimating Men and Women's Opinions

In 2013, 2000 men and 2000 women were randomly selected for a Pew survey. Interviewers reported that the proportion of men who felt that embryonic stem cell research was morally wrong was 0.23, and the proportion of women who felt it was morally wrong was 0.21 (Pew Foundation 2013). We wish to find a 95% confidence interval for the difference in proportions between men and women who feel this way in the population.

QUESTION Figure 7.19 shows the information that StatCrunch requires to calculate a 95% confidence interval for the difference in population proportions. Fill in the missing information. (Other statistical software packages require similar information.)



▶ **FIGURE 7.19** StatCrunch screenshot for a two-sample proportion test.

SOLUTION First, we must decide which is sample 1 and which is sample 2. It doesn't matter which choice we make, but we must be consistent, and we must remember our choice when we interpret the interval. We will choose to call the women sample 1 and the men sample 2.

Next, we must determine the number of successes in each sample. There were 2000 women in the sample, and the proportion of "successes" (those who agree that embryonic stem cell research is morally wrong) is 0.21. So the number of women who believe that it is morally wrong is $2000 \times 0.21 = 420$.

Similarly, the number of men who believe it is morally wrong is $2000 \times 0.23 = 460$.

We also must make sure that the "Confidence interval" button is checked and that the Level is set to 0.95.

CONCLUSION

Sample 1: number of successes, 420; number of observations, 2000

Sample 2: number of successes, 460; number of observations, 2000

TRY THIS! Exercise 7.65b

Checking Conditions

These calculations "work" only when conditions are met. In a nutshell, the conditions that must exist in order for us to apply the Central Limit Theorem must hold for both samples, and one more condition must be met: The samples must be independent of each other.

To summarize, before you can interpret a confidence interval for two population proportions, you must check:

1. Random and Independent. Both samples are *randomly* drawn from their populations, and observations are independent of each other. (*Note:* An important exception to this is discussed at the end of this section.)
2. Large Samples. Both sample sizes are large enough that at least 10 successes and 10 failures can be expected in both samples.
3. Big Populations. If the samples are collected without replacement, then both population sizes must be at least 10 times bigger than their samples.
4. (New!) Independent Samples. The samples must be independent of each other.

The new condition, condition 4, takes a little explanation. Condition 4 requires that there be no relationship between the objects in one sample and the objects in another. This condition would have been violated, for instance, if Pew had interviewed *the same people* in 2002 and again in 2009. (Note that it is not a bad idea to interview the same people twice across such time periods, if you can track them all down again. When that is done, however, the techniques presented here are not valid.)

Note that for condition 2 you now have four things you must check: (1) at least 10 successes in sample 1, (2) at least 10 failures in sample 1, (3) at least 10 successes in sample 2, and (4) at least 10 failures in sample 2. In symbols,

$$n_1\hat{p}_1 \geq 10,\ n_1(1-\hat{p}_1) \geq 10,\ n_2\hat{p}_2 \geq 10, \text{ and } n_2(1-\hat{p}_2) \geq 10$$

Details

Independent Samples
If the samples are not independent, then the standard error in the confidence interval will be the wrong value, and you'll have the wrong margin of error.

Example 13 Conditions for Men and Women

In Example 12 we did the preliminary steps for calculating the 95% confidence interval for the difference between the proportion of men who believe embryonic stem cell research is wrong and the proportion of women who believe it is wrong. The data came from a Pew study based on a random sample of men and women in the United States. This interval turned out to be

$$-0.046 \text{ to } 0.006$$

or -4.6 percentage points to $+0.6$ percentage point.

QUESTION Check that the conditions hold for interpreting this interval.

SOLUTION Whether or not the conditions hold depends, to a great extent, on whether the Pew researchers followed appropriate procedures. When in doubt, we will assume that they did. (The Pew website goes to some length to convince us that they do follow appropriate procedures.)

Condition 1: Random and Independent. We are told that the samples are random, and we must assume that the observations are independent of each other in both samples.

Condition 2: Large samples. We check all four:

$$2000 \times 0.21 = 420$$
$$2000 \times (1 - 0.21) = 1580$$
$$2000 \times 0.23 = 460$$
$$2000 \times (1 - 0.23) = 1540$$

All values are bigger than 10, so the samples are large enough.

Condition 3. Big Populations. Clearly, there are more than $10 \times 2000 = 20{,}000$ men and more than 20,000 women in the United States.

Condition 4. Because each was a random sample from different populations (the population of all women in the United States and the population of all men in the United States), the samples are independent.

CONCLUSION The conditions are satisfied.

TRY THIS! Exercise 7.65c

Interpreting Confidence Intervals for Two Proportions

The basic interpretation of a confidence interval for the difference of two proportions is the same as for one proportion: We are 95% confident (or whatever our confidence level is) that the true population value is within the interval.

One important difference, though, is that the reason we are looking at the difference of two proportions, $p_1 - p_2$, is that we want to compare them. We want to know which proportion is larger than the other and how much larger it is, or whether they are the same. Therefore, in examining a confidence interval for two proportions, we ask these questions:

1. Is 0 included in the interval? If so, then we can't rule out the possibility that the population proportions are equal.
2. What does a positive value mean? A negative value? What is the greatest plausible difference for the population proportions? The smallest plausible difference? To answer these questions, we have to know which population was assigned to be population 1 and which to be population 2.

In the Pew survey comparing attitudes towards embryonic stem cell research in 2002 and 2009, our confidence interval was 0.11 to 0.19. Note that 0 is not included in this interval. This tells us that we are confident that the population proportions really are different.

All of the values in this interval are positive. What does a positive number mean? It means $p_1 - p_2 > 0$. This can happen only if $p_1 > p_2$—in other words, if the proportion of people who support embryonic stem cell research is greater in population 1 than in population 2.

Now we need to know which is population 1 and which is population 2. Looking back, we see that we defined the 2009 survey as population 1. This tells us that the proportion of people who support embryonic stem cell research was greater in 2009 than in 2002.

How much greater? We are confident that the increase was no fewer than 11 percentage points and no more than 19 percentage points.

Details

Choosing Populations
When comparing proportions from different years, researchers usually choose the most recent year as their "population 1." This makes it possible to interpret the difference in proportions as a change across time.

EXAMPLE 14 Interpreting CIs for Two Proportions

In Example 13, we found that a 95% confidence interval for the difference in proportions between men who believe embryonic stem cell research is morally wrong and the proportion of women who believe it is wrong was −0.046 to 0.006. In percentages, this is −4.6 percentage points to 0.6 percentage point. We used women as population 1 and men as population 2.

QUESTION Interpret this confidence interval.

SOLUTION The interval contains 0. Thus we can't rule out the possibility that the proportion of men who believe this and the proportion of women who believe it are the same. In other words, we can't rule out the possibility that the same proportion of men as of women believe that embryonic stem cell research is morally wrong.

A positive value means that the percentage for women is greater than the percentage for men. We see that it is not implausible that the percentage for women is as much as 0.6 percentage point above the percentage for men. A negative value means that the proportion for women is less than that for men, and we see that the percent of women who feel embryonic stem cell research is morally wrong could plausibly be as much as 4.6 percentage points below that for men.

TRY THIS! Exercise 7.65d

> **Looking Back**
>
> **Controlled Experiments**
> An important feature of a well-designed controlled experiment is that subjects are assigned to treatment and control groups at random. If this doesn't happen, we cannot make cause-and-effect conclusions.

Random Assignment vs. Random Sampling

There is an important exception to condition 1 for confidence intervals of two proportions. Sometimes, a particular study is not concerned with generalizing to a larger population. Sometimes, the purpose is instead to determine whether there is a cause-and-effect relationship between two variables.

If the two samples are not random samples but, instead, objects are *randomly assigned* to groups, then if the other conditions are met, we can interpret a confidence interval for a difference in proportions. This is the situation in which we find ourselves when doing controlled experiments, as discussed in Chapter 1.

> **Looking Back**
>
> **Random Assignment**
> You learned in Section 1.4 that researchers randomly assign subjects to treatment groups in order to determine whether there is a cause-and-effect relationship between the treatment and the response variable.

EXAMPLE 15 Crohn's Disease Proportions

In Chapter 1 you learned about a study to determine (among other things) which of two treatments for Crohn's disease, Inflix injections or Azath pills, was better. Patients were randomly assigned to receive either Inflix or Azath. 169 patients received Inflix, and at the end of the study, 75 of them were in remission (a good outcome). 170 patients received Azath, and at the end of the study, 51 were in remission.

Let p_1 represent the proportion of Crohn's disease victims who would be in remission if they took Inflix, and let p_2 represent the proportion who would be in remission if they took Azath. A 95% confidence interval for the difference in population proportions is 0.04 to 0.25.

QUESTION Assume that conditions 2, 3, and 4 are satisfied. Interpret the confidence interval. Which treatment is better?

SOLUTION Even though the two samples are not randomly selected from the population, the fact that patients were randomly assigned to one of the two treatments, together with the fact that the other three conditions hold, means we can interpret the confidence interval.

The interval does *not* include 0 (although it comes close!) This tells us that we are confident that one treatment is different from the other.

The values of the confidence interval are all positive. A positive value means the proportion of people in remission is greater for population 1, which consists of those who took Inflix.

CONCLUSION Inflix is the better treatment. The percentage of people who will go into remission is at least 4 percentage points greater with Inflix than with Azath, and it could be as much as 25 percentage points greater.

TRY THIS! Exercise 7.69

Random sampling and random assignment are not the same thing. Keep this in mind:

> Random sampling allow us to make generalizations to the population from which the samples were taken.
>
> Random assignment allow us to make cause-and-effect conclusions.

SNAPSHOT CONFIDENCE INTERVAL FOR THE DIFFERENCE OF TWO SAMPLE PROPORTIONS

WHAT IS IT? ► The proportion of people/objects with a particular characteristic in one sample minus the proportion of people/objects with a particular characteristic in another sample.

WHAT IS IT USED FOR? ► To estimate the difference in proportions from two separate populations, for instance, men and women, Republicans and Democrats, or residents in 2018 with residents in 2015.

WHY DO WE USE IT? ► If the samples are independent of each other, and if both samples are randomly selected from their respective populations, then this statistic is an unbiased estimate of $p_1 - p_2$ and has standard error $\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}$.

HOW IS IT USED? ► If, in addition to everything above, both sample sizes are fairly large, then the sampling distribution is approximately Normal, and we can use the Normal distribution to find probabilities for this statistic.

CASE STUDY REVISITED

What was wrong with the American Medical Association's spring break survey? The AMA poll was actually based on an "online survey panel," which consists of a group of people who agree to take part in several different online surveys in exchange for a small payment. Marketing companies recruit people to join panels so that the marketers can investigate trends within various slices of the public. Such a sample may or may not be representative of the population we're interested in—we have no way of knowing. And because the sample is not chosen randomly, we also have no way of knowing how our estimate will behave from sample to sample.

For such a survey, it is impossible to find a confidence interval for the true proportion of women who "agree that spring break trips involve more or heavier drinking than occurs on college campuses" because (1) our estimate might be biased, and (2) the true percentage might lie much farther from our estimate than two standard errors. For this reason, the AMA ended up removing the margin of error from its website and no longer claimed that the figures were a valid inference for all college women who participated in spring break.

EXPLORING STATISTICS
CLASS ACTIVITY

Simple Random Sampling Prevents Bias

GOALS

In this activity, you'll see how the sampling method affects our estimation of a population mean.

MATERIALS

- A list of the first four amendments to the U.S. Constitution with each word numbered
- A random number table or other method of obtaining random numbers

ACTIVITY

James Madison (1751–1836), who became the fourth president of the United States, wrote the first ten amendments to the U.S. Constitution, which are known as the Bill of Rights. They went into effect on December 15, 1791.

Your teacher will give you a page where the first four amendments are printed, with each word numbered. Your goal is to estimate the mean length of the words that appear in these four amendments.

You will use two estimation procedures, compare them, and decide which method works better. One method is an informal method. The other is based on random sampling. Your instructor will give you detailed instructions about each method.

The first ten words of the Bill of Rights are shown in Figure A, with each word numbered.

001	002	003	004	005	006	007	008	009	010
Congress	shall	make	no	law	respecting	an	establishment	of	religion,

▲ **FIGURE A** The First Ten Words of the Bill of Rights

BEFORE THE ACTIVITY

1. What is the mean length of the ten words in the excerpt shown in Figure A?
2. Select any two words you wish from this excerpt. How different is the average length of the two words you selected from the mean length of all ten words?

AFTER THE ACTIVITY

Your instructor will give you data from the entire class. The data will consist of a list of estimates of mean word length based on the "informal" method, and a list based on the simple random sampling method. Make a dotplot of the "informal" estimates and a second dotplot of the estimates obtained using simple random sampling.

1. Compare the distributions from the two methods. (Comment on the shape, center, and spread of the distributions.)
2. Judging on the basis of your comparison of the dotplots, why is simple random sampling preferred over the informal method for collecting data?

This exercise is based on an activity from an INSPIRE workshop, which was based on an activity from Workshop Statistics, © 2004, Dr. Allan Rossman and Dr. Beth Chance, California Polytechnic State University.

CHAPTER REVIEW

KEY TERMS

population, *326*
parameter, *326*
census, *326*
sample, *326*
statistic, *326*
estimator, *326*
estimate, *326*
statistical inference, *327*
biased, *328*
sampling bias, *328*
measurement bias, *328*
convenience sample, *330*
simple random sample (srs), *330*
with and without replacement, *330*
sampling distribution, *335*
bias (accuracy), *336*
precision, *336*
standard error (SE), *336*
Central Limit Theorem (CLT), *341*
confidence interval, *347*
margin of error, *348*
confidence level, *348*

LEARNING OBJECTIVES

After reading this chapter and doing the assigned homework problems, you should

- Be able to estimate a population proportion from a sample proportion and quantify how far off the estimate is likely to be.
- Understand that random sampling reduces bias.
- Understand when the Central Limit Theorem for sample proportions applies and know how to use it to find approximate probabilities for sample proportions.
- Understand how to find, interpret, and use confidence intervals for a single population proportion.

SUMMARY

The Central Limit Theorem (CLT) for sample proportions tells us that if we take a random sample from a population, and if the sample size is large and the population size much larger than the sample size, then the sampling distribution of $\hat{p}$ is approximately

$$N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

This result is used to infer the true value of a population proportion on the basis of the proportion in a random sample. The primary means for doing this is with a confidence interval:

Formula 7.2: $\hat{p} \pm m \quad where \quad m = z^* \times SE_{\text{est}}$

$$\text{and} \quad SE_{\text{est}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where:

$\hat{p}$ is the sample proportion of successes, the proportion of people in the sample with the characteristic we are interested in
m is the margin of error
n is the sample size
z^* is a multiplier that is chosen to achieve the desired confidence level

An important first step is to make sure that the sample size is large enough for the CLT to work. This means that we need the sample size times the sample proportion to be at least 10 and that we need the sample size times (1 minus the sample proportion) to be at least 10.

A 95% confidence interval might or might not have the correct population value within it. However, we are confident that it does, because the method works for 95% of all samples.

Formula 7.3: $\hat{p}_1 - \hat{p}_2 \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

where $\hat{p}_1$ is the sample proportion of successes in the first group and $\hat{p}_2$ is the sample proportion of successes in the second group. Here n_1 is the sample size of the first group and n_2 is the sample size of the second group. Also, z^* is the multiplier chosen to achieve the desired level of confidence.

We can compare two population proportions by finding a confidence interval for their difference (subtract one from the other). If the confidence interval contains 0, it means the population proportions could be equal. If it does not contain 0, then we are confident that the population proportions are not equal, and we should note whether the values in the interval are all positive (the first population proportion is greater than the second) or all negative (the first is less than the second).

SOURCES

Colombel et al. 2010. Infliximab, azathioprine, or combination therapy for Crohn's disease. *New England Journal of Medicine* 362, 1383–1395.
Cornish, A. Do polls miss views of the young and mobile? 2007. National Public Radio. October 1. http://www.npr.org (accessed March 29, 2010).
Crowley, C. 2010. Soda tax or flat tax? Questions can influence poll results, Cornell expert says. *Albany Times Union*, February 5. http://www.timesunion.com (accessed February 6, 2010).
Gallup. Shrunken majority now favors stricter gun laws. 2007. http://www.galluppoll.com October 11 (accessed March 29, 2010).
James, W., et al. Effect of sibutramine on cardiovascular outcomes in overweight and obese subjects. *New England Journal of Medicine* 363, 905–917, September 2.
New York Times. 1993. 1 in 5 in new survey express some doubt about the Holocaust. April 20.
New York Times. 1994. Pollster finds error on Holocaust doubts. May 20.
Pasternak, B., et al. 2013. Ondansetron in pregnancy and risk of adverse fetal outcomes. *New England Journal of Medicine* 368, 814–823, February 28.

Pew Research Center for the People & the Press. 2010. The growing gap between landline and dual frame Election Polls. November 12, 2010, (accessed Dec. 20, 2010).

Pew Foundation. 2013. Abortion viewed in moral terms: Fewer see stem cell research and IVF as moral issues. http://www.pewforum.org/2013/08/15/abortion-viewed-in-moral-terms/ (accessed December 28, 2013).

Pew Research Center. 2013. 18% of Americans say they've seen a ghost. www.pewresearch.org/fact-tank/2013/10/30/18-of-americans-say-theyve-seen-a-ghost/ (accessed December 30, 2013).

Rassmussen Reports. 2013. 62% say their home is worth more than what they still owe. http://www.rasmussenreports.com/public_content/ (accessed December 20, 2013).

Schweinhart, L., et al. 2005. Lifetime effects: The High/Scope Perry Preschool Study through age 40. *Monographs of the High/Scope Educational Research Foundation, 14*. Ypsilanti, MI: High/Scope Press.

Trehan, I., et al. 2013. Antibiotics as part of the management of severe acute malnutrition. *New England Journal of Medicine* 368: 425–435, January 31.

Villanueva, C., et al. 2013. Transfusion strategies for acute upper gastrointestinal bleeding. *New England Journal of Medicine* 368, 11–21, January 3.

SECTION EXERCISES

SECTION 7.1

TRY **7.1 Parameter vs. Statistic (Example 1)** Explain the difference between a parameter and a statistic.

7.2 Sample vs. Census Explain the difference between a sample and a census. Every 10 years, the U.S. Census Bureau takes a census. What does that mean?

7.3 $\bar{x}$ vs. μ Two symbols are used for the mean: μ and $\bar{x}$.

a. Which represents a parameter and which a statistic?

b. In determining the mean age of all students at your school, you survey 30 students and find the mean of their ages. Is this mean $\bar{x}$ or μ?

7.4 $\bar{x}$ vs. μ The mean height of all 1000 employees at an office is 170 cm. A height sample of 50 people from this office has a mean of 185 cm. Which number is μ and which is $\bar{x}$?

* **7.5 GPAs** Suppose you knew the GPA of a random sample of 100 students of the school. Could you use this data to make inferences about the GPAs of the total 500 students in the school? Why or why not?

* **7.6 Sampling Weights** Suppose you want to estimate the mean weight of all the people in your locality. You set up a table outside a park asking for volunteers to tell you their weights. Do you think you would get a representative sample? Why or why not?

7.7 Sample vs. Census You are receiving a large shipment of batteries and want to test their lifetimes. Explain why you would want to test a sample of batteries rather than the entire population.

7.8 Salaries of Top-level Managers Suppose you find *all* the salaries of the top-level managers at a company. Could you use those data to make inferences about salaries of all employees at that office? Why or why not?

7.9 Sampling with and without Replacement Explain the difference between sampling with replacement and sampling without replacement. Suppose you have a deck of 52 cards and want to select two cards. Describe both procedures.

7.10 Simple Random Sampling Is simple random sampling usually done with or without replacement?

TRY **7.11 Finding a Random Sample (Example 2)** You need to select a simple random sample of four from eight friends who will participate in a survey. Assume the friends are numbered 1, 2, 3, 4, 5, 6, 7, and 8.

Select four friends, using the two lines of numbers in the next column from a random number table.

Read off each digit, skipping any digit not assigned to one of the friends. The sampling is without replacement, meaning that you cannot select the same person twice. Write down the numbers chosen. The first person is number 7.

0 <u>7</u> 0 3 3	7 5 2 5 0	3 4 5 4 6
7 5 2 9 8	3 3 8 9 3	6 4 4 8 7

Which four friends are chosen?

7.12 Finding a Random Sample You need to select a simple random sample of two from six friends who will participate in a survey. Assume the friends are numbered 1, 2, 3, 4, 5, and 6.

Use technology to select your random sample. Indicate what numbers you obtained and how you interpreted them.

If technology is not available, use the line from a random number table that corresponds to the day of the month on which you were born. For example, if you were born on the fifth day of any month, you would use line 05. Show the digits in the line and explain how you interpreted them.

* **7.13 Random Sampling** Assume your class has 30 students and you want a random sample of 10 of them. Describe how to randomly select 10 people from your class using the random number table.

7.14 Random Sampling with Coins Assume your class has 30 students and you want a random sample of 10 of them. A student suggests asking each student to flip a coin, and if the coin comes up heads, then he or she is in your sample. Explain why this is not a good method.

TRY **7.15 Questionnaire Response (Example 3)** A teacher at a community college sent out questionnaires to evaluate how well the administrators were doing their jobs. All teachers received questionnaires, but only 10% returned them. Most of the returned questionnaires contained negative comments about the administrators. Explain how an administrator could dismiss the negative findings of the report.

7.16 Survey on Tax Benefits A survey was conducted to ask whether tax benefits for senior citizens should be continued or stopped. Only clubs were visited to collect data. Do you think this would introduce bias? Explain.

7.17 Views on Capital Punishment In carrying out a study of views on capital punishment, a student asked a question two ways:

1. With persuasion: "My brother has been accused of murder and he is innocent. If he is found guilty, he might suffer capital punishment. Now do you support or oppose capital punishment?"
2. Without persuasion: "Do you support or oppose capital punishment?"

Here is a breakdown of her actual data.

Men

	With persuasion	No persuasion
For capital punishment	6	13
Against capital punishment	9	2

Women

	With persuasion	No persuasion
For capital punishment	2	5
Against capital punishment	8	5

a. What percentage of those persuaded against it support capital punishment?

b. What percentage of those not persuaded against it support capital punishment?

c. Compare the percentages in parts a and b. Is this what you expected? Explain.

7.18 Views on Capital Punishment Use the data given in Exercise 7.17.

Make the two given tables into one table by combining men for capital punishment into one group, men opposing it into another, women for it into one group, and women opposing it into another. Show your two-way table.

The student who collected the data could have made the results misleading by trying persuasion more often on one gender than on the other, but she did not do this. She used persuasion on 10 of 20 women (50%) and on 15 of 30 men (50%).

a. What percentage of the men support capital punishment? What percentage of the women support it?

b. On the basis of these results, if you were on trial for murder and did not want to suffer capital punishment, would you want men or women on your jury?

SECTION 7.2

7.19 Targets: Bias or Lack of Precision?

a. If a rifleman's gunsight is adjusted incorrectly, he might shoot bullets consistently close to 2 feet left of the bull's-eye target. Draw a sketch of the target with the bullet holes. Does this show lack of precision or bias?

b. Draw a second sketch of the target if the shots are both unbiased and precise (have little variation).

The rifleman's aim is not perfect, so your sketches should show more than one bullet hole.

7.20 Targets: Bias or Lack of Precision?, Again

a. If a rifleman's gunsight is adjusted correctly but he has shaky arms, the bullets might be scattered widely around the bull's-eye target. Draw a sketch of the target with the bullet holes. Does this show variation (lack of precision) or bias?

b. Draw a second sketch of the target if the shots are unbiased and have precision (little variation).

The rifleman's aim is not perfect, so your sketches should show more than one bullet hole.

7.21 Biased Sample? Suppose you go to a department store where one can shop both in-store and online. You want to know the average purchase volume per customer. You walk around the store asking the customers their order values. Would this result in a biased sample?

7.22 Bias? Suppose that, when taking a sample of five students' heights, you get a sample mean of 183 cm. This sample mean is far higher than the class-wide (population) mean. Does that prove that your sample is biased? Explain. What else could have caused this high mean?

7.23 Proportion of Multiples of 3 An observation of the outcomes of rolling a die has about 33.33% multiples of 3 and 66.67% non-multiples of 3, because two of the six outcomes are multiples of 3 (3 and 6) and four are not (1, 2, 4, and 5).

a. Find the proportion of multiples of 3 in the following observations from a random roll of a die. Count carefully.

3 1 2 5 6	1 2 3 5 1	3 4 2 3 5
2 6 3 4 6	4 3 1 5 1	6 1 3 2 2

b. Does the proportion found in part a represent $\hat{p}$ (the sample proportion) or p (the population proportion)?

c. Find the error in this estimate, the difference between $\hat{p}$ and p (or $\hat{p} - p$).

7.24 Proportion of Multiples of 3 In a die roll, 3 and 6 are multiples of 3 and 1, 2, 4, and 5 are not multiples of 3. Consider 90 rolls of a die on a random basis.

a. From how many outcomes of the 90 rolls would you expect to get multiples of 3, on average?

b. If you actually counted, would you get exactly the number you predicted in part a? Explain.

TRY **7.25 Juice (Example 4)** According to a leading juice company, mixed fruit juice makes up 35% of the total sales. Suppose we examine 200 random customers.

a. How many customers should we expect for our sample percentage of mixed fruit juice?

b. What is the standard error?

c. Use your answer to fill in the blanks:
We expect ____% mixed fruit juice buyers, give or take _____%.

7.26 Random Letters Samuel Morse suggested in the nineteenth century that the letter "t" made up 9% of the English language. Assume this is still correct. A random sample of 1000 letters is taken from a randomly selected, large book and the t's are counted.

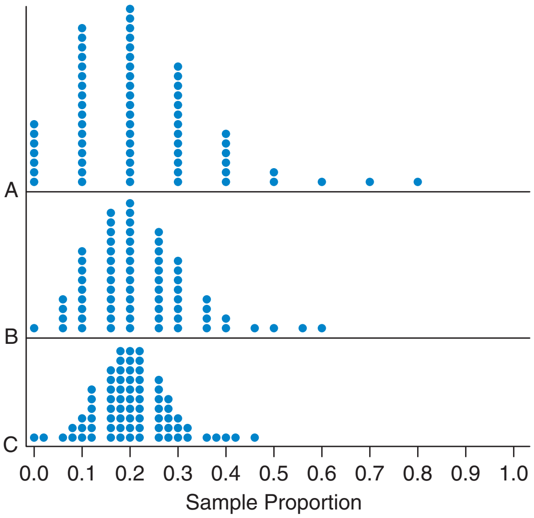
a. What value should we expect for our sample percentage of t's?

b. Calculate the standard error.

c. Use your answers to fill in the blanks:
We expect ____% t's, give or take _____%.

7.27 ESP A Zener deck of cards has cards that show one of five different shapes with equal representation, so that the probability of selecting any particular shape is 0.20. A card is selected randomly, and a person is asked to guess which card has been chosen. The graph below shows a computer simulation of experiments in which a "person" was asked to guess which card had been selected for a large number of trials. (If the person does not have ESP, then his or her proportion of successes should be about 0.20, give or take some amount.) Each dot in the dotplots represents the proportion of success for one person. For instance, the dot in Figure A farthest to the right represents a person with an 80% success rate. One dotplot represents an experiment in which each person had 10 trials; another shows 20 trials; and a third shows 40 trials.

Explain how you can tell, from the widths of the graphs, which has the largest sample ($n = 40$) and which has the smallest sample ($n = 10$).

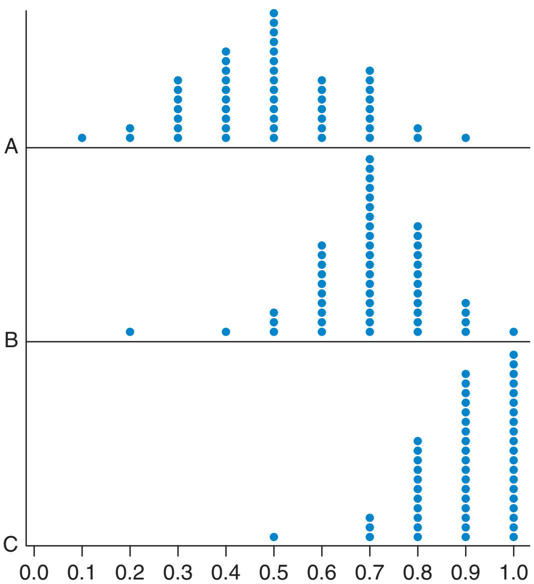


7.28 ESP Again In the graph for Exercise 7.27, explain how you can tell from the *shape* of the graphs which has the largest sample size and which has the smallest sample size.

7.29 Standard Error Which of the dotplots given in Exercise 7.27 has the largest standard error, and which has the smallest standard error?

7.30 Bias? Assuming that the true proportion of success for the trials shown in the graph for Exercise 7.27 is 0.2, explain whether any of the graphs shows bias.

7.31 Fair Coin? One of the graphs shows the proportion of heads from flipping a fair coin 10 times, repeatedly. The others do not. Which graph represents the coin flips? Explain how you know.



7.32 Far from Fair Which of the graphs in Exercise 7.31 is centered farthest from 0.50?

TRY * **7.33 What Is the Proportion of Seniors?** A population of college students is taking an advanced math class. In the class are three juniors and two seniors. Using numbers 1, 2, and 3 to represent juniors and 4 and 5 to represent seniors, sample without replacement. Draw a sample of two people four times (once in each of parts a, b, c, and d), and then fill in the following table.

a. Use the first line (reprinted here) from the random number table to select your sample of two. (The selections are underlined.)

0 **2** 7 7 9   7 2 6 **4** 5   3 2 6 9 9   8 6 0 0 9

Report the percentage of seniors in the sample. (Count the number of 4's and 5's and divide by the sample size 2.)

b. Use the next line to select your sample of two.

3 1 8 6 7   8 5 8 7 2   9 1 4 3 0   4 5 5 5 4

Report the percentage of seniors in the sample.

c. Use the next line to select your sample of two.

0 7 0 **3** 3   7 5 2 5 0   3 4 5 4 6   7 5 2 9 8

Report the percentage of seniors in the sample.

d. Use the last line to select your sample of two.

0 9 0 8 4   9 8 9 4 8   0 9 5 4 1   8 0 6 2 3

Report the percentage of seniors in the sample.

e. Fill in the rest of the table below, showing the results of the four samples.

Repetition	p (Population Proportion of Seniors)	$\hat{p}$ (Sample Proportion of Seniors)	Error: $\hat{p} - p$
1 (from part a)	$2/5 = 0.4$	$1/2 = 0.5$	$0.5 - 0.4 = 0.1$
2 (from part b)			
3			
4			

7.34 Simulation From a very large (essentially infinite) population, of which half are men and half are women, you take a random sample, with replacement. Use the following random number table and assume each single digit represents selection of one person; the odd numbers represent women and the even numbers (0, 2, 4, 6, 8) men.

a. Start on the left side of the top line (with 118) and count 10 people. What percentage of the sample will be men?

1 1 8 4 8	8 0 8 0 9	2 5 8 1 8	3 8 8 5 7
2 3 8 1 1	8 0 9 0 2	8 5 7 5 7	3 3 9 6 3
9 3 0 7 6	3 9 9 5 0	2 9 6 5 8	0 7 5 3 0

b. Start in the middle of the second line (with 857) and count 20 people. What percentage of the sample will be men?

c. If you were to repeat parts a and b many times, which sample would typically come closer to 50%—the sample of 10 or the sample of 20? Why?

SECTION 7.3

* **7.35 Juice (Example 5)** Return to Exercise 7.25 and find the approximate probability that out of 100 random buyers, 38% or more will buy mixed fruit juice.

* **7.36 Random Letters** Return to Exercise 7.26 and find the approximate probability that the random sample of 1000 letters will contain 8.1% or fewer t's.

TRY **7.37 Habitat Protection (Example 6)** Natural habitats must be protected to maintain the ecological balance. According to indexmundi.com's survey in 2010, about 26% of the land in Brazil is a habitat-protected area. Suppose a geologist randomly selects 200 regions to study soil types found in the country.

Use the Central Limit Theorem (and the Empirical Rule) to find the approximate probability that the proportion of protected regions is more than one standard error from the population value of 0.26.

The conditions for using the Central Limit Theorem are satisfied because the sample is random; the population is more than 10 times 1000; *n* times *p* is 52, and *n* times (1 minus *p*) is 148, and both are more than 10.

* **7.38 Millionaires with Master's Degrees** According to an article in randomhistory.com, only 18% of all millionaires in the world have a master's degree. Suppose a conclave of millionaires contains 150 millionaires that were randomly sampled from the population of millionaires.

Use the Central Limit Theorem (and the Empirical Rule) to find the approximate probability that the conclave will have a proportion of millionaires with master's degrees that is more than two standard errors below 0.18.

You can use the Central Limit Theorem because the millionaires were randomly sampled; the population is more than 10 times 150; and *n* times *p* is 27 and *n* times (1 minus *p*) is 123, and both are more than 10.

TRY * g **7.39 The Oregon Bar Exam (Example 7)** According to the Oregon Bar Association, approximately 65% of the people who take the bar exam to practice law in Oregon pass the exam. Find the approximate probability that at least 67% of 200 randomly sampled people taking the Oregon bar exam will pass. (In other words, find the probability that at least 134 out of 200 will pass.) *See page 373 for guidance.*

* **7.40 Feeding Vegans** A survey of eating habits showed that approximately 4% of people in Portland, Oregon, are vegans. Vegans do not eat meat, poultry, fish, seafood, eggs, or milk. A restaurant in Portland expects 300 people on opening night, and the chef is deciding on the menu. Treat the patrons as a simple random sample from Portland and the surrounding area, which has a population of about 600,000.

If 14 vegan meals are available, what is the approximate probability that there will not be enough vegan meals—that is, the probability that 15 or more vegans will come to the restaurant? Assume the vegans are independent and there are no families of vegans.

* **7.41 Overweight Children** The *Ventura County Star* (June 20, 2012) reported on a study of children in public schools in California that looked at the proportion of overweight or obese children. In Huntington Park (a small city outside Los Angeles), 53% of the children were overweight or obese; this was the highest rate found in any city in California. Suppose a random sample of 200 public school children is taken from Huntington Park. Assume the sample was taken in such a way that the conditions for using the Central Limit Theorem are met. We are interested in finding the probability that the proportion of overweight/obese children in the sample will be greater than 0.50.

a. Without doing any calculations, determine whether this probability is greater than 50% or if it is less than 50%. Explain.

b. Calculate the probability that 50% or more are overweight or obese.

7.42 Living in Poverty The *Ventura County Star* article mentioned in Exercise 7.41 also reported that 25% of the residents of Huntington Park lived in poverty. Suppose a random sample of 400 residents of Huntington Park is taken. We wish to determine the probability that 30% or more of our sample will be living in poverty.

a. Before doing any calculations, determine whether this probability is greater than 50% or less than 50%. Why?

b. Calculate the probability that 30% or more of the sample will be living in poverty Assume the sample is collected in such a way that the conditions for using the CLT are met.

* **7.43 Passing a Test by Guessing** A true/false test has 40 questions. A passing grade is 60% or more correct answers.

a. What is the probability that a person will guess correctly on one true/false question?

b. What is the probability that a person will guess incorrectly on one question?

c. Find the approximate probability that a person who is just guessing will pass the test.

d. If a similar test were given with multiple-choice questions with four choices for each question, would the approximate probability of passing the test by guessing be higher or lower than the approximate probability of passing the true/false test? Why?

* **7.44 Age: Randomly Chosen?** A large community college district has 1500 teachers, of whom 50% are below 45 years and 50% are above 45 years of age. In this district, administrators are promoted from among the teachers. There are currently 80 administrators; 75% of these administrators are above 45 years of age.

a. If administrators are selected randomly from faculty, what is the approximate probability that the percentage of administrators above 45 years of age is 75% or more?

b. If administrators are selected randomly from the faculty, what is the approximate probability that the percentage of administrators below 45 years of age is 25% or less?

c. How are part a and part b related?

d. Do your answers suggest that it is reasonable to reject the claim that the administrators have been selected randomly from the teachers? Answer yes or no, and explain your answer.

SECTION 7.4

7.45 Death Penalty In November 2011, a Pew Poll showed that 1241 out of 2001 randomly polled people in the United States favor the death penalty for those convicted of murder. Assuming the conditions for using the CLT were met, answer these questions.

Sample	X	N	Sample p	95% CI
1	1241	2001	0.620190	(0.598925, 0.641455)

Minitab Output

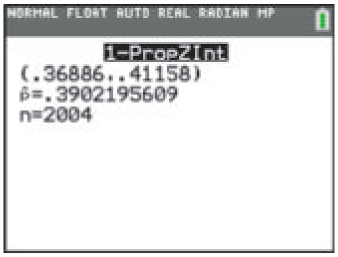
a. Using the Minitab output given, write out the following sentence, filling in the blanks. I am 95% confident that the population proportion favoring the death penalty is between _______ and _______. Report each number correct to three decimal places.

b. Is it plausible to claim that a majority favor the death penalty? Explain.

7.46 Is Marriage Becoming Obsolete? When asked whether marriage is becoming obsolete, 782 out of 2004 randomly selected adults who responded to a 2010 Pew Poll said yes. Refer to the output given. Assume the conditions for using the CLT were met.

a. Report a 95% confidence interval for the proportion. Use correct wording (see Exercise 7.45).

b. Is it plausible to claim that a majority think marriage is becoming obsolete?



7.47 East Germany According to a Gallup Poll taken in East Germany, when adults were asked whether they were thriving, struggling, or suffering, 261 out of 435 said they were struggling. (The number that said they were suffering was 50. In West Germany the number of those who reported suffering was less than half of that.)

a. What is the value of $\hat{p}$, the sample proportion that say they are struggling?

b. Check the conditions to determine whether you can apply the CLT to find a confidence interval.

c. Find a 95% confidence interval for the population proportion that say they are struggling.

7.48 View of Immigration In June 2012, a Gallup Poll asked U.S. adults whether immigration was a good thing or a bad thing for the country. Out of 1004 respondents, 663 said it was a good thing.

a. What percentage of those taking the poll said immigration was a good thing?

b. Check the conditions to determine whether you can use the CLT to find a confidence interval.

c. Find a 95% confidence interval for the proportion who believe immigration is a good thing.

7.49 Higher Education A random sample of likely students for higher studies showed that 28% would want to pursue economics. The margin of error is 4.5 percentage points with a 95% confidence level.

a. Using a carefully worded sentence, report the 95% confidence interval for the percentage of students who plan to choose economics.

b. Is there evidence that there will not be enough students for economics?

c. Suppose the survey was conducted in one section out of 12 sections of the classes eligible to participate in the survey. Explain how that would affect your conclusion.

7.50 Higher Education A random sample of likely students for higher studies showed that 72% want to pursue medicine, with a margin of error of 3.5 percentage points and with a 95% confidence level.

a. Use a carefully worded sentence to report the 95% confidence interval for the percentage of students who plan to choose medicine.

b. Is there evidence that there will not be enough students for medicine?

c. Suppose the survey was conducted in one section out of 12 sections of the class eligible to participate in the survey. Explain how that would affect your conclusion.

TRY **7.51 Social Service (Example 8)** In a simple random sample of 2000 residents, 77% have volunteered for a cleanliness drive in the locality.

a. What is the standard error for this estimate of the percentage of all residents who have volunteered for a cleanliness drive?

b. Find the margin of error, using a 95% confidence level, for estimating the percentage of all residents who have volunteered for a cleanliness drive.

c. Report the 95% confidence interval for the percentage of all residents who have volunteered for a cleanliness drive.

d. Suppose that in the past, 60% of all the residents had volunteered for a cleanliness drive. Does the confidence interval you found in part c support or refute the claim that the percentage of residents who have volunteered for a cleanliness drive has increased? Explain.

7.52 Osteoarthritis In a random sample of 1800 persons aged 65 and over, the proportion with osteoarthritis was found to be 0.225 (or 22.5%).

a. What is the standard error for the estimate of the proportion of all person aged 65 and over with osteoarthritis?

b. Find the margin of error, using a 95% confidence level, for estimating this proportion.

c. Report the 95% confidence interval for the proportion of all persons of aged 65 and over with osteoarthritis.

d. According to patient.info, 50% of all persons aged 65 and over have osteoarthritis. Does the confidence interval you found in part c support or refute this claim? Explain.

7.53 Confidence in Public Schools A 2012 Gallup Poll reported that only 581 out of a total of 2004 U.S. adults said they had a "great deal of confidence" or "quite a lot of confidence" in the public school system. This was down 5 percentage points from the previous year. Assume the conditions for using the CLT are met.

a. Find a 99% confidence interval for the proportion that express a great deal of confidence or quite a lot of confidence in the public schools, and interpret this interval.

b. Find a 90% confidence interval and interpret it.

c. Find the width of each interval by subtracting the lower proportion from the upper proportion, and state which interval is wider.

d. How would a 95% interval compare with the others in width?

7.54 Confidence in Doctors In 2014, Equiniti 360 Clinical reported that 52,250 out of a total of 55,000 patients in Wales said they had a great deal of confidence in their doctors. Assume the conditions for using the Central Limit Theorem are met.

a. Find a 99% confidence interval for the proportion that has a great deal of confidence in the doctors and interpret this interval.

b. Find a 90% confidence interval and interpret it.

c. Which interval is wider?

d. How would a 95% interval compare with the others in width?

* **7.55 Social Media** A survey showed that when a sample group was asked whether they used Facebook for networking, staying connected, or playing games, about 80% said they used it for staying connected.

a. Assuming the sample size was 5000, how many in the sample said they were using it to stay connected?

b. Is the sample size large enough to apply the Central Limit Theorem? Explain. Assume the other conditions are met.

c. Find a 95% confidence interval for the proportion that said they were using it to stay connected.

d. Find the width of the interval you found in part c by subtracting the lower limit from the upper limit. Round your answer to the nearest whole percentage.

e. Now assume the sample size was 20,000. How many would have said they were using it stay connected?

f. Find a 95% confidence interval for the population that said they were using it to stay connected using the numbers from part e.

g. What is the width of the interval you found in part f?

h. When the sample size is multiplied by 4 (as it was here, going from 5000 to 20,000), is the interval width decreased by a factor of 4? In other words, is the interval for the larger sample size one-fourth (1/4) of the interval for the smaller sample size? If not, what is it divided by?

(Source: http://www.statcrunch.com/5.0/viewreport.php?reportid=59990. Accessed via StatCrunch. Owner: daniellerudd@hotmail.com)

* **7.56 Benefits of FDI** In April 2003, a CRISIL survey showed that 17.59% of the labor market is fully satisfied with FDI in India.

a. Assuming the sample size was 800, how many would have said that the labor market is satisfied with FDI?

b. Is the sample size large enough to apply the Central Limit Theorem (CLT)? Explain. Assume the other conditions for using the CLT are met.

c. Find a 95% confidence interval for the percentage that is satisfied with FDI, using the numbers from part a.

d. Find the width of the interval you found in part c, by subtracting the lower boundary from the upper boundary.

e. Now assuming the sample size was multiplied by 5 ($n = 4500$) and the percentage was still 17.59%, how many would have been satisfied by FDI?

f. Find a 95% confidence interval, using the numbers from part e.

g. What is the width of the interval you found in part f?

h. When the sample size is multiplied by 5, is the width of the interval divided by 5? If not, what *is* it divided by?

TRY **7.57 Understanding the Meaning of Confidence Levels: 90% (Example 9)** Each student in a class of 40 was randomly assigned one line of a random number table. Each student then counted the odd-numbered digits in a 30-digit line. (Remember that 0, 2, 4, 6, and 8 are even.)

a. *On average*, in the list of 30 digits, how many odd-numbered digits would each student find?

b. If each student found a 90% confidence interval for the percentage of odd-numbered digits in the entire random number table, how many intervals (out of 40) would you expect *not* to capture the population percentage of 50%?

7.58 Understanding the Meaning of Confidence Levels: 80% Each student in a class of 30 was assigned one random line of a random number table. Each student then counted the even-numbered digits in a 30-digit line.

a. *On average*, in the list of 30 digits, how many even-numbered digits would each student find?

b. If each student found an 80% confidence interval for the percentage of even-numbered digits, how many intervals (out of 30) would you expect *not* to capture 50%? Explain how you arrived at your answer.

7.59 Russian Presidential Vote In the 2012 Russian presidential election, 45,513,001 people voted for Vladimir Putin, 12,288,624 voted for Gennady Zyuganov, and 12,885,159 voted for other candidates.

(Source: Central Election Commission of the Russian Federation)

a. What percentage of voters chose Vladimir Putin?

b. Would it be inappropriate to find a confidence interval for the proportion of voters choosing Putin? Why or why not?

7.60 Human Cloning In a Gallup Poll, 441 of 507 adults said it was "morally wrong" to clone humans.

a. What proportion of the respondents believed it morally wrong to clone humans?

b. Find a 95% confidence interval for the population proportion who believed it is morally wrong to clone humans. Assume that Gallup used a simple random sample.

c. Find an 80% confidence interval (using a z^* of 1.28 if you are calculating by hand).

d. Which interval is wider, and why?

TRY **7.61 Do People Think Astrology Is Scientific? (Example 10)** In the 2012 General Social Survey, people were asked their opinions on astrology—whether it was very scientific, somewhat scientific, or not at all scientific. Of 1974 who responded, 101 said astrology was very scientific.

a. Find the proportion of people in the survey who believe astrology is very scientific.

b. Find a 95% confidence interval for the population proportion with this belief.

c. Suppose a TV news anchor said that 5% of people in the general population think astrology is very scientific. Would you say that is plausible? Explain your answer.

7.62 Do People Think the Sun Goes around the Earth? In the 2012 General Social Survey, people were asked whether they thought the sun went around the planet Earth or vice versa. Of 1974 people, 203 thought the sun went around Earth.

a. What proportion of people in the survey believed the sun went around Earth?

b. Find a 95% confidence interval for the proportion of all people with this belief.

c. Suppose a scientist said that 30% of people in the general population believe the sun goes around Earth. Using the confidence interval, would you say that was plausible? Explain your answer.

SECTION 7.5

7.63 Good News on Jobs The Pew Research Center reports on a survey taken in late 2013 in which they asked whether respondents have heard "good news" about the job market. They compared those making $30,000 or less per year with those making between $31,000 and $74,000. We'll label the population of those making less than $30,000 as population 1 (low income), and those making between $31,000 and $74,000 as population 2 (middle income). A 95% confidence interval for the difference in proportions, 1 (low) minus 2 (middle), is (−0.08 to 0.02). Interpret this confidence interval. If the interval contains 0, indicate what this means. Explain the meaning of positive and/or negative values.

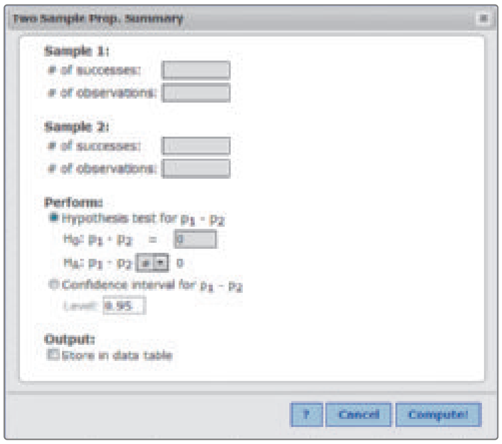
7.64 Go Green A survey report shows that in an office 47% of the senior management and 34% of executives implemented the motto "Save Water, Save Trees, Go Green" in routine office activities. A 95% confidence interval for the difference in these proportions is (−0.15 to −0.09), or, in other terms of percentage points, (−15% to −9%). (Use the executives as population 1.) Interpret this confidence interval. If the interval contains 0, indicate what this means. Explain the meaning of positive and/or negative values.

TRY **7.65 Stressed Moms (Examples 11, 12, 13, and 14)** An April 2012 Gallup Poll of low-income moms showed that 54%

of stay-at-home moms experienced stress, and 49% of employed moms experienced stress. All respondents had an annual household income of less than $36,000.

a. (Example 11) Can we conclude, on the basis of these two percentages alone, that in the United States, a greater percentage of stay-at-home moms than of employed moms experienced stress? Explain.

b. (Example 12) Assume that the sampling was done with 1000 stay-at-home moms and 1000 working moms. Report the numbers you would need to enter into StatCrunch as shown in the figure. Let sample 1 be the stay-at-home moms.



c. (Example 13) Check that the conditions for using the two-proportion confidence interval hold. Note that polls such as Gallup do not ask poll questions of related people. For example, they do not ask questions of both husbands and wives.

d. (Example 14) Find a 95% confidence interval for the difference in proportions (stay-at-home minus working), and interpret the interval.

7.66 European Union According to a Eurobarometer survey, only 27% of the respondents in Cyprus in 2013 favored European Union (EU) membership, whereas in 2012, 35% of the respondents favored it.

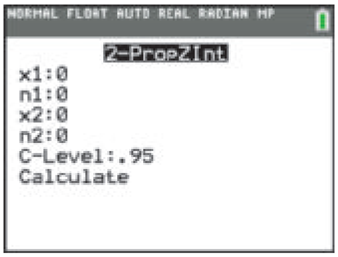
a. Can we conclude, from these two percentages alone, that in Cyprus, a lower percentage favored EU membership in 2013 than in 2012? Explain.

b. Assume that the sampling was done with 1000 random people in 2012 and 1000 random people in 2013. How many of the people in 2012 and how many of the people in 2013 favored EU membership?

c. Assume again that the sampling was done with 1000 random people in 2012 and 1000 random people in 2013. Report the numbers you would need to put in the TI-84 input screen to replace the zeroes. Let sample 1 be the proportion supporting EU membership in 2013.

d. Check that the conditions for using the two-proportion confidence interval hold.

e. Find a 95% confidence interval for the difference in proportions supporting EU membership (2013 minus 2012), and interpret the interval.

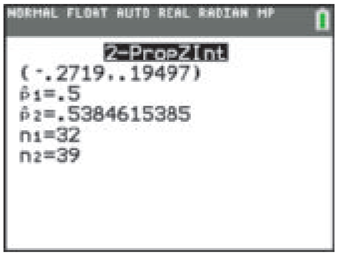


g 7.67 Perry Preschool and Graduation from High School The Perry Preschool Project was created in the early 1960s by David Weikart in Ypsilanti, Michigan. In this project, 123 African American children were randomly assigned to one of two groups: One group enrolled in the Perry Preschool, and the other group did not. Follow-up studies were done for decades. One research question was whether attendance at preschool had an effect on high school graduation. The table shows whether the students graduated from regular high school or not and includes both boys and girls (Schweinhart et al. 2005). Find a 95% confidence interval for the difference in proportions, and interpret it. *Refer to page 373 for guidance.*

	Preschool	No Preschool
Grad HS	37	29
No Grad HS	20	35

7.68 Preschool: Just the Boys Refer to Exercise 7.67 for information. This data set records results just for the boys.

	Preschool	No Preschool
Grad HS	16	21
No Grad HS	16	18



a. Find and compare the percentages that graduated for each group, descriptively. Does this suggest that preschool was linked with a higher graduation rate?

b. Verify that the conditions for a two-proportion confidence interval are satisfied.

c. Indicate which one of the following statements is correct.

i. The interval does not capture 0, suggesting that it is plausible that the proportions are the same.

ii. The interval does not capture 0, suggesting that it is not plausible that the proportions are the same.

iii. The interval captures 0, suggesting that it is plausible that the population proportions are the same.

iv. The interval captures 0, suggesting that it is not plausible that the population proportions are the same.

d. Would a 99% confidence interval be wider or narrower?

TRY 7.69 Antibiotics for Malnutrition (Example 15) A study was done of children from Malawi (in southeastern Africa) with severe acute malnutrition (Trehan et al. 2013). Of the 922 children randomized to receive amoxicillin in addition to food, 817 recovered. Of the 922 children randomized to receive the placebo and food, 785 recovered. The trial was double-blind.

a. Find the percentage that survived in each group, and compare the percentages descriptively.
b. Check that the conditions are met.
c. Find a 95% confidence interval for the difference in proportions. It is plausible that the recovery rates are the same. Also state whether receiving the antibiotic treatment (and food) *causes* a better result than receiving the placebo (and food).

7.70 Transfusions for Bleeding in the Stomach Should patients who are bleeding from the stomach get transfusions when their hemoglobin level falls below 7 grams per deciliter (restrictive strategy) or when it falls to 9 grams per deciliter (liberal strategy)? A study used random assignment with 461 patients assigned to the restrictive strategy and 460 to the liberal strategy (Villanueva et al. 2013). The table shows survival after six weeks in each group.

	Restrictive	Liberal
Alive	438	419
Not alive	23	41

a. Calculate the percentage alive in each of the treatment groups, and write a sentence comparing the percentages in the context of this study. Call the restrictive group 1 and the liberal group 2, and find the difference, $p_1 - p_2$.
b. Find a 95% confidence interval for the difference in proportions alive in the two groups, and explain whether the interval captures 0 and what that means. Use the same difference as in part a. Assume that the conditions are met, although we have random assignment and not a random sample.
c. For the 95% confidence interval from part b, explain what the signs of the boundary values mean and whether we can conclude that one treatment causes the better result.
d. If you had a relative with bleeding and his hemoglobin level dropped to 8.5 grams per deciliter, would you argue for a transfusion or wait until it dropped further?

7.71 Gender and Use of Turn Signals Statistics student Hector Porath wanted to find out whether gender and the use of turn signals when driving were independent. He made notes when driving in his truck for several weeks. He noted the gender of each person that he observed and whether he or she used the turn signal when turning or changing lanes. (In his state, the law says that you must use your turn signal when changing lanes, as well as when turning.) The data he collected are shown in the table.

	Men	Women
Turn signal	585	452
No signal	351	155
	936	607

a. What percentage of men used turn signals, and what percentage of women used them?
b. Assuming the conditions are met (although admittedly this was not a random selection), find a 95% confidence interval for the difference in *percentages*. State whether the interval captures 0, and explain whether this provides evidence that the proportions of men and women who use turn signals differ in the population.
c. Another student collected similar data with a smaller sample size:

	Men	Women
Turn Signal	59	45
No Signal	35	16
	94	61

First find the percentage of men and the percentage of women who used turn signals, and then, assuming the conditions are met, find a 95% confidence interval for the difference in *percentages.* State whether the interval captures 0, and explain whether this provides evidence that the percentage of men who use turn signals differs from the percentage of women who do so.

d. Are the conclusions in parts b and c different? Explain.

7.72 Diet Drug (Meridia) A randomized, placebo-controlled study of a diet drug (Meridia) was done on overweight or obese subjects and reported in the *New England Journal of Medicine* (James et al. 2010). The patients were all 55 years old or older with a high risk of cardiovascular events. Those who had a heart event had either a heart attack, a stroke, or death from heart-related factors. The table gives the counts.

	Drug	Placebo
Heart Event	559	490
No Heart Event	4347	4408
	4906	4898

a. Compare the rates of heart events for the drug group and for the placebo group.
b. Assuming the conditions are met (although we have random assignment and not a random sample), find a 95% confidence interval for the difference in proportions. In particular, does the interval capture 0, and what does that show?
c. Say you collected a new sample 10 times larger. Assuming the proportions came out the same, would the interval change? Would the new interval be narrower or wider?

7.73 Drug for Nausea Ondansetron (Zofran) is a drug used by some pregnant women for nausea. There was some concern that it might cause trouble with pregnancies. An observational study was done of women in Denmark (Pasternak et al. 2013). An analysis of 1849 exposed women and 7376 unexposed women showed that 2.1% of the women exposed to ondansetron and 5.8% of the unexposed women had miscarriages during weeks 7–22 of their pregnancies.

a. Was the rate of miscarriages higher with Zofron, as feared?
b. Which of the conditions for finding the confidence interval do not hold?

7.74 Preschool: Just the Girls The Perry Preschool Project was created in the early 1960s by David Weikart in Ypsilanti, Michigan. In this project, 123 African American children were randomly assigned to one of two groups: One group enrolled in the Perry Preschool, and one group did not enroll. Follow-up studies were done for decades. One research question was whether attendance at preschool had an effect on high school graduation.

The table shows whether the students graduated from regular high school or not and includes *girls* only (Schweinhart et al. 2005).

	Preschool	No Preschool
HS Grad	21	8
No HS Grad	4	17

a. Find the percentages that graduated for both groups, and compare them descriptively. Does this suggest that preschool was associated with a higher graduation rate?

b. Which of the conditions fail so that we cannot use a confidence interval for the difference between proportions?

CHAPTER REVIEW EXERCISES

7.75 Banning Super-Size Sugary Drinks A June 2012 Rasmussen Poll showed that 65% of its randomly selected U.S. adults were opposed to banning super-size sugary soft drinks, with a margin of error of 3 percentage points and a 95% confidence interval. Assume the conditions hold.

a. Report the confidence interval for the population *percentage* that opposed banning super-size sugary soft drinks in 2012, using a carefully worded sentence.

b. If the sample size were larger and the sample proportion stayed the same, would the interval be wider or narrower than the one obtained in part a?

c. If the confidence level were 99% instead of 95%, and nothing else were changed, would the interval be wider or narrower than the one obtained in part a?

d. The total adult population of the United States in 2012 was about 240 million. Suppose the population had been half that. Would this have changed any of your answers?

7.76 Friendly Neighbors A September 2015 Pew Research Center survey showed that 71% of the randomly selected people living in Asia-Pacific view Japan as a favorable neighbor. (This was up from 68% in October 2014.) Suppose the margin of error was 2 percentage points with a 95% confidence interval. Assume that the conditions hold for all parts.

a. Report the confidence interval for the population percentage that held Japan as a favorable neighbor in 2015 using a carefully worded sentence.

b. If the sample size were larger and sample proportions stayed the same, would the interval be wider or narrower than the one obtained in part a.

c. If the confidence level were 99% and nothing else changed, would the interval be narrower or wider than the one obtained in part a.

d. The total population of adults in Asia-Pacific is 36.84 billion. Suppose the population had been one-third of that. Would that have changed any of your answers?

7.77 Sample Proportion A poll on a proposition showed that we are 95% confident that the population proportion of voters supporting it is between 40% and 48%. Find the sample proportion that supported the proposition.

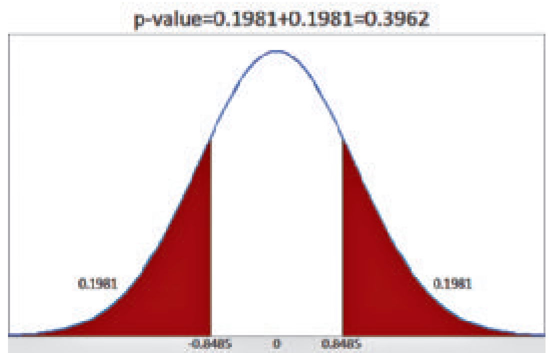
7.78 Sample Proportion A poll on a proposition showed that we are 99% confident that the population proportion of voters supporting it is between 52% and 62%. Find the sample proportion that supported the proposition.

7.79 Margin of Error A poll on a proposition showed that we are 95% confident that the population proportion of voters supporting it is between 40% and 48%. Find the margin of error.

7.80 Margin of Error A poll on a proposition showed that we are 99% confident that the population proportion of voters supporting it is between 52% and 62%. Find the margin of error.

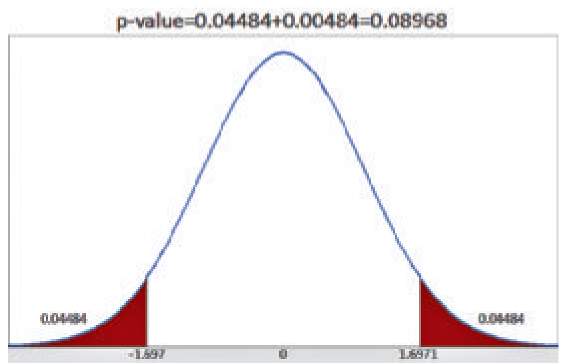
7.81-7.82

(A)



28 heads out of 50 tosses

(B)



112 heads out of 200 tosses

Figure (A) shows the probability of getting 28 or more heads OR 22 or fewer heads in 50 tosses of a fair coin. Figure (B) shows the probability of getting 112 or more heads OR 88 or fewer heads in 200 tosses of a fair coin. We will refer to the shaded areas as “tail areas.” Both show the probability of getting more than 56% or less than 44% heads. Use these figures to help you complete Exercises 7.81 and 7.82. below.

* **7.81 Tail Areas** Refer to figures (A) and (B) above.

a. Compare the percentages of heads in figures (A) and (B).

b. Calculate the standard error for the proportion of heads when tossing a fair coin 50 times. Then find the standard error for the proportion of heads when tossing a fair coin 200 times.

c. When using the CLT for one sample proportion, If you increase the sample size, the standard error will _______________ (increase or decrease), which makes the z-score _________________________ _________ (closer to zero or farther from 0), and that makes the tail area(s) _______________ (larger or smaller).

* **7.82 Tail Areas, Again** Refer to figures (A) and (B) on the previous page.

When using the CLT for one sample proportion, if you decrease the sample size, the standard error _______________ (increases or decreases), which makes the z-score _______________________ ___________ (closer to zero or farther from 0), and that makes the tail area(s) _______________ (larger or smaller).

7.83 Dreaming in Color According to studies done in the 1940s, 29% of people dream in color. Assuming this is still true, find the probability that in a random sample of 200 independent people, 50% or more will report dreaming in color. Start by checking the conditions to see whether the Central Limit Theorem applies.

7.84 Hand Washing Ignaz Semmelweiss (1818–1865) was the doctor who first encouraged other doctors to wash their hands with disinfectant before touching patients. Before the new procedure was established, the rate of infection at Dr. Semmelweiss's hospital was about 10%. Afterwards the rate dropped to about 1%. Assuming the population proportion of infections was 10%, find the probability that the sample proportion will be 1% or less, assuming a sample size of 200. Start by checking the conditions required for the Central Limit Theorem to apply.

* **7.85 Exercise vs. Meditation** During a health survey, a poll estimated that 28% of the surveyed people exercise daily and 26% of the surveyed people meditate daily. However, the surveyors claimed that the options were in a "statistical tie." Explain what this means at the level of someone who has taken an introductory statistics course.

7.86 Sampling Error In March 2012, President Obama's approval ratings were the highest they had been in almost a year: 46% of those that Gallup sampled approved of the job Obama was doing as president (http://www.gallup.com/poll viewed 4/9/12). Gallup reports a margin of *sampling* error of "plus or minus 1 percentage point." Explain in everyday terms what this means. Assume a 95% confidence level.

7.87 Young Women's Career Goals Are the career goals of young women changing? A Pew Poll asked young women about the importance they place on having a high-paying career or profession. Those women who said being successful in a high-paying career is "one of the most important things" or "very important" in their lives are included in the row labeled Career in the table. Pew asked this question in 1997 and again in 2011.

a. Find and compare the sample proportions of those who had high-paying career goals in 2011 and those who had such goals in 1997.

b. Find a 95% confidence interval for the difference in proportions who said a high-paying career was important, state whether the interval captures 0, and explain what that means. Also, if applicable, state clearly whether the women in 2011 were more or less career-oriented than those in 1997. Assume that the conditions necessary for using the confidence interval are met.

	1997	2011
Career	105	403
No	83	207

7.88 Young Men's Career Goals Are the career goals of young men changing? A Pew Poll asked young men about the importance they place on having a high-paying career or profession. Those men who said being successful in a high-paying career is "one of the most important things" or "very important" in their lives are included in the row labeled Career in the table. Pew asked this question in 1997 and again in 2011.

a. Find and compare the sample proportions of those who had high-paying career goals in 2011 and those who had such goals in 1997.

b. Find a 95% confidence interval for the difference in proportions who said a high-paying career was important, state whether the interval captures 0, and explain what that means. Assume that the conditions necessary for using the confidence interval are met.

	1997	2011
Career	113	415
No	82	288

7.89–7.90 Sample Size from Margin of Error The margin-of-error term in the confidence interval for one proportion can be solved for n. If we assume 95% confidence, use a z-value of 2, and assume the largest possible standard error to get a conservative answer, the formula simplifies to $n = \frac{1}{m^2}$.

The resulting value for n is the approximate sample size required to get a margin of error of m, assuming the confidence level is 95%. The margin of error must be entered as a decimal, so a margin of error of 2 percentage points is entered as 0.02.

* **7.89 Voters Poll: Sample Size** A polling agency wants to determine the sample size required to get a margin of error of no more than 3 percentage points (0.03). Assume the pollsters are using a 95% confidence level. How large a sample should they take? See above for the formula.

* **7.90 Ratio of Sample Sizes** Find the sample size required for a margin of error of 3 percentage points, and then find one for a margin of error of 1.5 percentage points; for both, use a 95% confidence level. Find the ratio of the larger sample size to the smaller sample size. To reduce the margin of error to half, by what do you need to multiply the sample size?

7.91 Criticize the Sampling Marco is interested in whether Proposition P will be passed in the next election. He goes to the university library and takes a poll of 100 students. Since 58% favor Proposition P, Marco believes it will pass. Explain what is wrong with his approach.

7.92 Criticize the Sampling Maria opposes capital punishment and wants to find out if a majority of voters in her state support it. She goes to a church picnic and asks everyone there for their opinion. Because most of them oppose capital punishment, she concludes that a vote in her state would go against it. Explain what is wrong with Maria's approach.

7.93 Random Sampling? If you walked around your school campus and asked people you met how many keys they were carrying, would you be obtaining a random sample? Explain.

7.94 Biased Sample? You want to find the mean weight of the students at your college. You calculate the mean weight of a sample of members of the football team. Is this method biased? If so, would the mean of the sample be larger or smaller than the true population mean for the whole school? Explain.

* **7.95 Bias?** Suppose that, when taking a random sample of 4 from 123 women, you get a mean height of only 60 inches (5 feet). The procedure may have been biased. What else could have caused this small mean?

* **7.96 Bias?** Four women selected from a photo of 123 were found to have a sample mean height of 71 inches (5 feet 11 inches). The population mean for all 123 women was 64.6 inches. Is this evidence that the sampling procedure was biased? Explain.

GUIDED EXERCISES

g * **7.39 The Oregon Bar Exam** According to the Oregon Bar Association, approximately 65% of the people who take the bar exam to practice law in Oregon pass the exam.

QUESTION Find the approximate probability that at least 67% of 200 randomly sampled people who take the Oregon bar exam will pass it.

Step 1 ▶ Population proportion
The sample proportion is 0.67. What is the population proportion?

Step 2 ▶ Check assumptions
Because we are asked for an approximate probability, we might be able to use the Central Limit Theorem. In order to use the Central Limit Theorem for a proportion, we must check the assumptions.

a. Randomly sampled: Yes
b. Sample size: If a simple random sample of 200 independent people take the bar exam, how many of them would you expect to pass, on the average? Calculate n times p. Also calculate how many you would expect to fail, n times $(1 - p)$. State whether these are both more than 10.
c. Assume the population size is at least 10 times the sample size, which would be at least 2000.

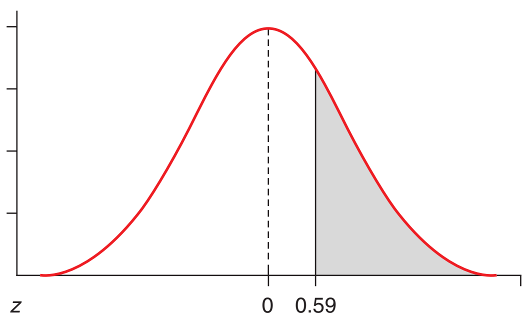
Step 3 ▶ Calculate
Part of the standardization follows. Finish it, showing all the numbers.
First find the standard error:

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.65(0.35)}{?}} = ?$$

Then standardize:

$$z = \frac{\hat{p} - p}{SE} = \frac{0.67 - 0.65}{?} = ?\text{ standard units}$$

Find the approximate probability that at least 0.67 pass by finding the area to the right of the z-value of 0.59 in the Normal curve. Show a well-labeled curve, starting with what is given in the figure below.



Step 4 ▶ Explain
Explain why the tail area in the accompanying figure represents the correct probability.

Step 5 ▶ Answer the question
Find the area of the shaded region in this figure.

g **7.67 Perry Preschool and Graduation from High School** The Perry Preschool Project was created in the early 1960s by David Weikart in Ypsilanti, Michigan. In this project, 123 African American children were randomly assigned to one of two groups: One group enrolled in the Perry Preschool and the other did not. Follow-up studies were done for decades. One research question was whether attendance at preschool had an effect on high school graduation. The table shows whether the students graduated from regular high school or not and includes both boys and girls (Schweinhart et al. 2005).

	Preschool	No Preschool
Grad HS	37	29
No Grad HS	20	35

QUESTION Find a 95% confidence interval for the difference in proportions, and interpret it. Follow the steps below.

Step 1 ▶ Calculate percentages
Looking at the children who went to preschool, 37/57, or 64.9%, graduated from high school. Looking at the children who did not go to preschool, what percent graduated from high school?

Step 2 ▶ Compare
In this sample, are the children who attend preschool more or less likely to graduate than the children who don't attend preschool?

Step 3 ▶ Verify conditions
Although we don't have a random sample of children, we do have random assignment to groups, and the two groups are independent.

We must verify that the sample sizes are large enough.

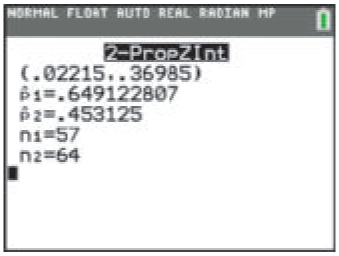
$n_1\hat{p}_1 = 57(0.649) = 37$
$n_1(1 - \hat{p}_1) = 57(0.351) = 20$
$n_2\hat{p}_2 = 64(0.453) = ??$
$n_2\hat{p}_2 = ??$

As you can see, because we are using the estimated values of $\hat{p}$ (p-hat), for our expected values we simply get the numbers in the table. Do the remaining two calculations showing that you get 29 and 35. Because all of these numbers are larger than 10, we can proceed.

Step 4 ▶ Calculate intervals

Refer to the TI-84 output, and fill in the blanks in this sentence:

I am 95% confident that the difference in proportions graduating (Preschool rate minus No Preschool rate) is between ____ and ____. You may round each number to three decimal digits.



Step 5 ▶ Draw conclusions

Indicate which one of the following statements is correct?

i. The interval does not capture 0, suggesting that it is plausible that the proportions are the same.
ii. The interval does not capture 0, suggesting that it is not plausible that the proportions are the same.
iii. The interval captures 0, suggesting that it is plausible that the population proportions are the same.
iv. The interval captures 0, suggesting that it is not plausible that the population proportions are the same.

Step 6 ▶ Generalize

Can we generalize to a larger population from this data set? Why or why not?

Step 7 ▶ Determine causation

Can we conclude from this data set that preschool caused the difference? Why or why not?

TechTips

For generating random numbers, see the TechTips for Chapter 5.

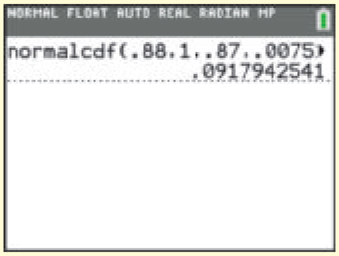
EXAMPLE A: ONE-PROPORTION PROBABILITY USING NORMAL TECHNOLOGY ► A U.S. government survey in 2007 said that 87% of young Americans earn a high school diploma. If you took a simple random sample of 2000 young Americans, what is the approximate probability that 88% or more of the sample will have earned their high school diploma? To use the "Normal" steps from Chapter 6, we just need to evaluate the mean and standard error (the standard deviation). The population mean is 0.87. The standard error is

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.87(1-0.87)}{2000}} = \sqrt{\frac{0.1131}{2000}} = 0.00752$$

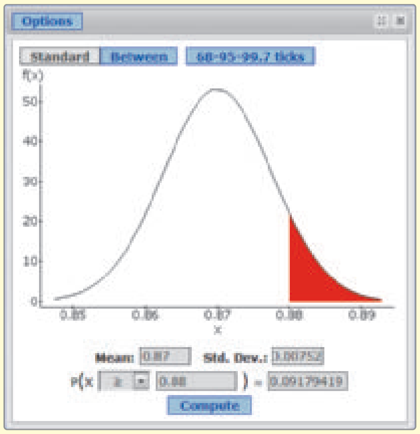
Now you can use the "Normal" TechTips steps (for TI-84, Minitab, Excel, or StatCrunch) from Chapter 6 starting on page 319.

Figure 7A shows TI-84 output, and Figure 7B shows StatCrunch output.

Thus the probability is about 9%.



▲ **FIGURE 7A** TI-84 Normal Output



▲ **FIGURE 7B** StatCrunch Normal Output

EXAMPLE B: FINDING A CONFIDENCE INTERVAL FOR A PROPORTION ► Find a 95% confidence interval for a population proportion of heads when obtaining 22 heads in a sample of 50 tosses of a coin.

EXAMPLE C: FINDING A CONFIDENCE INTERVAL FOR (THE DIFFERENCE OF) TWO PROPORTIONS ► Find a 95% confidence interval for the difference between the two population proportions using the Perry Preschool Project results. The project reported that 37 of 57 children sampled who had attended preschool graduated from high school, whereas 29 of 64 children sampled who had not attended preschool graduated from high school.

EXAMPLE D: USING SIMULATIONS TO UNDERSTAND THE BEHAVIOR OF ESTIMATORS ► Use simulated sampling on a population of 1000, consisting of 250 cat lovers and 750 dog lovers, to demonstrate the effects of sample size on sample proportion as an estimator of population proportion.

TI-84

Confidence Interval for One Proportion

1. Press **STAT**, choose **Tests, A: 1-PropZInt**.
2. Enter: **x, 22**; **n, 50**; **C-level**, **.95**.
3. Press **ENTER** when **Calculate** is highlighted.

Figure 7C shows the 95% confidence interval of (0.302, 0.578).

◄ **FIGURE 7C** TI-84 Output for a One-Proportion *Z*-interval

Confidence Interval for Two Proportions

1. Press **STAT**, choose **Tests**, **B: 2-PropZInt**.
2. Enter **x1: 37**; **n1: 57**; **x2: 29**; **n2: 64**; **C-Level: .95**.
3. Press **Enter** when **Calculate** is highlighted.
4. The output screen shows a 95% CI of (0.02215, 0.36985).

MINITAB

Confidence Interval for One Proportion

1. **Stat > Basic Statistics > 1 Proportion**.
2. Choose **Summarized data**.

   Enter: **Number of events**, **22**; **Number of trials**, **50**.
3. Click **Options** and, in the **Method:** box, select **Normal approximation**.
4. In the **Confidence level:** box, if you want a confidence level different from 95%, change it.
5. Click **OK**: click **OK**.

The relevant part of the output is

95% CI
(0.302411, 0.577589)

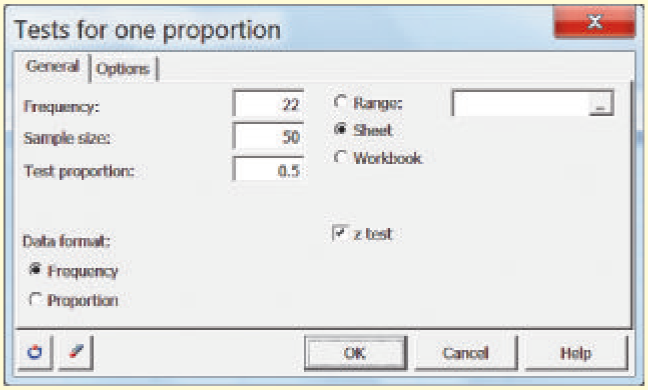
Confidence Interval for Two Proportions

1. **Stat** > **Basic Statistics** > **2 Proportions**.
2. Choose **Summarized data**. Enter: **Sample 1**, **Number of events**, **37**, **Number of trials**, **57; Sample 2**, **Number of events**, **29**, **Number of trials**, **64**. Click **OK**.
3. The output shows a 95% CI for difference of (0.0221483, 0.369847).

EXCEL

Confidence Interval for One Proportion

1. Click **Add-Ins**, **XLSTAT**, **Parametric Test**, **Tests for one proportion**.
2. See Figure 7D. Enter: **Frequency, 22; Sample size**, **50**; **Test Proportion**, **0.5.** Leave the other checked options as given and click **OK**.



▲ **FIGURE 7D** XLSTAT Input for One Proportion

(If you wanted an interval other than 95%, you would click Options and change the significance level. For a 99% interval you would use 1. For a 90% interval you would use 10.)

The relevant part of the output is

95% confidence interval on the proportion (Wald):
(0.302, 0.578)

Confidence Interval for Two Proportions

1. Click **Add-Ins**, **XLSTAT**, **Parametric Test**, **Test for two proportions**.
2. Enter: **Frequency 1: 37**; **Sample size 1: 57**; **Frequency 2: 29**; **Sample size 2: 64**.
3. Leave other checked options as given, and click **OK**.
4. The relevant part of the output is

95% confidence interval on the difference between the proportions:
(0.0221, 0.36989)

STATCRUNCH

Confidence Interval for One Proportion

1. **Stat** > **Proportion Statistics** > **One Sample** > **With Summary**
2. Enter: **# of successes, 22; # of observations, 50**.
3. Select the **Confidence Interval** option.

   Leave the default 0.95 for a 95% interval or change the **Level**. For **Method**, leave the default **Standard-Wald**.
4. Click **Compute!**

The relevant part of the output is shown. "L. Limit" is the lower limit of the interval, and "U. Limit" is the upper limit of the interval.

L. Limit	U. Limit
0.30241108	0.5775889

Confidence Interval for Two Proportions

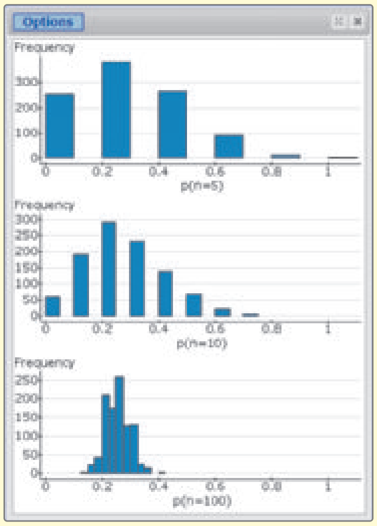
1. **Stat** > **Proportion Statistics** > **Two Sample** > **With Summary.**
2. Enter: **Sample 1: # of successes: 37; # of observations: 57; Sample 2: # of successes: 29; # of observations: 64.**
3. Select **Confidence interval for $p_1 - p_2$,** leave the **Level** at **0.95,** and click **Compute!**
4. The output will show **L. Limit** (for lower limit of confidence interval) **0.022148349**, and **U. Limit** (for upper limit of confidence interval) **0.36984727**.

Simulated Sampling

Enter the population data. This can be done either by hand or by loading the data file from the CD accompanying this text. To enter by hand, you can "brute-force it" by entering **Cat** in first 250 rows of the first column, which is labeled **var1** and then entering **Dog** in the next 750 rows. But *if you do it carefully*, it is quicker to do **Data** > **Compute Expression** and then type the expression **concat(rep("cat",250),rep("dog",750)).** Change the column label to **Lover**. Click **Compute!**

1. In this step you will generate 1000 samples of size 5 from the population, Lover; then you will count and list the number of cat lovers in each sample in a new column labeled Cfive. **Data** > **Sample**. **Select columns: Lover**. Enter **Sample size: 5**, **Number of samples: 1000**. Select **Compute statistic for each sample**. *Now be careful*! Enter **sum("Sample(Lover)" = "Cat")**. Enter **Column name: Cfive**. Click **Compute!**
2. In this step you will calculate the proportion of cat lovers in each sample and then list the 1000 sample proportions in a new column labeled p(n = 5). **Data** > **Compute Expression**. Enter **Expression: Cfive/5**. Enter **Column label: p(5 = n).** Click **Compute!**

3. To generate a column of 1000 sample proportions of sample size 10, repeat steps 2 and 3, but substitute **Sample size: 10**, **Column name: Cten**, **Expression: Cten/10**, and **Column label: p(n = 10).**
4. To generate a column of 1000 sample proportions of sample size 100, repeat steps 2 and 3, but change to **Sample size: 100, Column name: Chundred, Expression: Chundred/100, and Column label: p(n = 100).**
5. Now graph the results. **Graphs > Histogram**. Select **p(n = 5)**, **p(n = 10)**, and **p(n = 100);** do this by holding the **Ctrl** key down while clicking on **p(n = 5)**, **p(n = 10)**, and **p(n = 100).** Scroll down and check **Draw horizontal grid lines** and **Use same X-axis**. Set **Rows per page** to **3.** Click **Compute!** Your output should look somewhat like Figure 7E.



▲ **FIGURE 7E** StatCrunch Output

8 Hypothesis Testing for Population Proportions

THEME

In many scientific and business contexts, decisions must be made about the values of population parameters, even though our estimates of these parameters are uncertain. Hypothesis testing provides a method for making these decisions, while controlling for the probability of making certain types of mistakes.

In science, business, and everyday life, we often have to make decisions on the basis of incomplete information. For example, a chewing gum company might need to decide what percent of customers will buy a new flavor. The company can't test the new flavor on everyone in the country—it has to base its decision on a small sample. An educational psychologist wonders whether kids who receive music training become more creative than other children. The test she uses to measure creativity does in fact show an increase, but might this increase be explained by chance alone? A sample of people who watched violent TV when they were children turn out to exhibit more violent behavior than a comparison sample made up of people who did not watch violent TV when young. Could this difference be due to chance, or is something else going on?

Hypothesis testing is a type of statistical inference. In Chapter 7, we used confidence intervals to estimate parameters and provide a margin of error for our estimates. In this chapter we make decisions on the basis of the information provided by our sample. If we knew everything about the population, we would definitely know what decision to make. But seeing only a sample from the population makes this decision harder, and mistakes are inevitable. Just as we measured our uncertainty in Chapter 7, our next task is to measure our mistake rate when testing hypotheses. In this chapter, we continue to work with population proportions. In the next chapter, we'll see how to find confidence intervals and perform hypothesis tests for means.

CASE STUDY

Dodging the Question

Have you ever become frustrated watching a politician in a debate, or maybe in an interview, completely dodge a difficult question? One way a politician might do this is by answering a different, easier question. Researchers have found that we are not always good at noticing when a question is dodged, and that our attitudes toward the speaker and our situation help determine whether we are good "dodge detectors." In order to help viewers become better dodge detectors while watching political debates, some TV stations have started posting, at the bottom of the TV screen, the question that was asked. Does this practice really improve people's ability to notice question dodging?

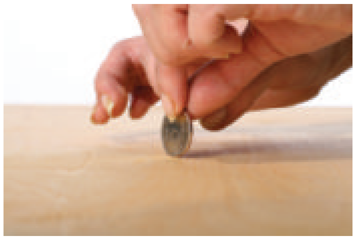
Some researchers filmed hired actors to play politicians having a formal debate. In this video, one of the "politicians" is asked a question about health care—but he dodges the question by instead answering a question about illegal drugs. Test subjects were recruited to watch the video. Half of the subjects, randomly chosen, saw the video with the question about health care posted at the bottom of the screen. The other half saw nothing on the bottom of the screen. For those who saw the question posted, 88% noticed the dodge. For those who did not see the question posted, only 39% noticed that the politician evaded the question (Rogers and Norton 2011).

The difference between these two groups is 49 percentage points. Does this difference occur because we really and truly listen differently when the question is posted on the screen? Or might this difference be due merely to chance? After all, the subjects were randomly assigned either to see the post or not to see the post, so there was an element of chance involved. In other words, is it possible that even if posting the question had no effect at all, we would see a difference as large as 49 percentage points?

In this chapter we'll show you how to make this decision: Is the difference real, or could it be due to chance? At the end of the chapter, we'll return to this case study and see what decision the researchers made and why they made it.

SECTION 8.1

The Essential Ingredients of Hypothesis Testing



Football games and tennis matches begin with a coin toss to determine which team or player gets to play offense first. Coin tosses, in which the coin is flipped high into the air, are used for this because flipping a coin is believed to be fair. "Fair" means the coin is equally likely to land heads or tails, so both sides have an equal chance of winning.

But what if we spin the coin (on a hard, flat surface) rather than flipping it in the air? We claim that, because the "heads" side of a coin bulges out, the lack of symmetry will cause the spinning coin to land on one side more often than on the other. In other words, we believe a spun coin is not fair. Some people—maybe most—will find this claim to be outrageous and will insist it is false.

Suppose for the moment that, as evidence of our claim, we take a coin and spin it 20 times and that we get 20 heads. Reflect on what your reaction would be. We're betting that you'd be surprised. In your personal experience, this sort of thing, 20 heads in 20 spins, is rare when dealing with fair coins. You were expecting about 10 heads, give or take, because you didn't really believe that spinning the coin would matter. The fact that all 20 spins landed heads means that something surprising has happened, and you will probably think that we were right—that spun coins are biased.

If that describes your thought process, then you've just informally done a **hypothesis test**. Hypothesis testing is a procedure that enables us to choose between two claims when we have variability in the outcomes. We will teach a particular procedure that we call formal hypothesis testing. We call this procedure formal because it is based on particular terminology and four well-specified steps. However, we hope to show you that this "formal" procedure is not too different from the common sense you just applied when thinking about getting 20 heads in 20 spins.

The formal procedure looks like this:

- In the first step, you state a hypothesis—a claim about the world—which, in this case, was the claim that the spun coin is biased. In a formal hypothesis test, this claim will be weighed against a neutral, skeptical claim, which in this case would be the claim that a spun coin is *not* biased.
- The second step in the formal process is a preparation step: You determine how you'll use data to make your decision and make sure you have enough data to minimize the probability of making mistakes.
- In the third step, you collect data and compare them to your expectations. When you saw that the sample proportion of heads in 20 spins was 1.0 (20 heads out of 20 spins), you mentally compared this to your preconceived notion that the sample proportion was going to be close to 0.5. You might have even computed the probability of getting such an extreme outcome if, in fact, spinning a coin

were fair. (The probability of getting either 20 heads or 0 heads for a fair coin is about 0.000002, about 2 in one million. That's why you would feel surprised if all 20 spins were heads and you were expecting a 50–50 outcome.)

- In the fourth and final step of a formal hypothesis test, you state your conclusion: Do you believe the claim, or do you find that the claim doesn't have enough evidence to back it? If we did in fact get 20 heads in 20 spins, we would be willing to conclude that spinning a coin is biased.

The example we have just walked through has an extreme outcome: exactly 100% heads. But what about in-between outcomes? What if we saw 7 out of 20 heads (0.35)? This outcome is less than 0.5, but is it so much less that we would think spinning the coin is biased? Hypothesis testing helps us make decisions for in-between cases such as this.

Before going further, take out a coin. Spin it 20 times, and record the number of heads. We're going to illustrate the basic concepts of hypothesis testing using this "experiment." When *we* did this, we got 7 heads, and we'll use this outcome in the following discussion.

We're first going to show you some important concepts that are the essential ingredients in the recipe for preparing hypothesis tests: hypotheses, minimizing mistakes, a test statistic, and surprise. Then, in Section 8.2, we will show how these ingredients are combined into the four steps of a formal hypothesis test. Section 8.3 covers some details of calculations and some issues that help you better interpret the results of hypothesis tests. The final section shows how our template for hypothesis testing can be expanded to fit the new context of comparing two population proportions.

Details

Pennies
In fact, spinning a coin *does* produce bias, but as far as we know, only for some coins. If you can find a 1962 penny, you might need to spin it about 50 times, but you should see reason to reject the null hypothesis.

Main Ingredient: A Pair of Hypotheses

In a formal hypothesis test, hypotheses are statements about population parameters. A hypothesis begins as a statement about the real world, but it must then be rephrased in terms of population parameters.

For example, in the coin-spinning example, our claim about the world is that a spun coin is not fair. We now must restate this in terms of a parameter.

The parameter we are making a claim about is the probability of the coin landing heads. Let's call that probability p. For a fair coin, $p = 0.5$. If the coin is not fair, then p is *not* equal to 0.5. In symbols, $p \neq 0.5$. These are statements about the population parameter p, the probability of getting heads.

The claim that the coin is fair is probably not that outlandish to you. If you're even a little bit skeptical, it's probably what you already believe. Such a hypothesis is called a **null hypothesis**. Our null hypothesis is that $p = 0.50$.

On the other hand, the hypothesis that we hope to convince you and the world is true is called the **alternative hypothesis**. Ours is $p \neq 0.5$.

Hypotheses will come in pairs like the pair you've just seen:

> The **null hypothesis**, which we write H_0 (and pronounce "H-naught" or simply "the null hypothesis"), is the neutral, status quo, skeptical statement about a population parameter. In the context of researching new ideas, the null hypothesis often represents "no change," "no effect," or "no difference." In this text, the null hypothesis will always have an = sign.
>
> The **alternative hypothesis**, H_a (pronounced "H-A"), is the research hypothesis. It is a statement about the value of a parameter that we intend to demonstrate is true.

For our coin-spinning experiment, we write the hypotheses as

$$H_0: p = 0.5$$
$$H_a: p \neq 0.5$$

We emphasize that in a formal hypothesis test, hypotheses are *always* statements about population parameters. In this case, our population consists of infinitely many spins of our coin, and p represents the probability of getting heads.

Details

Definition of *Hypothesis*
Merriam-Webster's online dictionary defines a hypothesis as "a tentative claim made in order to draw out and test its logical or empirical consequences."

Looking Back

Statistic vs. Parameter
In Chapter 7 you learned that $\hat{p}$, a statistic, represents the proportion of successes in a sample, whereas p, a parameter, represents the proportion of successes in the population.

Caution

Pronunciation
Do not say "Ho" or "Ha," as Santa Claus might.

KEY POINT Hypotheses are always statements about population parameters; they are never statements about sample statistics.

EXAMPLE 1 Marriage

Historically, about 70% of all U.S. adults were married. A sociologist who believes that marriage rates in the United States have declined will take a random sample of U.S. adults and record whether or not they are married.

QUESTIONS

a. State the null and alternative hypotheses in words.

b. Let p represent the proportion of all adults in the United States who are married. State the null and alternative hypotheses in terms of this population parameter.

SOLUTIONS

The alternative hypothesis is the claim that the sociologist wishes to make, and the null hypothesis is the skeptical claim that things have not changed.

a. Null hypothesis: The same proportion of adults are married now as in the past. Alternative hypothesis: The proportion of adults who are married now is *less than* it was in the past.

b. Because the historical proportion is $p = 0.70$, we have

$$H_0: p = 0.70$$
$$H_a: p < 0.70$$

TRY THIS! Exercise 8.3

The alternative hypothesis in Example 1 has a "less than" sign, rather than the "not equal to" sign that we used in our coin-spinning experiment. This is because the alternative hypothesis must reflect the claim made about the real world. Here, the sociologist wants to establish that the proportion of married adults has *decreased* from its previous value. Therefore, he believes the proportion of married adults is less than the historical value of 0.70. Such hypotheses are called **one-sided hypotheses**. For the coin-spinning experiment, we wanted to show only that the spun coin is *not* fair; we didn't care whether it was biased towards heads or biased towards tails. A hypothesis with a "not equal to" sign is called a **two-sided hypothesis**.

There are three basic pairs of hypotheses, as Table 8.1 shows.

TABLE 8.1 Three pairs of hypotheses that can be used in a hypothesis test.

Two-Sided	One-Sided (Left)	One-Sided (Right)
$H_0: p = p_0$	$H_0: p = p_0$	$H_0: p = p_0$
$H_a: p \neq p_0$	$H_a: p < p_0$	$H_a: p > p_0$

EXAMPLE 2 Internet Advertising

An Internet retail business is trying to decide whether to pay a search engine company to upgrade its advertising. In the past, 15% of customers who visited the company's webpage by clicking on the advertisement bought something (this is called a "click-through"). If the business decides to purchase premium advertising, then the search engine company will make that company's ad more prominent.

The search engine company offers to do an experiment: For one day, customers will see the retail business's ad in the more prominent position. The retail business can then decide whether the advertising improves the percentage of click-throughs. The retailer agrees to the experiment, and when it is over, 17% of the customers have bought something.

A marketing executive wrote the following hypotheses:

$$H_0\colon \hat{p} = 0.15$$
$$H_a\colon \hat{p} = 0.17$$

where $\hat{p}$ represents the proportion of the sample that bought something from the website.

QUESTION What is wrong with these hypotheses? Rewrite them so that they are correct.

SOLUTION First, these hypotheses are written about the sample proportion, $\hat{p}$. We *know* that 17% of the sample bought something, so there is no need to make a hypothesis about it. What we don't know is what proportion of the entire population of people who will click on the advertisement will purchase something. The hypotheses should be written in terms of p, the proportion of the population that will purchase something.

A second problem is with the alternative hypothesis. The research question that the company wants to answer is not whether 17% of customers will purchase something. It wants to know whether the percentage of customers who do so has increased over what has happened in the past.

The correct hypotheses are

$$H_0\colon p = 0.15$$
$$H_a\colon p > 0.15$$

where p represents the proportion of all customers who click on the advertisement and purchase a product.

TRY THIS! Exercise 8.9

Hypothesis tests are like criminal trials. In a criminal trial, two hypotheses are placed before the jury: The defendant either is not guilty or is guilty. These hypotheses are not given equal weight, however. The jury is told to assume the defendant is not guilty until the evidence overwhelmingly suggests this is not so. (Defendants charged with a crime in the United States must be found guilty "beyond all reasonable doubt.")

Hypothesis tests follow the same principle. The statistician plays the role of the prosecuting attorney, who hopes to show that the defendant is guilty. The hypothesis that the statistician or researcher hopes to establish plays the role of the prosecutor's charge of guilt. The null hypothesis is chosen to be a neutral, noncontroversial statement (such as the claim that the defendant is not guilty). Just as in a jury trial, where we ask the jury to believe that the defendant is not guilty unless the evidence against this belief is overwhelming, we will believe that the null hypothesis is true in the beginning. But once we examine the evidence, we may reject this belief if the evidence is overwhelmingly against it.

Details

p Can Be Either a Proportion or a Probability

The parameter p can represent both. For example, if we were describing the voters in a city, we might say that $p = 0.54$ are Republican. Then p is a proportion. But if we selected a voter at random, we might say that the probability of selecting a Republican is $p = 0.54$.

KEY POINT The null hypothesis always gets the benefit of the doubt and is assumed to be true throughout the hypothesis-testing procedure. If we decide at the last step that the observed outcome is extremely unusual under this assumption, then *and only then* do we reject the null hypothesis.

The most important step of a formal hypothesis test is choosing the hypotheses. In fact, there are really only two steps of a formal hypothesis test that a computer cannot do, and this is one of those steps. (The other step is checking to make sure that the conditions necessary for the probability calculations to be valid are satisfied. Also, computers can't interpret the findings, as you will be asked to do.)

Add In: Making Mistakes

Mistakes are an inevitable part of the hypothesis-testing process. The trick is not to make them too often.

One mistake we might make is to reject the null hypothesis when it is true. For example, even a fair coin *can* fall heads in 20 out of 20 flips. If that happened, we might conclude that the coin was unfair when it really was fair. We can't prevent this mistake from happening, but we can try to make it happen infrequently.

The **significance level** is the name for a special probability: It is the probability of making the mistake of rejecting the null hypothesis when, in fact, the null hypothesis is true. The significance level is such an important probability that it has its own symbol, the Greek lowercase alpha: α.

In our experiment with spinning the coin, the significance level is the probability that we will conclude that spinning a coin is *not* fair when, in fact, it really *is* fair. In a criminal justice setting, the significance level is the probability that we conclude that the suspect is guilty when he is actually innocent.

EXAMPLE 3 Significance Level for Internet Advertising

In Example 2, an Internet retail business gave a pair of hypotheses about p, the proportion of customers who click on an advertisement and then purchase a product from the company. Recall that in the past, the proportion of customers who bought the product was 0.15, and the company hopes this proportion has increased. It intends to test these hypotheses with a significance level of 5%. In other words, $\alpha = 0.05$.

$$H_0: p = 0.15$$
$$H_a: p > 0.15$$

QUESTION Describe the significance level in context.

SOLUTION The significance level is the probability of rejecting H_0 when in fact it is true. In this context, this means that the probability is 5% that the company will conclude that the proportion of customers who will buy its product is bigger than 0.15 when, in fact, it is 0.15.

TRY THIS! Exercise 8.11

Naturally, we want a procedure with a small significance level, because we don't want to make mistakes too often. How small? Most researchers and statisticians use a significance level of 0.05. In some situations it makes sense to allow the significance level to be bigger, and some situations require a smaller significance level. But $\alpha = 0.05$ is a good place to start.

KEY POINT The significance level, α (Greek lowercase alpha), represents the probability of rejecting the null hypothesis when the null hypothesis is true. For many applications, $\alpha = 0.05$ is considered acceptably small, but 0.01 and 0.10 are also sometimes used.

Mix with: The Test Statistic

A **test statistic** compares the real world with the null hypothesis world. It compares our observed outcome with the outcome the null hypothesis says we should see.

For example, in our coin-spinning experiment, we saw 7 out of 20 heads, for a proportion of 0.35 heads. The null hypothesis tells us that we should expect half to be heads: 0.5. The test statistic tells us how far away our observation, 0.35, is from the null hypothesis value, 0.5.

To do this comparison, we use the **one-proportion *z*-test statistic**.

Formula 8.1: The one-proportion *z*-test statistic

$$z = \frac{\hat{p} - p_0}{SE}, \text{ where } SE = \sqrt{\frac{p_0(1 - p_0)}{n}}$$

The symbol p_0 represents the value of p that the null hypothesis claims is true. For example, for the coin-spinning example, p_0 is 0.5. Most of the other test statistics you will see in this text have the same structure as Formula 8.1:

$$z = \frac{\text{observed value} - \text{null value}}{SE}$$

For our coin-spinning example, because we observed 0.35 heads in 20 spins, the observed value of our test statistic is

$$z_{\text{observed}} = \frac{0.35 - 0.50}{\sqrt{\dfrac{0.50(1 - 0.50)}{20}}} = \frac{-0.15}{\sqrt{0.0125}} = \frac{-0.15}{0.1118} = -1.34$$

Looking Back

Standard Error
The standard error of the sample proportion, given in Chapter 7, is

$$SE = \sqrt{\frac{p(1 - p)}{n}}$$

Formula 8.1 uses the symbol p_0 to remind us to use the value that the null hypothesis claims to be correct.

Why Is the z-Statistic Useful? The *z*-test statistic has the same structure as the *z*-score introduced in Chapter 3, and it serves the same purpose. By subtracting the value the null hypothesis expects from the observed value, $\hat{p} - p_0$, we learn how far away the actual sample value was from the expected value. A positive value means the outcome was greater than what was expected, and a negative value means it was smaller than what was expected.

If the test statistic value is 0, then the observed value and the expected value are the same. This means we have little reason to doubt the null hypothesis. The null hypothesis tells us that the test statistic should be 0, give or take some amount. If the value is far from 0, then we doubt the null hypothesis.

By dividing this distance by the standard error, we convert the distance into "standard error" units, and we learn how many standard errors away our outcome lies from what was expected. Our spun coin resulted in a *z*-statistic of –1.34. This tells us that we saw fewer heads than the null hypothesis expected, and that our sample proportion was 1.34 standard errors below the null hypothesis proportion of 0.5.

Looking Back

***z*-Scores**
A *z*-score has the structure

$$\frac{\textit{observed value} - \textit{mean}}{\textit{standard deviation}}$$

KEY POINT If the null hypothesis is true, then the *z*-statistic will be close to 0. Therefore, the farther the *z*-statistic is from 0, the more the null hypothesis is discredited.

EXAMPLE 4 Test Statistic for e-Readers

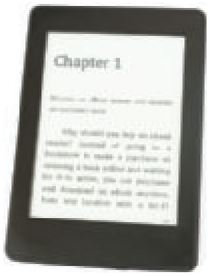
Does ownership of an e-reader (such as the Nook or Kindle) increase after the holidays? A publishing company wants to know in order to plan its advertising budget. Based partly on conventional wisdom and partly on combining results from various studies, the publisher believes that in early December 2011, about 10% of Americans owned an e-reader. In January 2012, after the holidays, the Pew Research Center conducted a survey of 1377 randomly selected adults and found that 19% of them owned an e-reader (http://libraries.pewinternet.org/2012/04/04/the-rise-of-e-reading/). Can the publisher conclude that e-reader ownership increased over the holidays? Or is the apparent change simply due to the fact that Pew surveyed only 1377 randomly selected people? If we let p represent the proportion of people in the population who owned an e-reader in January 2012, then our hypotheses are

$$H_0: p = 0.10$$
$$H_a: p > 0.10$$

QUESTION Calculate the observed value of the test statistic, and explain the value in context.

SOLUTION The observed proportion of people in the sample who owned e-readers is 0.19. Therefore, the observed value of the test statistic is

$$z_{abc} = \frac{0.19 - 0.10}{\sqrt{\frac{0.10(0.90)}{1377}}} = \frac{0.09}{0.0080845} = 11.13$$



CONCLUSION The observed value of the test statistic is 11.13. This means that while the null hypothesis claims a proportion of 0.10, the observed proportion was much higher—11.13 standard errors higher—than what the null hypothesis claims.

TRY THIS! Exercise 8.15

! Caution

Standard Errors in Hypothesis Tests
Remember that the standard error in the denominator is calculated using the *null hypothesis* value for the population proportion, *not* using the observed proportion.

The Final Essential Ingredient: Surprise!

No, the final essential ingredient in hypothesis testing is not *a* surprise. The main ingredient is surprise itself.

Surprise happens when something unexpected occurs. The null hypothesis tells us what to expect when we look at our data. If we see something unexpected—that is, if we are surprised—then we should doubt the null hypothesis, and if we are really surprised, we should reject it altogether.

Figure 8.1 shows all possible outcomes of our coin-spinning experiment in terms of the test statistic and shows the "surprising" outcomes in red. If spinning a coin is fair, as the null hypothesis claims, we can compute the probability of every outcome. A fair coin will produce about 10 heads, give or take, which is associated with a test statistic of 0. If the null hypothesis is true, getting 5 or fewer heads OR 15 or more heads, is rare. (In fact, the probability that you will get 0 to 5 OR 15 to 20 heads—the outcomes shown in red in Figure 8.1—is less than 5%.) If we had spun a coin 20 times and saw one of these red outcomes, we would be surprised and would probably reject the null hypothesis that the spun coin was fair.

Because we are statisticians, we have a way of measuring our surprise. The **p-value** is a number that measures our surprise by reporting the probability that if the null hypothesis is true, a test statistic will have a value as extreme as or more extreme than the value we actually observe. Small p-values (closer to 0) mean we have received a surprise. Large p-values (closer to 1) mean no surprise: The outcome happens fairly often. Any of the outcomes in red in Figure 8.1 has a small p-value (all are less than 0.05).

H_0 says rarely happens------>|<------------H_0 says happens often------------>|<------H_0 says rarely happens

–4.5	–4.0	–3.6	–3.1	–2.7	–2.2	–1.8	–1.3	–0.9	–0.4	0.0	+0.5	+0.9	+1.3	+1.8	+2.2	+2.7	+3.1	+3.6	+4.0	+4.5
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

▲ **FIGURE 8.1** All possible outcomes for spinning a coin 20 times in terms of the test statistic and the number of heads. The red values are those that might be considered unusual and unexpected if the null hypothesis is true. If the coin is truly fair, "red" outcomes will happen less than 5% of the time. The "black" outcomes will happen a little more than 95% of the time. The lowest row shows the outcomes in terms of the number of heads. The row above it shows the outcomes in terms of the z-statistic.

EXAMPLE 5 Judging p-Values

Suppose you spin a coin 20 times and count the number of heads. The null hypothesis is that the coin is fair. The alternative hypothesis is that the coin is not fair.

QUESTION Which of these outcomes, 3 heads or 9 heads, has the smaller p-value? Why? (You might refer to Figure 8.1.)

SOLUTION If the null hypothesis is right, then we should get about half heads. This means we expect 10 heads, give or take. Because 9 heads is much closer to 10 heads than 3 heads is to 10, 3 is a more surprising outcome and will have a smaller p-value. We can see, by referring to Figure 8.1, that the test statistic value for 3 heads is –3.1 and that this is a more extreme outcome with a smaller p-value than 9 heads (with $z = -0.4$).

TRY THIS! Exercise 8.17

KEY POINT The p-value is a probability. Assuming that the null hypothesis is true, the p-value is the probability that if the experiment were repeated, you would get a test statistic as extreme as or more extreme than the one you actually got. A small p-value suggests that a surprising outcome has occurred and discredits the null hypothesis.

EXAMPLE 6 e-Book Publishing

The publishers in Example 4 were interested in knowing whether more than 10% of their target demographic owned e-readers. They carried out a hypothesis test and found that the observed value of the test statistic was 11.13. We can calculate that the p-value associated with this is nearly 0. (In the next section, you'll learn how to calculate this.)

QUESTION Explain the meaning of the p-value in this context. If we believed that the null hypothesis was true, would we be surprised? (The null hypothesis was that $p = 0.10$, where p was the proportion of the population that owned e-readers.)

SOLUTION The p-value is very small (nearly 0). This tells us that if it is true that 10% of the population owns e-readers, then getting a test statistic as extreme as or more extreme than 11.13 is very improbable (maybe even close to impossible!). If you believed that the null hypothesis was true and $p = 0.10$, then you should be *very* surprised, because what you saw is nearly impossible.

TRY THIS! Exercise 8.19

SECTION 8.2

Hypothesis Testing in Four Steps

Now that you know the main ingredients (hypotheses, minimizing mistakes, test statistic, and surprise), it's time to learn the recipe. The hypothesis tests you'll see in this book consist of four steps that combine the ingredients you've just studied into a useful, logical structure.

The four steps are

Step 1: Hypothesize.
State your hypotheses about the population parameter.

Step 2: Prepare.
Get ready to test: State a significance level. Choose a test statistic appropriate for the hypotheses. State and check conditions required for future computations, and state any assumptions that must be made.

Step 3: Compute to compare.
Compute the observed value of the test statistic, and compare it to what the null hypothesis said you would get. Find the p-value in order to measure your level of surprise.

Step 4: Interpret.
Do you reject or not reject the null hypothesis? What does this mean in the context of the data?

A Few Details

Step 1 we have covered in detail, but the other steps still require some explanation. After we've explained, we'll show some examples of how these steps work together.

Detail for Step 2: Check Conditions to Find Probabilities The significance level and the p-value are both probabilities, and both are calculated assuming the null hypothesis is true. To be able to find these probabilities, we need to know the probability distribution for our test statistic.

Looking Back

The Sampling Distribution
Recall from Chapter 7 that the probability distribution for a statistic is called the sampling distribution.

The sampling distribution for the one-proportion test statistic is, approximately, the standard Normal distribution if the following conditions are met *and* if the null hypothesis is true:

1. *Random Sample.* The sample is collected randomly from the population.
2. *Large Sample Size.* The sample size, n, is large enough that the sample has 10 or more expected successes and 10 or more expected failures; in other words, $np_0 \geq 10$ and $n(1 - p_0) \geq 10$.
3. *Large Population.* If the sample is collected without replacement, then the population size is at least 10 times bigger than the sample size. (If the sample is drawn with replacement, then any size population will work.)
4. *Independence.* Each observation or measurement must have no influence on any others.

KEY POINT Under the appropriate conditions, the sampling distribution of the z-statistic is approximately a standard Normal distribution, $N(0, 1)$.

It is very important that in step 2 ("Prepare") we make sure these conditions hold or that we can reasonably assume that they do. If not, then we can't find the p-value. If we can't find the p-value, there is no reason to proceed with the other steps.

EXAMPLE 7 Checking Conditions for e-Reader Poll

The observed value of the test statistic for the hypothesis test of Example 4 was 11.3. If we know that the sampling distribution for the test statistic is a standard Normal distribution, $N(0, 1)$, then we will know that this is an extremely rare outcome. The poll was based on a simple random sample of 1377 adults living in the United States, and the sample proportion who own an e-reader was 0.19. The hypotheses were

$$H_0: p = 0.10$$
$$H_a: p > 0.10$$

QUESTION Check the conditions to show that the test statistic for the e-reader survey approximately follows a standard Normal distribution.

SOLUTION We check each of the four conditions to see whether they are satisfied.

1. *Random Sample.* We were told that the sample was random.
2. *Large Sample Size.* Because $p_0 = 0.10$, we expect $1377 \times 0.10 = 137.7$ successes, which is bigger than 10. We know that if 10% of 1377 is bigger than 10, then 90% is definitely bigger, and so we know that we can expect more than 10 failures ($1377 \times 0.90 = 1239.3$). The sample size is large enough.
3. *Large Population.* The sample is a simple random sample, so it is taken without replacement. Thus we must check that the population is at least 10 times larger than the sample. The population of all U.S. adults is easily 10 times larger than 1377.
4. *Independence.* We assume that the pollsters surveyed people in such a way that their responses were independent of each other.

CONCLUSION The conditions are verified. We can use the standard Normal distribution, $N(0, 1)$, to find probabilities for the test statistic.

NOW TRY! Exercise 8.23

Detail for Step 3: Calculating the p-Value

After finding the observed value of the test statistic in step 3, we next must determine whether this value "surprises" the null hypothesis. The p-value measures this surprise. But to find the p-value, you need to know what the phrase *as extreme as or more extreme than* means when we say, "The p-value is the probability that, if the null hypothesis were true, we would get a statistic as extreme as or more extreme than the observed test statistic."

Think of our coin-spinning example. We expected a proportion of 0.50 heads, but we saw 0.35 heads. Our observed value of the test statistic was therefore found to be $z = -1.34$. We wish to find the p-value to answer this question: If we were to do this again, and if coin-spinning is really and truly fair, what's the probability that we would get a test statistic value *as extreme as or more extreme than* -1.34?

The meaning of this phrase depends on which of our three alternative hypotheses we're using.

If the alternative hypothesis is two-sided:

$H_a: p \neq p_0$ (The true value of p is either bigger or smaller than what the null hypothesis claims.)

then *as extreme as or more extreme than* means "even farther away from 0 than the value you observed." This corresponds to finding the probability in both tails of the $N(0, 1)$ distribution. This is called a **two-tailed p-value**.

Our coin-spinning example will have a two-tailed p-value because the alternative is two-sided ($p \neq 0.5$).

There are two different types of one-sided hypotheses.

If the alternative hypothesis is

$H_a: p < p_0$ (The true value is less than the value claimed by the null hypothesis.)

Details

Tails
Many statisticians use the terms "one-tailed hypotheses" and "one-sided hypotheses" interchangeably, and likewise "two-sided" and "two-tailed."

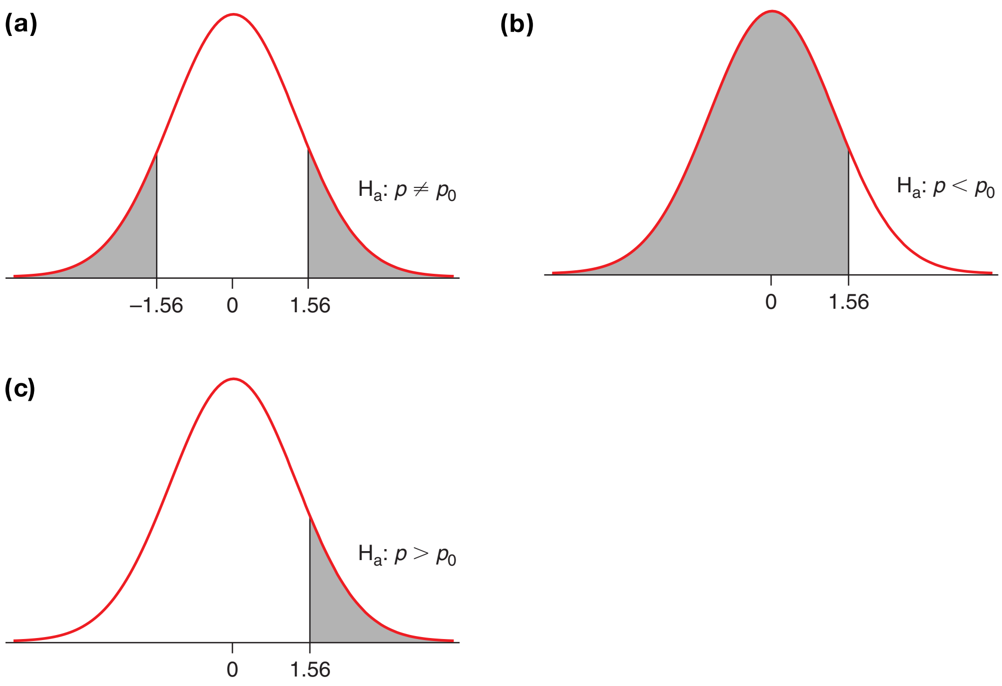
then *as extreme as or more extreme than* means "less than or equal to the observed value." This corresponds to finding the probability in the left tail of $N(0, 1)$ and so is called a left-tailed p-value. Example 1, a test to see whether the marriage rate had decreased, uses a left-tailed p-value because the alternative hypothesis is H_a: $p < 0.70$.

Finally, if the alternative hypothesis is

H_a: $p > p_0$ (The true value is greater than the value claimed by the null hypothesis.)

then *as extreme as or more extreme than* means "greater than or equal to the observed value." This corresponds to finding the probability in the right tail of $N(0, 1)$. This p-value is called a right-tailed p-value. Example 2 (H_a: $p > 0.15$) and Example 7 (H_a: $p > 0.10$) both use right-tailed p-values.

Once we've determined which extremities to use, we can use Table 2 (Normal table) in Appendix A, a statistical calculator, or other technology. These three cases are illustrated in Figure 8.2, which uses an observed test statistic value of $z = 1.56$.



▶ **FIGURE 8.2** The shaded areas represent p-values for three different alternative hypotheses when the observed value of the test statistic is $z = 1.56$. **(a)** The p-value for a two-sided alternative hypothesis (0.119). **(b)** The p-value for a left-sided hypothesis (0.941). **(c)** The p-value for a right-sided hypothesis (0.059).

EXAMPLE 8 p-Value for Coin Spinning

We claimed that spinning a coin leads to a biased outcome. Our hypotheses were

$$H_0: p = 0.5$$
$$H_a: p \neq 0.5$$

We spun the coin 20 times and saw a sample proportion of $\hat{p} = 0.35$, which led to an observed test statistic of $z_{observed} = -1.34$. Assume the required conditions are satisfied, so that the sampling distribution of the test statistic is approximately the standard Normal distribution.

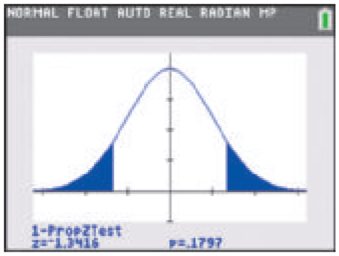
QUESTION Find the approximate p-value, and explain what it means.

SOLUTION The alternative hypothesis is two-sided, so we will find a two-tailed p-value. The situation is similar to Figure 8.2a, but with $z = -1.34$. We can use a statistical calculator, such as the Normal Calculator in StatCrunch or the TI-84, or Table 2 in Appendix A. Figure 8.3 shows the output from a TI-84.

The p-value is 0.1797.

CONCLUSION The p-value is about 0.18. This tells us that if spinning a coin is fair, then the probability of seeing an outcome as extreme as or more extreme than 0.35 heads is about 18%.

NOW TRY! Exercise 8.25



▲ **FIGURE 8.3** TI-84 output showing the probability that a test statistic will be as extreme or more extreme than -1.34 if the null hypothesis is true.

Compare the approximate p-value of 0.18 that we calculated in Example 8 with Figure 8.1. There we see that the test statistic value of -1.34 falls into the "black" region, which indicates an outcome that is not surprising if the coin is fair. Indeed, our p-value of 18% is relatively large: For fair coins, outcomes like this happen about 1 time in 5.

Detail for Step 4: Making a Decision In the final step, you must make a decision between hypotheses and explain what the decision means. How, then, do we choose between the two hypotheses? The p-value measures our surprise (or lack of surprise) at the outcome of our test, but what should we do about this number? When is the outcome so unusual that we should reject the null hypothesis?

We apply a simple rule: Reject the null hypothesis if the p-value is smaller than (or equal to) the value chosen for the significance level, α. If the p-value is larger than the significance level, do not reject the null hypothesis. For most applications, this means you reject the null hypothesis if the p-value is less than or equal to 0.05.

Following this rule ensures that our significance level is achieved. In other words, by following this rule and rejecting the null hypothesis when the p-value is less than or equal to 0.05, we ensure that there is only a 5% chance (at most) that we are mistakenly rejecting the null hypothesis (rejecting H_0 even though H_0 is true).

For our coin-spinning experiment, we found a p-value of about 0.18. If we wished to achieve a significance level of 0.05, then we would *not* reject the hypothesis that spinning a coin is fair, because 0.18 is greater than 0.05.

KEY POINT To achieve a significance level of α, reject the null hypothesis if the p-value is less than (or equal to) α. If the p-value is greater than α, do not reject the null hypothesis.

The Four-Step Approach

With those details taken care of, we're now ready to perform a hypothesis test. We'll work completely through one example to show you how the steps fit together.

In one Florida election, 47% of all registered voters voted. A researcher studying the behavior of academics was curious about whether political scientists voted in the same proportion as the rest of the people in the state. To answer this question, the researcher took a random sample, without replacement, of 54 political scientists who live in Florida and interviewed them to determine whether they voted in this election. As it turned out, 40 of the 54 political scientists in the sample voted, which is a sample proportion of 0.74 (Schwitzgebel 2010).

We wish to carry out a hypothesis test to determine whether the proportion of all political scientists who voted in this election differed from the proportion of the general public.

Step 1: Hypothesize The null and alternative hypotheses are stated both in words and in symbols.

H_0: Political scientists vote in the same proportion as the public, 0.47.

H_a: Political scientists do not vote in the same proportion as the public.

$$H_0: p = 0.47$$
$$H_a: p \neq 0.47$$

The parameter p represents the proportion of all political scientists in Florida who voted in this election.

Step 2: Prepare First, we choose a significance level. We will use the standard value of $\alpha = 0.05$. This means that if it is true that political scientists voted at the same rate as the rest of the population of Florida, there is a 5% chance that we will mistakenly conclude that they did *not* vote at the same rate.

Next we choose a test statistic. Choosing the correct test statistic is not a big deal at this point, because you have seen only one test statistic to choose from: the one-proportion z-test statistic. In Section 8.4 you'll study a new test statistic for comparing two population proportions. Later still, you will see test statistics for comparing means.

Finally, we must make sure that conditions are met so that the distribution of the test statistic is approximately Normal. To do this, we must check the four conditions.

> **Details**
>
> **Other Test Statistics**
> In Chapter 9 you will learn about the t-test statistic, used for testing means, and in Chapter 10 you will learn about the chi-square test statistic.

Random Sample. We are told that the data come from a random sample of 54 political scientists, and this satisfies the first condition.

Large Sample. We must next check that the sample size of 54 is large enough to produce at least 10 successes and 10 failures.

If the null hypothesis is true, the probability of success is $p_0 = 0.47$. Because $n = 54$,

$np_0 = 54 \times 0.47 = 25.38$, which is more than 10, and

$n(1 - p_0) = 54 \times (1 - 0.47) = 28.62$, which is also more than 10.

Large Population. The third condition is true if the population—all political scientists registered to vote in Florida—is more than 10 times bigger than the sample size; that is, if the population size is greater than $10 \times 54 = 540$. We confess that we do not know the number of political scientists in Florida. We cannot check this condition, but we will assume that it is true. (Remember, we need only check this if the sampling was done without replacement.)

Independence. The final condition to check is that the observations are independent. We assume that, because the researcher used a random sample of political scientists, their responses were independent of one another.

Because the conditions are verified, if the null hypothesis (that the proportion of all political scientists who voted is 0.47) is true, then the sampling distribution of the one-proportion z-test statistic is $N(0, 1)$.

Step 3: Compute to compare At this point, we are ready to turn to technology to complete step 3. For instance, Figure 8.4 shows the input that StatCrunch requires to carry out the hypothesis test. First we must enter the number of observed successes and the sample size ("# of observations"). Then we must enter the hypotheses by choosing the correct value for the null hypothesis ($p_0 = 0.47$) and the correct form for the alternative hypothesis (two-sided). Other statistical software packages are very similar.

The software will now calculate the observed value of the test statistic and the p-value. Before we look at the output, though, let's go over the calculations ourselves.

Find the observed value of the test statistic. Remember that our test statistic will compare the value of the statistic provided by the data, $\hat{p}$, with the value that the null hypothesis says we should see, p_0.

The researcher reports that in his sample of 54 political scientists, 40 of them voted. The sample proportion therefore is $\hat{p} = \dfrac{40}{54} = 0.7407$ (after rounding.) How far away,

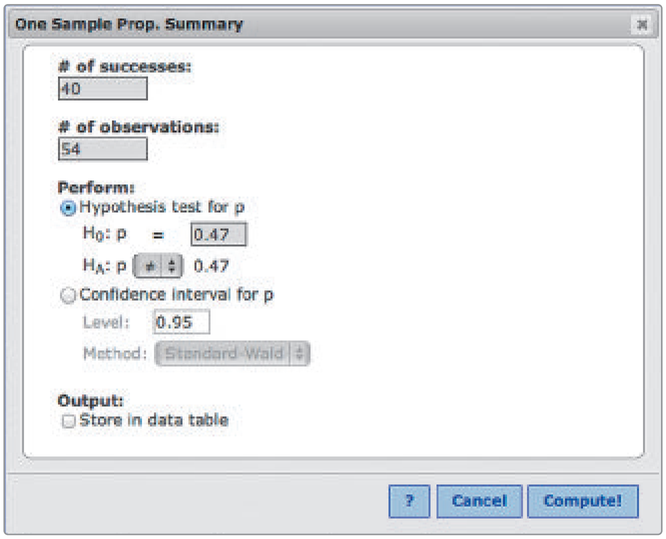


FIGURE 8.4 The procedure for carrying out a hypothesis test in StatCrunch is similar to that in many statistical software packages. You must enter the data and the correct hypotheses. (StatCrunch will also compute the number of successes directly from the raw data if provided.)

in terms of standard errors, is 0.7407 from 0.47? To answer, we must first find the standard error.

$$SE = \sqrt{\frac{p_0(1 - p_0)}{n}} = \sqrt{\frac{0.47 \times (1 - 0.47)}{54}} = 0.0679188$$

The observed value of our test statistic is

$$z_{\text{observed}} = \frac{\hat{p} - p_0}{SE} = \frac{0.740741 - 0.47}{0.0679188} = 3.99$$

We see that the observed proportion is just less than 4 standard errors above the null hypothesis claim.

Find the p-value. Because the alternative hypothesis is two-sided ($p \neq 0.47$), we will find a two-tailed p-value. Using Table 2 in Appendix A or a statistical calculator, we find that the p-value is about 0.00006. This calculation is illustrated in Figure 8.5.

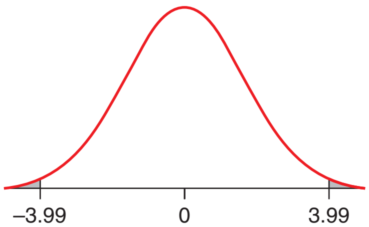
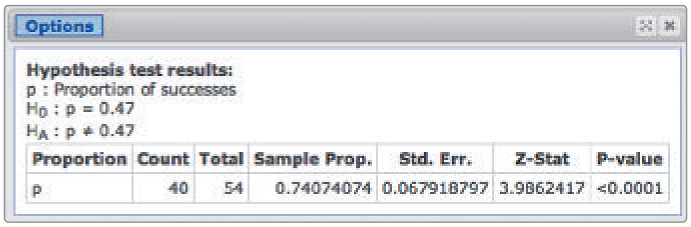


FIGURE 8.5 The p-value as a shaded area. This value is from a two-tailed test for which z is 3.99. The area has been enlarged a bit so that it can be seen readily.

Our calculations confirm the StatCrunch output, shown in Figure 8.6.



Options

Hypothesis test results:
p : Proportion of successes
H_0 : p = 0.47
H_A : p ≠ 0.47

Proportion	Count	Total	Sample Prop.	Std. Err.	Z-Stat	P-value
p	40	54	0.74074074	0.067918797	3.9862417	<0.0001

FIGURE 8.6 The StatCrunch output gives us the test statistic value of 3.986 and tells us that the p-value is smaller than 0.0001.

Note that StatCrunch follows a common convention for very small p-values. Rather than stating the precise value, it simply states that the value is very small. StatCrunch does this for p-values less than 0.0001. However, to be consistent with other technologies, in this text we'll do it whenever the p-value is less than 0.001.

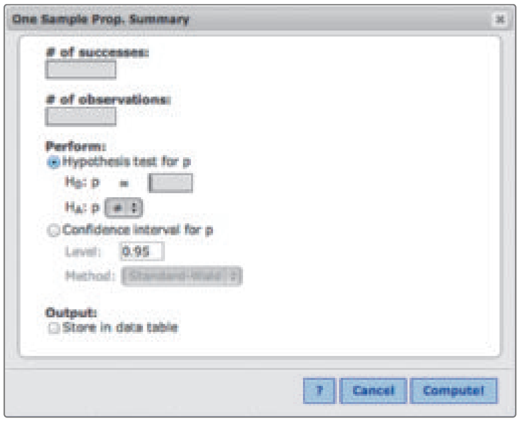
Step 4: Interpret Because the p-value is less than our stated significance of 0.05, we reject the null hypothesis. We conclude that political scientists did *not* vote in the same proportion as the rest of the population.

Often, technology is used to carry out a hypothesis test. Examples 9 and 10 illustrate how this is done using the statistical software package StatCrunch, but other packages are similar.

EXAMPLE 9 Men's Health

Health professionals are often concerned about our lifestyles and how they affect our well-being. A group of medical researchers knew from previous studies that in the past, about 39% of all men between the ages of 45 and 59 were regularly active. Because regular activity is good for our health, researchers were concerned that this percentage had declined over time. For this reason, they selected a random sample, without replacement, of 1927 men in this age group and interviewed them. Out of this sample, 680 said that they were regularly active (Elwood et al. 2013).

QUESTION Carry out the first two steps of a hypothesis test that will test whether the proportion of regularly active men in this age group has declined. Figure 8.7 shows the input required by a statistical software package. Explain how you would fill in the required entries shown in Figure 8.7. Use a significance level of 5%.



▶ **FIGURE 8.7** Most statistical software packages require input such as this in order to perform a hypothesis test based on summary statistics.

SOLUTION We will let p represent the proportion of all men in this age group who would say that they are regularly active.

Step 1: Hypothesize
In the past, the population proportion, p, was 0.39. The researchers wish to know whether this proportion has *decreased,* which means we have a left-sided alternative hypothesis. Thus

$$H_0: p = 0.39$$

$$H_a: p < 0.39$$

Step 2: Prepare
We will do a one-proportion z-test. The problem statement tells us to use a significance level of $\alpha = 0.05$.

We now check the conditions.

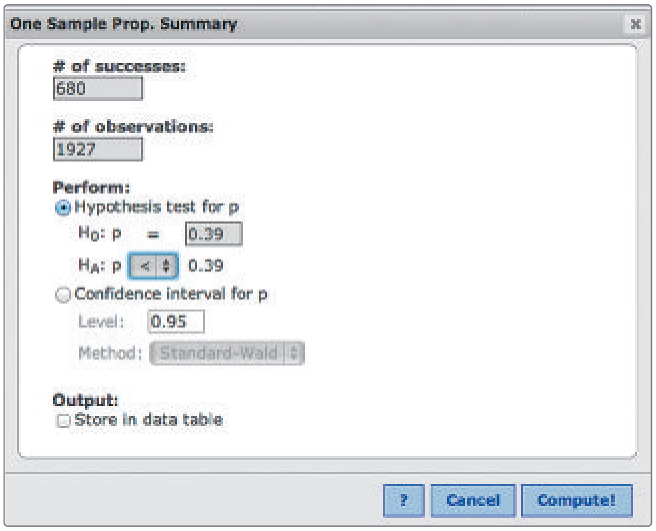
Random Sample. We are told the sample is random.

Large Sample. We have $p_0 = 0.39$. Thus we expect $1927 \times 0.39 = 752$ successes, which is bigger than 10. Because $(1 - 0.39)$ times 1927 will lead to an even larger number of expected failures, the number of expected failures is also bigger than 10. And so the sample size is large enough.

Large Population. If sampling was done without replacement, we need a large population. The population of all men in this age group is certainly larger than 10×1927.

Independence. As long as the sample was random and the men were interviewed independently, this condition is satisfied.

The required input to compute the test statistic and p-value using StatCrunch are shown in Figure 8.8. Similar inputs are required by other statistical software packages.

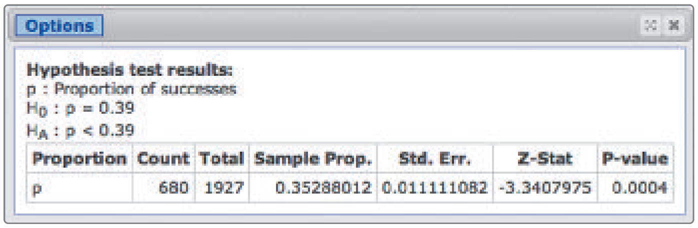


◀ **FIGURE 8.8** Required input, using StatCrunch, to test the hypothesis that the proportion of regularly active men in this age group has declined.

TRY THIS! Exercise 8.29

EXAMPLE 10 Men's Health, Continued

Using the output provided in Figure 8.9, carry out steps 3 and 4 of a hypothesis test to test whether the proportion of men aged 45–59 who say they are regularly active has declined from 0.39. Use a significance level of 5%.



Options

Hypothesis test results:
p : Proportion of successes
$H_0 : p = 0.39$
$H_A : p < 0.39$

Proportion	Count	Total	Sample Prop.	Std. Err.	Z-Stat	P-value
p	680	1927	0.35288012	0.011111082	-3.3407975	0.0004

◀ **FIGURE 8.9** StatCrunch output to test whether the proportion of regularly active men has declined from historical levels.

Details

Small p-Values
When small p-values, such as 0.0001, occur, many software packages round off and report the p-value as "$p < 0.001$" (or use some other small value). We will follow that practice in this book.

SOLUTION

Step 3: Compute to compare
The observed value of the test statistic is -3.34, which tells us that our observed sample proportion was 3.34 standard errors below the value of 0.39. The p-value is 0.0004, which is quite small.

Step 4: Interpret
Because the p-value is less than 0.05, we reject the null hypothesis. We conclude that the proportion of all men in this age group who are regularly active was smaller in 2009 than in 1979.

TRY THIS! Exercise 8.31

SNAPSHOT ONE-PROPORTION *z*-TEST

WHAT IS IT? ▶ A procedure for choosing between two hypotheses about the true value of a single population proportion. The test statistic is

$$z = \frac{\hat{p} - p_0}{SE}, \text{ where } SE = \sqrt{\frac{p_0(1 - p_0)}{n}}$$

WHAT DOES IT DO? ▶ Because estimates of population parameters are uncertain, a hypothesis test gives us a way of making a decision while knowing the probability that we will incorrectly reject the null hypothesis.

HOW DOES IT DO IT? ▶ The test statistic z compares the sample proportion to the hypothesized population proportion. Large values of the test statistic tend to discredit the null hypothesis.

HOW IS IT USED? ▶ When proposing hypotheses about a single population proportion. The data must be from an independent, random sample, and the sample size must be sufficiently large.

SECTION 8.3

Hypothesis Tests in Detail

In this section, we cover a variety of concepts that are important in correctly using and interpreting hypothesis tests.

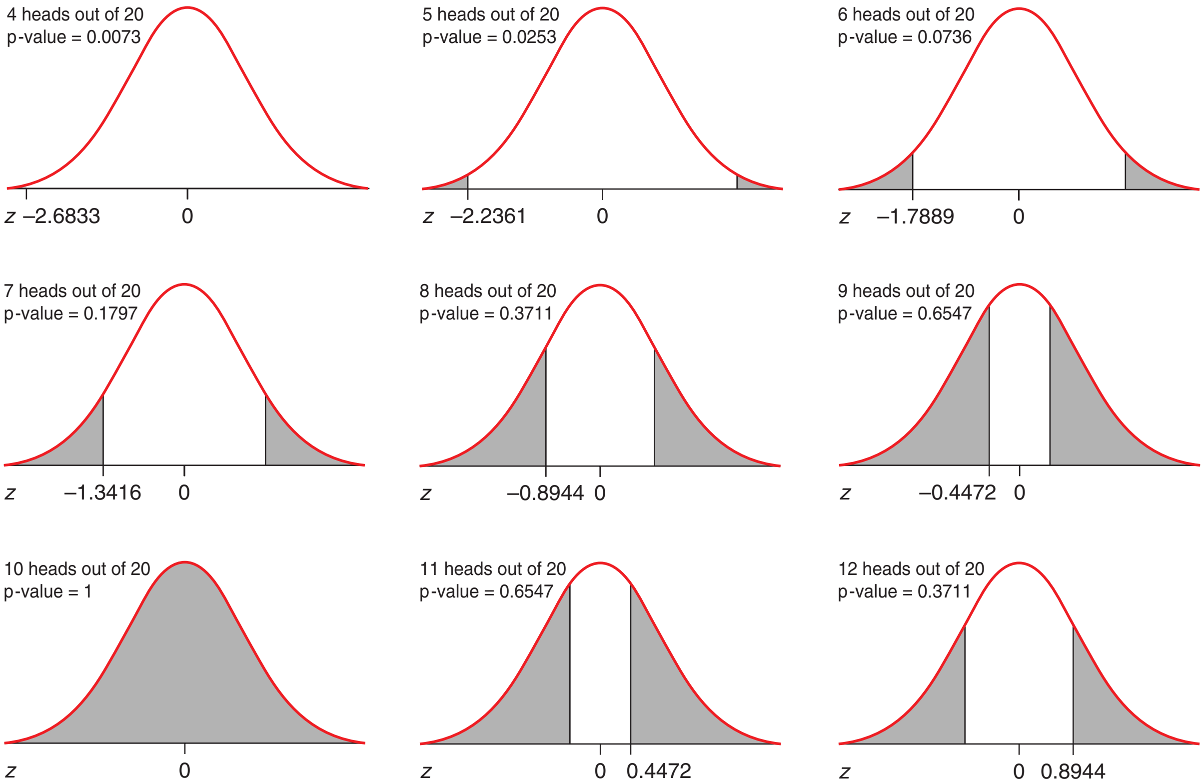
Xtreme Stats!

For many people, it seems a little odd that a small p-value leads to such a major action as rejecting the null hypothesis. But it is important to realize that when we see a small p-value, it means our test statistic is extreme. And an extreme test statistic means something unusual, and therefore unexpected, has happened.

Figure 8.10 illustrates how the p-value depends on the observed outcome of our coin-spinning study. Each graph represents the p-value for a different outcome, with the coin spun 20 times in each case. The null hypothesis in all cases is $p = 0.5$, and the alternative is the two-sided hypothesis that the probability of heads is not 0.5. Note that the closer the number of heads is to 10, the closer the z-value is to 0 and the larger the p-value is. Also note that the p-value for an outcome of 11 heads is the same as for 9 heads, and the p-value for an outcome of 12 heads is the same as for 8 heads. This happens because the alternative hypothesis is two-sided and the Normal distribution is symmetric.

! Caution

So Many *p*'s!
p is the population proportion.
p_0 is the value of the population proportion according to the null hypothesis.
$\hat{p}$ is the sample proportion.
The p-value is the probability that if the null hypothesis is true, our test statistic will be as extreme as or more extreme than the value we actually observed.



▲ **FIGURE 8.10** Each graph shows the p-value (shaded) for a different number of heads out of 20 spins of a coin, with the assumption that the coin is fair and using a two-sided alternative hypothesis. Note that the closer the number of heads is to 10 (out of 20), the closer the observed test statistic is to 0 and the larger the p-value is. Also, as the number of heads gets farther from 10, the observed test statistic gets farther from 0 and the p-value gets smaller.

EXAMPLE 11 p-Values for Coin Spinning

Two different students each did the coin-spinning experiment with a two-sided alternative hypothesis. Their test statistics follow.

$$\text{Study 1: } z = 1.98$$

$$\text{Study 2: } z = -2.02$$

QUESTION Which of these test statistics has the smaller p-value, and why?

SOLUTION If the null hypothesis is correct, then the test statistic should be close to 0. Values farther from 0 are more surprising and so have smaller p-values. Because -2.02 is farther from 0 than is 1.98, the area under the Normal curve in the tails is smaller for -2.02 than for 1.98. Thus -2.02 has the smaller p-value.

TRY THIS! Exercise 8.43

If Conditions Fail

If the conditions concerning the sampling distribution of the z-statistic fail to be met, then we cannot find the p-value using the Normal curve (using Table 2 in Appendix A or technology). However, other approaches often exist.

Sample Size Is Too Small The Normal distribution is only an approximation to the true distribution of the z-statistic. If the sample size is large enough, then the approximation is very good. If the sample size is too small, then the approximation may not be good, but other tests can be used. (These tests are not covered in this text.)

Samples Are Not Randomly Selected If samples are not selected randomly, then it is not possible to make inferences about the populations the samples came from. That having been said, random samples are relatively rare in medical studies. For example, medical researchers cannot take a random sample of people with the medical condition they wish to study; they must rely on recruiting patients who come into hospitals. Psychologists at universities often study students, particularly students who are willing to submit to a study for a small amount of money or the chance to win a prize in a raffle. They cannot take a random sample of all people in the population they wish to study and fly them to the university to participate in their experiment.

These convenience samples are problematic for the statistical techniques in this text—and indeed for any statistical technique. Sometimes we get around this by assuming that the samples are random or, at the very least, representative of the population. However, we have no guarantee that the conclusions we make with this assumption are valid or useful.

In many situations, though, researchers are not that interested in generalizing to the larger population. For example, in randomized controlled experiments, the emphasis is usually on understanding whether the treatment—maybe a new sleeping pill—works for anyone at all. A random sample of insomniacs is not available, but by randomizing the patients on hand to the treatment or a placebo, researchers can tell whether the pill is effective for *this* group. Research with other groups is still needed to see whether the results obtained are replicable, but the study is at least an encouraging start, because it can inform researchers that the therapy works with some patients. You'll see an example of this reasoning in Section 8.4.

! Caution

Cause and Effect
In Chapter 1 you learned that we can conclude that there is a cause-and-effect relationship between a treatment variable and a response variable only when we have a controlled experiment that uses random assignment, includes a placebo (or comparison) treatment, and is double-blind.

Balancing Two Types of Mistakes

One of the main ingredients in hypothesis testing is the probability of making a certain type of mistake. This type of mistake occurs when we reject the null hypothesis even though it is true. The probability of making this mistake is called the significance level. In step 2 of our hypothesis test, we deliberately set this probability to a small value, typically 5%.

If it is so important to have a small probability of making this mistake, why don't we choose an even smaller probability? Why not set it to 0?

The reason is that there is a tradeoff. As the probability of mistakenly rejecting the null hypothesis is made smaller, the probability of making another type of mistake gets bigger. This other mistake is to *fail* to reject the null hypothesis, even though it is wrong. For instance, we might conclude that there is no reason to think spinning a coin is biased, when, in fact, it *is* biased. Mistakes like this can be costly, because we might fail to make an important discovery. For instance, medical researchers might fail to recognize that a new medical procedure is effective, and as a result, many people will not have this potential cure available to them.

To understand this tradeoff, think about the criminal justice system. The null hypothesis, as the jury is told to believe, is that the defendant is innocent. The first type of mistake occurs when we convict an innocent person (mistakenly reject the null hypothesis). The probability of making this mistake is what we call the significance level. The second type of mistake occurs when we free a guilty person (fail to reject the null hypothesis even though it is false).

We can make the significance level (the probability of convicting an innocent man) 0 by following a simple rule: Free every defendant. If everyone goes free, then it is impossible to convict an innocent person because we will convict no one. But now the probability of freeing a guilty person is 100%, since every guilty person will be set free.

Of course, we could lower the probability of freeing guilty people to 0% by simply convicting everyone. But now the significance level has gone to 100% as well, because every innocent person will be convicted.

There is only one way out if we want to lower the probability of *both* types of mistakes:

Increase the sample size. Increasing the sample size improves the precision of the test and so we make mistakes less often.

KEY POINT We cannot make the significance level arbitrarily small, because doing so increases the probability that we will mistakenly fail to reject the null hypothesis.

Table 8.2 shows the two types of mistakes.

	Reject H_0	Fail to Reject H_0
H_0 True	Bad (The probability of doing this is called the significance level.)	Good
H_0 False	Good	Bad

◀ **TABLE 8.2** The two types of mistakes. If the null hypothesis is true, we might reject it. If the null is false, we might fail to reject it.

EXAMPLE 12 Describing Mistakes

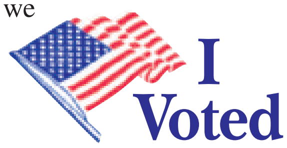
In Section 8.2, we considered whether political scientists vote in the same proportions as the general public. Our hypotheses were

$$H_0: p = 0.47$$
$$H_a: p \neq 0.47$$

where p is the proportion of all political scientists in Florida who voted.

QUESTION Describe the two types of errors we might make in conducting this hypothesis test. Your descriptions should be in the context of this problem. Explain what it means to set the significance level to 5%.

SOLUTION The first type of mistake is to reject the null hypothesis when it is true. In the present context, this means concluding that political scientists vote in different proportions than the public even though they actually have the same voter turnout as the general public. The second type of mistake is to fail to reject the null hypothesis when it is false. In the present context, this means concluding that there is no difference between political scientist voter turnout and the general turnout, even though there really is. The 5% significance level means that there is only a 5% chance that we will mistakenly conclude that the political scientists are different from the general public when, in fact, they are not.



TRY THIS! Exercise 8.45

So What? Statistical Significance vs. Practical Significance

Researchers call a result "statistically significant" when they reject the null hypothesis. This means that the difference between their data-estimated value for a parameter and the null hypothesis value for the parameter is so large that it cannot be convincingly explained by chance. However, just because a difference is statistically significant does not mean it is useful or meaningful.

A *practically* significant result is both statistically significant and meaningful. For example, suppose that the proportion of people who get a certain type of cancer is 1 in ten million. However, a statistical analysis finds that those who talk on their cell phones every day have a statistically significant greater risk of getting that cancer, and that the risk is doubled. It may be true that using your cell phone is therefore more dangerous than not using it, but would you really stop talking on the phone if your risk would change from 2 in 10 million to 1 in 10 million? That's a pretty big change of habit for a pretty small change in risk. Most people would conclude that the difference in risk is statistically significant, but not practically significant.

KEY POINT Statistically significant findings do not necessarily mean that the results are useful.

Don't Change Hypotheses!

A researcher sets up a study to see whether caffeine affects our ability to concentrate. He has a large number of subjects, and he gives them a task to complete when they have not had any caffeine. The task takes some concentration to complete, and he records how long it takes them. Later, he asks them to complete the same task, only this time the subjects have had a dose of caffeine. Again he records the time, and he's interested in the proportion who take longer to complete the task with caffeine than without.

He isn't sure just what the effect of caffeine will be. It might help people concentrate, in which case only a small proportion of people will take longer. On the other hand, it might make people jittery so that a large proportion will take longer to complete the task. If caffeine has no effect, probably half will take the same amount of time or more, and half will take the same amount of time or less.

The researcher chooses a significance level *of* $\alpha = 0.05$ to test this pair of hypotheses:

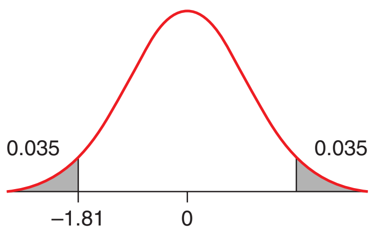
$$H_0\text{: } p = 0.50$$
$$H_a\text{: } p \neq 0.50$$

The parameter p represents the proportion of all people who would take longer to complete the task with caffeine than without. His alternative hypothesis is two-sided because he does not know what the effect will be—that is, whether caffeine will increase or decrease concentration.

He collects his data and gets a z-statistic of -1.81. This leads to a p-value of 0.07—and to a moral dilemma! (Figure 8.11 illustrates this p-value.) The researcher needs a p-value less than or equal to 0.05 if he is to publish this paper, because no one wants to hear about an insignificant result.

However, it occurs to this researcher that if he had a different alternative hypothesis, his p-value would be different. Specifically, if he had used

$$H_a\text{: } p < 0.50$$



▶ **FIGURE 8.11** The shaded areas represent the p-value of 0.070 for a test statistic of $z = -1.81$ in a two-sided hypothesis test.

then the p-value would have been the area in just the lower tail. In that case, his p-value would be 0.035, and he would reject the null hypothesis.

What the researcher has thought of doing is the sort of thing that small children (and politicians) do in a contest: They change the rules midway through so that they can win. As they say in the western movies, "You gotta dance with the gal/guy that brung ya." You can't choose your hypotheses to fit the data. By doing so, you are increasing the true significance level of your test above the 5% "advertised" significance level.

Hypothesis-Testing Logic

Statisticians and scientists are rather touchy when it comes to talk about "proving" things. They often use softer words, such as "Our data demonstrate that …" or "Our data are consistent with the theory that . . ." One reason is that in mathematical and scientific circles, the word *prove* has a very precise and very definite meaning.

If something is proved, then it is absolutely, positively, and without any doubt true. However, in real life, and particularly in statistics and science (which we consider to be part of real life), you can never be completely certain. In fact, as you've seen, mistakes are built into the hypothesis-testing procedure. We design the procedure so that mistakes are rare, but we know that they happen. For this reason, we avoid saying, for example, that we have *proved* that drinking caffeine changes the proportion of people who take longer to complete a particular task.

On a similar note, it is improper (maybe even impolite!) to say that you have "accepted" the null hypothesis when your p-value is bigger than 0.05. Instead, we say, "We have failed to reject H_0" or "We cannot reject H_0." This is because several factors might make it difficult to determine whether the null hypothesis is false.

It could be that our sample size was too small for us to detect that the null hypothesis was wrong. Basically, the amount of variability in a test statistic based on a small sample is so large that it washes out our ability to see the relatively small difference between the null hypothesis value and the true value. With a larger sample size, we'd have less sampling variability and maybe see that the true population proportion was different from the null hypothesis's claim.

In our long-running coin-spinning experiment, we spun the coin 20 times, got 7 heads, and found a p-value of 0.18. We failed to reject the null hypothesis that the coin was fair, and so we conclude that there is no evidence that spinning a coin is biased. Could we make a stronger statement? Could we claim that spinning a coin is fair?

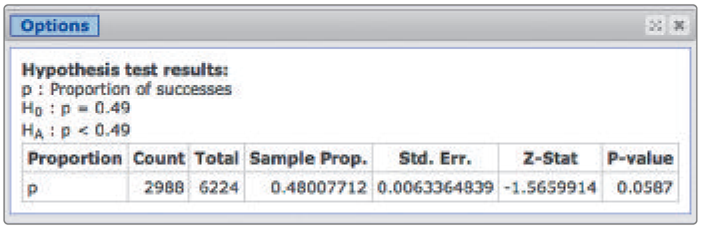
No. For instance, we might have had too small a sample size. That is, even though the sample size was large enough for the conditions of the Central Limit Theorem to give us a good approximation for the p-value, the sampling variability was still too great for us to see whether the probability of heads is different from 0.50 when spinning a coin.

KEY POINT Don't say you "proved" something with statistics. Say you "demonstrated" it or "showed" it. Similarly, don't say you "accept the null hypothesis"; say, rather, that you "cannot reject the null hypothesis" or that you "failed to reject the null hypothesis" or that "there is insufficient evidence to reject the null hypothesis."

EXAMPLE 13 Find the Flaws

Are public libraries in the United States an endangered species? In the past years, suppose that it was believed that roughly 49% of Americans had visited a library. (In truth, the percentage was higher, but we're using this value to illustrate a point.) Imagine that a professional association of librarians wishes to know whether attendance is declining. They examine a Pew survey conducted in 2013, in which only 48% of those in a random sample had visited a library in the last year (Pew Research Center 2013). The librarians decide to use a strict significance level of $\alpha = 0.01$. They carry out a hypothesis test, and the result is shown in Figure 8.12. Based on this, they send out a press release that says, "We have proved that there is no decline in library attendance."

▶ **FIGURE 8.12** StatCrunch output for the librarians' hypothesis test.



Options

Hypothesis test results:
p : Proportion of successes
$H_0 : p = 0.49$
$H_A : p < 0.49$

Proportion	Count	Total	Sample Prop.	Std. Err.	Z-Stat	P-value
p	2988	6224	0.48007712	0.0063364839	-1.5659914	0.0587

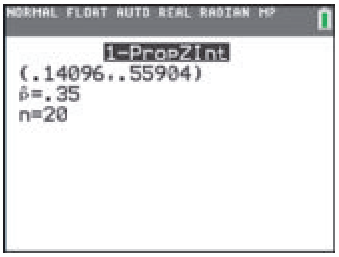
QUESTION What mistakes did this professional association make? How would you correct them?

SOLUTION The professional association concluded that they had *proved* that attendance did not decline. However, they have *not* proved that the null hypothesis is true. At best, they have not found sufficient evidence to reject it, but this is very different from saying that they have found evidence that proves it is true.

The association should conclude that, using a 1% significance level, there is not enough evidence to conclude that library membership has declined.

TRY THIS! Exercise 8.49

Confidence Intervals and Hypothesis Tests



▲ **FIGURE 8.13** TI-84 confidence interval output for 7 heads out of 20 spins.

Confidence intervals and hypothesis tests are closely related, even though they are used to answer (slightly) different questions. Confidence intervals are used to answer the question "What is the value of this parameter?" For instance, suppose we spin a coin 20 times and get 7 heads. An approximate 95% confidence interval for the probability of getting heads is 0.14 to 0.56, as shown in Figure 8.13. This tells us that, on the basis of our data, we are highly confident that the true probability of getting heads is between 14% and 56%.

The hypothesis test answers a slightly different question: "Are the data consistent with the parameter being one particular value, or might the parameter be something else?" These hypothesis tests are a little more vague: We are not really asking what the value is; we simply want to know whether it is one thing or another. For instance, for the coin-spinning we ask, "Are the data consistent with the coin being fair? That is, $p = 0.50$? Or is the coin not fair?"

Even though they are designed to answer different questions, they are similar enough that you can often use a confidence interval to reach the same types of conclusions you would reach with a hypothesis test using a two-sided alternative hypothesis. In most situations, doing a hypothesis test with a two-sided alternative hypothesis and significance level α (in percentage points) will lead to the same conclusion as finding a confidence interval with a $(1 - \alpha)$ confidence level and rejecting the null hypothesis if its value is not captured by the interval. Table 8.3 shows this relationship between confidence level and significance level.

Looking Back

Confidence Intervals for Proportions
You learned how to find a confidence interval for a population proportion in Section 7.4.

For instance, if we wish to test whether the spun coin is fair with a significance level of $\alpha = 0.05$, and our sample proportion after 20 spins is 0.35, then we find the $(1 - \alpha) = 1 - 0.05 = 0.95$, or 95%, confidence interval based on this outcome. The interval is 0.14 to 0.56, as shown in Figure 8.13. Because this interval includes 0.5, we would not reject the null hypothesis that the coin is fair.

Confidence Level = $(1 - \alpha)$	Alternative Hypothesis	Significance Level α
90%	Two-sided ($\neq$)	10%
95%	Two-sided ($\neq$)	5%
99%	Two-sided ($\neq$)	1%

◀ **TABLE 8.3** The relationship between confidence levels and significance levels for hypothesis tests. There are rare occasions when the conclusions of a hypothesis test would be different from what we might conclude from a confidence interval, because of the different method of calculating the standard error.

If we wished to do this test with a 10% significance level, we would find the 90% confidence interval. This turns out to be 0.17 to 0.53, as shown in the StatCrunch output in Figure 8.14. Again, the interval includes 0.50, the null hypothesis value, so we do not reject the null hypothesis.

Options

90% confidence interval results:
p : Proportion of successes
Method: Standard-Wald

Proportion	Count	Total	Sample Prop.	Std. Err.	L. Limit	U. Limit
p	7	20	0.35	0.10665365	0.17457037	0.52542963

▲ **FIGURE 8.14** StatCrunch output for a 90% confidence interval for the proportion of heads for spinning a coin.

KEY POINT A confidence interval answers the question "What is the value of the parameter?" A hypothesis test is used to judge between two different claims about the value.

Details

Rare Differences

It is possible to reach different conclusions using confidence intervals and hypothesis tests, because the standard errors are computed using slightly different values for the population proportion p. The confidence interval estimates p with the sample proportion ($\hat{p}$), while the hypothesis test uses the value claimed by the null hypothesis, p_0.

SECTION 8.4

Comparing Proportions from Two Populations

You have now seen how to carry out a hypothesis test for a single population proportion. With very few changes, this procedure can be altered to accommodate a more interesting situation: comparing proportions from two populations.

Consider as an example the opinion polls on embryonic stem cell research introduced in Section 7.5. This medical research shows great promise in the treatment of several major diseases but is controversial because it goes against many people's religious convictions. The Pew Forum on Religion & Public Life has, over time, conducted surveys to assess Americans' support for stem cell research. In 2002, 43% of Americans expressed support for stem cell research (Pew Forum 2008). Later, in 2009, 58% supported this research (Pew Forum 2009). Can we conclude that support has changed in the population of all Americans? Or could this difference be due to chance variation during the sampling procedure?

This problem involves two populations. One population consists of all Americans in 2009, and the second population consists of all Americans in 2002. Each population has a true proportion who support stem cell research, but in each case we cannot know this true value. Instead, we have a random sample taken from both populations, and we must estimate the two proportions from these two random samples.

Here are the changes we need to make to our "ingredients" in order to compare proportions from two populations.

Changes to Ingredients: The Hypotheses

Because we now have two population proportions to consider, we need some new notation. We'll let p_1 represent the proportion of Americans who supported stem cell research in 2009, and we'll let p_2 represent the proportion who supported it in 2002.

We are not interested in the actual numerical values of p_1 and p_2, as we were when dealing with just one population proportion. We are interested only in their relation to each other. The conservative, status-quo, not-worth-a-headline hypothesis is that these proportions are the same. In other words, there has been no change in support. We write this as

$$H_0: p_1 = p_2$$

In words, the null hypothesis says that the proportion of Americans who support stem cell research was the same in 2009 as it was in 2002.

The alternative hypothesis is that the proportion of Americans who support stem cell research has changed. If so, the two proportions are no longer equal.

$$H_a: p_1 \neq p_2$$

One-sided hypotheses are also possible. Our research question might instead have been "Has support for stem cell research decreased?" If that had been our question, then we would have used

$$H_a: p_1 < p_2$$

And if we had wished to answer the question "Has support for stem cell research increased?" we would have used this alternative:

$$H_a: p_1 > p_2$$

These options lead to three pairs of hypotheses, as shown in Table 8.4. You choose the pair that corresponds to the research question your study hopes to answer. Note that the null hypothesis is always $p_1 = p_2$ because the neutral position is always that the two proportions are the same.

▶ **TABLE 8.4** Possible hypotheses for a two-proportion hypothesis test.

Hypothesis	Symbols	The Alternative in Words
Two-sided	$H_0: p_1 = p_2$ $H_a: p_1 \neq p_2$	The proportions are different in the two populations.
One-sided (left)	$H_0: p_1 = p_2$ $H_a: p_1 < p_2$	The proportion in population 1 is less than the proportion in population 2.
One-sided (right)	$H_0: p_1 = p_2$ $H_a: p_1 > p_2$	The proportion in population 1 is greater than the proportion in population 2.

Changes to Ingredients: The Test Statistic

We are interested in how p_1 and p_2 differ, so our test statistic is based on the difference between our sample proportions from the two populations. The test statistic we will use has the same structure as the one-sample z-statistic:

$$z = \frac{\text{estimator} - \text{null value}}{SE}$$

However, the estimator for the **two-proportion z-test** is $\hat{p}_1 - \hat{p}_2$ because we are estimating the difference $p_1 - p_2$. Here $\hat{p}_1$ and $\hat{p}_2$ are just the sample proportions for the different samples. In our case, $\hat{p}_1$ is the sample proportion for the people surveyed in 2009 (reported as 0.58), and $\hat{p}_2$ is the sample proportion for the people surveyed in 2002 (reported as 0.43).

The null value is 0, because the null hypothesis claims these proportions are the same, so $p_1 - p_2 = 0$.

The standard error, SE, is more complicated than in the one-sample case, because the null hypothesis no longer tells us the value of the population proportion. All it tells us is that both populations have the same value. For this reason, when we estimate this single value, we pool the two samples together. Formula 8.2 shows you how to do this.

Details

Finding Differences
When we talk about "the difference" between two quantities, we mean "How far apart are they?" We answer this by subtracting one from the other.

Details

No Difference
If two quantities are the same, their difference is 0.

Formula 8.2: The two-proportion *z*-test statistic

$$z = \frac{\hat{p}_1 - \hat{p}_2 - 0}{SE}$$

where

$$SE = \sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

n_1 = sample size in sample 1

n_2 = sample size in sample 2

$$\hat{p} = \frac{\text{number of successes in sample 1} + \text{number of successes in sample 2}}{n_1 + n_2}$$

$$\hat{p}_1 = \text{proportion of successes in sample 1} = \frac{\text{number of successes in sample 1}}{n_1}$$

$$\hat{p}_2 = \text{proportion of successes in sample 2} = \frac{\text{number of successes in sample 2}}{n_2}$$

Formula 8.2 is perhaps the most elaborate formula we have shown you so far. As usual, it is much more important to be able to use technology to perform this test than to apply the formula. Still, studying the formula does help us understand why the test statistic is useful.

Caution

Minding Your *p*'s
It usually doesn't matter which population you call "1" and which you call "2," but once you've made the choice, you must stick with it.

EXAMPLE 14 Pew Survey on Stem Cell Research

The researchers from the Pew study interviewed two random samples. Both samples, the one taken in 2002 and the one taken in 2009, had 1500 people. In 2002, 645 people expressed support for stem cell research. In 2009, 870 expressed support. These data are summarized in Table 8.5.

	2002	2009	Total
Support Stem Cell Research	645	870	1515
Do Not Support Stem Cell Research	855	630	1485
Total	1500	1500	3000

◀ **TABLE 8.5** Data for the Pew study.

Looking Back

Two-way Tables
Two-way tables, such as the one used to summarize the data in Example 14, were first presented in Chapter 1.

QUESTION Find the observed value of the test statistic to test the hypotheses

$$H_0: p_1 = p_2$$
$$H_a: p_1 \neq p_2$$

where p_1 represents the proportion of Americans who supported stem cell research in 2009, and p_2 represents the proportion who supported this research in 2002.

SOLUTION We must bring all the pieces together and assemble them into the test statistic:

$$\hat{p}_1 = \frac{870}{1500} = 0.58 \text{ (a value we knew already from the original report)}$$

$$\hat{p}_2 = \frac{645}{1500} = 0.43 \text{ (another value we knew from the original report)}$$

$$\hat{p} = \frac{870 + 645}{1500 + 1500} = 0.505 \text{ (a pooled estimate of the sample proportion)}$$

Details

The Sign of *z*
In a two-proportion test, whether *z* is positive or negative depends on which population you call "1" and which you call "2." It's important to pay attention to which proportion is subtracted from which!

$$SE = \sqrt{0.505(1 - 0.505)\left(\frac{1}{1500} + \frac{1}{1500}\right)} = 0.018257$$

Now we assemble the pieces:

$$z_{\text{observed}} = \frac{0.58 - 0.43}{0.018257} = 8.22$$

TRY THIS! Exercise 8.65

Changes to Ingredients: Checking Conditions

The conditions that we need to check for a two-sample test of proportions are similar to those for a one-sample test, but with some additional things to consider.

1. *Large Samples*: Both sample sizes must be large enough. Because we don't know the value of p_1 or p_2, we must use an estimate. The null hypothesis says that these two proportions are the same, so we use $\hat{p}$, the pooled sample proportion, to check this condition. Do not use $\hat{p}_1$ or $\hat{p}_2$. This means that we need
   a. $n_1\hat{p} \geq 10$ and $n_1(1 - \hat{p}) \geq 10$
   b. $n_2\hat{p} \geq 10$ and $n_2(1 - \hat{p}) \geq 10$
2. *Random Samples*: The samples are drawn randomly from the appropriate population. In practice, this condition is often impossible to check unless we were present when the data were collected. If we were not told explicitly the sample was randomly drawn, we may have to assume the condition is satisfied.
3. *Independent Samples*: The samples are independent of each other. This condition is violated if, for example, the same individuals are in both samples that we are comparing.
4. *Independent within Samples*: The observations within each sample must be independent of one another.

If these four conditions hold, then, if the null hypothesis is true, z follows a $N(0, 1)$ distribution.

EXAMPLE 15 Right of Way

Psychologists at the University of California, Berkeley, were interested in studying whether people driving "high-status" cars behaved differently than those driving other cars. State law requires that cars come to a complete halt and not enter a crosswalk that contains a pedestrian. The researchers used an accomplice to step into a crosswalk on a busy street in California. The researchers took careful note of whether the driver illegally cut off the pedestrian accomplice by entering the crosswalk. They also rated the status of the car with a number between 1 and 5; a 1 meant "low" status, and a 5 meant "high" (such as a Rolls Royce). A vehicle was recorded only if there were no cars in front of or behind it. This helped ensure that observations were independent. (For instance, if one car stops, a car behind it might be more likely to stop.) Although the actual study used an advanced statistical model to control for potential confounding factors, we can perform a simple analysis by combining the cars into two groups. The researchers observed 33 "low-status" cars (rated 1 or 2), and 119 "high-status" cars (rated 3 or higher). Of the low-status cars, 8 cars failed to stop. Of the high-status cars, 45 failed to stop (Piff et al. 2012).

QUESTION Find and state the proportion of cars that failed to stop for both groups. Perform a four-step hypothesis test to test the hypothesis that high-status cars are more likely to cut off a pedestrian. Use a significance level of 0.10.

SOLUTION The proportion of low-status cars that cut off the pedestrian was $8/33 = 0.242$. The proportion of high-status cars that cut off the pedestrian was $45/119 = 0.378$.

As always, our hypotheses are about populations. We witnessed only 33 low-status and 119 high-status cars, but we take these as representative of a population and hypothesize about *all* low-status and high-status cars that might pass through this crosswalk.

Let p_1 represent the proportion of all low-status cars that would fail to stop if they were passing through this crosswalk when a pedestrian stepped into the crosswalk. Let p_2 represent the proportion of all high-status cars that would fail to stop.

Step 1: Hypothesize
The null hypothesis is neutral; it states that both groups of cars are the same.

$$H_0: p_1 = p_2$$

The alternative hypothesis is that the high-status cars cut off pedestrians more often:

$$H_a: p_1 < p_2$$

Step 2: Prepare
We use a significance level of $\alpha = 0.10$. We are comparing two population proportions, so our test statistic will be the two-proportion z-statistic. We must check to see whether conditions are satisfied so that we can use the Normal distribution as an approximate sampling distribution.

1. Are both samples large enough?
   We find that

$$\hat{p} = \frac{8 + 45}{33 + 119} = 0.34868$$

   First sample: $n_1 \times 0.34868 = 33 \times 0.34868 = 11.5$, which is greater than 10, and $33 \times (1 - 0.34868) = 21.5$, which is also greater than 10.

   Second sample: $n_2 \times 0.34868 = 119 \times 0.34868 = 41.5$ and $119 \times (1 - 0.34868) = 77.5$, both of which are greater than 10.

2. Are the samples drawn randomly from their respective populations?
   Frankly, probably not (but we hope they are representative). However, to proceed, we will assume that they are.

3. Are the samples independent of each other?
   Yes; there is no reason to suppose that the actions of low-status cars will affect high-status cars, or the other way around.

4. Are observations within each sample independent?
   Yes; the researchers took care to make sure this was so by recording only cars that were not following other cars.

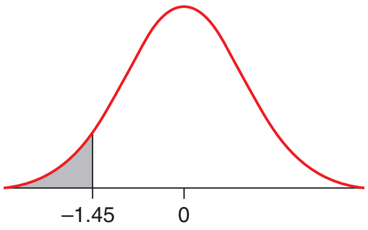
With these three conditions checked, we can proceed to step 3.

Step 3: Compute to compare
We must find the individual pieces of Formula 8.2. We defined p_1 to be the proportion of low-status cars that cut off pedestrians. So let $\hat{p}_1$ be the proportion of low-status cars *in the sample* that cut off pedestrians. The sample size in this group was $n_1 = 33$. Earlier, we calculated

$$\hat{p}_1 = \frac{8}{33} = 0.2424$$

The sample proportion for the high-status cars was

$$\hat{p}_2 = \frac{45}{119} = 0.3782$$



▲ **FIGURE 8.15** The shaded area represents the probability to the left of a test statistic value of −1.45.

To find the standard error, we need to use $\hat{p}$, the proportion of bad events that happen in the sample if we ignore the fact that the cars belong to two different groups. Above we found that $\hat{p} = 0.34868$. The standard error is then

$$SE = \sqrt{0.34868(1 - 0.34868)\left(\frac{1}{33} + \frac{1}{119}\right)} = 0.0938$$

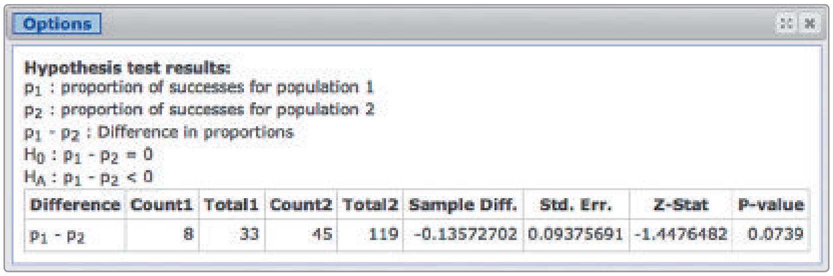
Putting it all together yields

$$z_{\text{observed}} = \frac{0.2424 - 0.3782}{0.0938} = -1.45$$

Now that we know the observed value, we must measure our surprise. The null hypothesis assumes that the two population proportions are the same and that our sample proportions differed only by chance. The p-value will measure the probability of getting an outcome as extreme as or more extreme than −1.45, assuming that the population proportions are the same.

The p-value is calculated exactly the same way as with a one-proportion z-test. Our alternative hypothesis is a left-sided hypothesis, so we need to find the probability of getting a value less than the observed value (Figure 8.15).

▶ **FIGURE 8.16** StatCrunch output shows the observed value of the two-proportion z-test statistic, −1.45, and the p-value of 0.0739.


Options

Hypothesis test results:
p_1 : proportion of successes for population 1
p_2 : proportion of successes for population 2
$p_1 - p_2$: Difference in proportions
$H_0 : p_1 - p_2 = 0$
$H_A : p_1 - p_2 < 0$

Difference	Count1	Total1	Count2	Total2	Sample Diff.	Std. Err.	Z-Stat	P-value
$p_1 - p_2$	8	33	45	119	-0.13572702	0.09375691	-1.4476482	0.0739

Figure 8.16 provides the z-value and the p-value from StatCrunch:

$$\text{p-value} = 0.074$$

Step 4: Interpret

The p-value is less than our stated significance level of 0.10, so we reject the null hypothesis. We conclude that the high-status cars were more likely to cut off the pedestrian.

Note that because this was an observational study, we can't conclude cause and effect, which means that we can't conclude that driving a high-status car makes a person less concerned about pedestrians. Also, our sample was not randomly selected from a larger population, so perhaps the results would differ in different parts of the country. The p-value was relatively high; if we had used a significance level of 0.05, we would *not* have rejected the null hypothesis. The original researchers used more sophisticated methods that enabled them to control for potential confounders and allowed them to see how the proportion of cutting off a pedestrian varied for each of the five status levels. With these more refined methods, they got a smaller p-value.

TRY THIS! Exercise 8.69

SNAPSHOT TWO-PROPORTION z-TEST

WHAT IS IT? ► A hypothesis test.

WHAT DOES IT DO? ► Provides a procedure for comparing two population proportions. The null hypothesis is always that the proportions are the same, and this procedure gives us a way to reject or fail to reject that hypothesis.

HOW DOES IT DO IT? ► The test statistic z compares the differences between the sample proportions and the value 0 (which is what the null hypothesis says this difference should be):

$$z = \frac{\hat{p}_1 - \hat{p}_2 - 0}{SE}, \text{ where } SE = \sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Values of z that are far from 0 tend to discredit the null hypothesis.

HOW IS IT USED? ► When comparing two proportions, each from a different population. The data must come from two independent, random samples, and each sample must be sufficiently large. Then $N(0, 1)$ can be used to compute the p-value for the observed test statistic.

CASE STUDY REVISITED

During political debates, candidates sometimes dodge questions by answering a question different from the question asked. Can posting the correct question at the bottom of the TV screen make it more likely that viewers will notice this evasion? Two researchers randomly assigned viewers to one of two conditions. In both conditions, viewers watched actors in a debate in which one of the actors dodged the question. Under one condition, though, the question was posted on the screen while the viewers watched. In this group, 88% of viewers noticed the dodge. The other group saw nothing posted, and only 39% of them noticed the dodge. Can this difference of 88% − 39% = 49 percentage points be due to chance? Or does it suggest that we can become better dodge detectors if we are reminded of the question?

To find out, the researchers carried out a hypothesis test. The data are summarized in Table 8.6 and Figure 8.17.

	Question Posted	No Question Posted	Total
Detected	63	28	91
Not Detected	9	43	52
Total	72	71	143

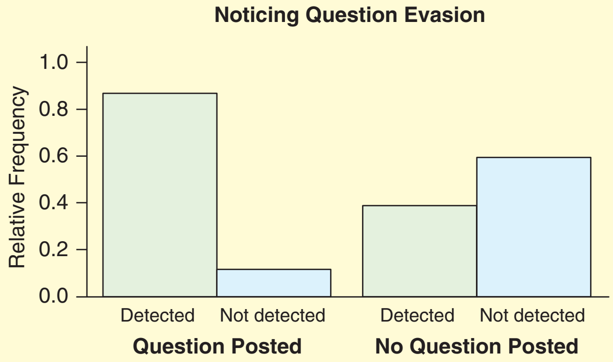
▲ **TABLE 8.6** Two-way table for the relationship between posting of a question and dodge detection.

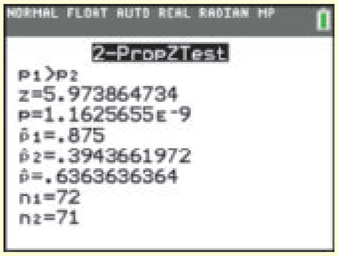
We carried out a two-proportion z-test to see whether posting questions improves dodge detection. We'll used a significance level of 0.05. Say we let p_1 represent the proportion of people who will notice the dodge when the question is posted during the debate, and we let p_2 represent the proportion of people who will notice the dodge when no question is posted. Then our hypotheses are

H_0: $p_1 = p_2$ (The same percentage in both groups will recognize the dodge.)

H_a: $p_1 > p_2$ (A greater percentage will recognize the dodge when the question is posted.)

▶ **FIGURE 8.17** Relationship between detecting a dodge and whether or not the viewer saw the question on their TV screen.





▲ **FIGURE 8.18** TI-84 output for a two-sample *z*-test using a two-sided alternative hypothesis. The p-value is 0.00000000116.

A quick check shows that the sample sizes are large enough and that the other necessary conditions hold. The *z*-statistic, calculated with technology (see Figure 8.18), is 5.97. We know from the Empirical Rule that the p-value will be very small, since *z*-statistics are almost never that far from 0. And indeed, if we were to calculate the p-value, we would find that if the null hypothesis were true, then the probability of getting a test statistic as large as 5.97 or larger would be 0.000000001.

With such a small p-value, we reject the null hypothesis and conclude that posting the question on the screen really did help viewers notice the "candidate's" evasion of the question.

EXPLORING STATISTICS
CLASS ACTIVITY

Identifying Flavors of Gum through Smell

GOALS	MATERIALS
To use a hypothesis test to determine how well a person can distinguish between flavors based on smell alone.	• Gum (or candy) in two different flavors. We will call these flavor A and flavor B. Each student will need one piece of each flavor. • A paper towel for each student.

ACTIVITY

Pair up. One student will take the role of the sniffer, and the other will act as the researcher. Both students must know the two flavors that are being tested.

Researcher: Ask the sniffer to turn his or her back, then select a piece of gum at random and record the flavor of the gum on a sheet of paper. Place the gum on a sheet of paper towel.

Hold the gum about 2 inches below the sniffer's nose and ask him or her to identify which of the two flavors is in the hand. Sniffers may take as long as they like. Record whether the sniffer's response was correct next to the flavor that you wrote down for this trial.

Do 20 trials, then work together to determine the proportion of correct responses. If time permits, change roles.

BEFORE THE ACTIVITY

1. Let p represent the probability that the sniffer makes a correct identification. If the sniffer is simply guessing and cannot tell the difference, what value would p have?
2. If the sniffer is not guessing and really can tell the difference, would p be bigger than, less than, or the same as the value you found in Question 1?
3. Suppose the sniffer cannot tell the difference between the two flavors. In 20 trials, about how many would you expect the sniffer to get right? What proportion is this? What is the greatest number of trials you'd expect the sniffer to get right if she or he were just guessing? What proportion is this?
4. How many trials would the sniffer have to get right before you would believe that he or she can tell the difference? All 20? 19? Explain.
5. Write a pair of hypotheses to test whether the sniffer is just guessing or can really tell the difference. Write the hypotheses in both words and symbols, using the parameter p to represent the probability that the sniffer correctly identifies the scent.

AFTER THE ACTIVITY

1. Report the proportion of trials the sniffer got right.
2. Does this show that the sniffer could tell the difference in the scents? Explain.
3. If the sniffer is just guessing, what is the probability that he or she would have gotten as many right as, or more right than, the actual number recorded?

   a. Pretty likely: between 50% and 100% b. Maybe: between 10% and 50%

   c. Fairly unlikely: between 5% and 10% d. Very unlikely: between 0% and 5%

CHAPTER REVIEW

KEY TERMS

hypothesis testing, *380*
null hypothesis, H_0, *381*
alternative hypothesis, H_a, *381*
two-sided hypothesis, *382*
one-sided hypothesis, *382*
significance level, α (alpha), *384*
test statistic, *385*
one-proportion z-test, *385*
p-value, *386*
two-proportion z-test, *404*

LEARNING OBJECTIVES

After reading this chapter and doing the assigned homework problems, you should

- Know how to test hypotheses concerning a population proportion and hypotheses concerning the comparison of two population proportions.
- Understand the meaning of p-value and how it is used.
- Understand the meaning of significance level and how it is used.
- Know the conditions required for calculating a p-value and significance level.

SUMMARY

Hypothesis tests are performed in the following four steps.

Step 1: Hypothesize.
Step 2: Prepare.
Step 3: Compute to compare.
Step 4: Interpret.

Step 1 is the most important, because it establishes the entire procedure. Hypotheses are *always* statements about parameters. The alternative hypothesis is the hypothesis that the researcher wishes to convince the public is true. The null hypothesis is the skeptical, neutral hypothesis. Each step of the hypothesis test is carried out assuming that the null hypothesis is true. For all tests in this book, the null hypothesis will always contain an equals ($=$) sign, whereas the alternative hypothesis can contain the symbol for "is greater than" ($>$), the symbol for "is less than" ($<$), or the symbol for "is not equal to" ($\neq$).

Step 2 requires that you decide what type of test you are doing. In this chapter, this means you are either testing the value of a proportion from a single population (one-proportion z-test) or comparing two proportions from different populations (two-proportion z-test). You must also check that the conditions necessary for using the standard Normal distribution as the sampling distribution are all met.

Step 3 is where the observed value of the statistics are compared to the null hypothesis. This step is most often handled by technology, which will compute a value of the test statistic and the p-value. These values are valid only if the conditions in step 2 are satisfied.

Step 4 requires you to compare the p-value, which measures our surprise at the outcome if the null hypothesis is true, to the significance level, which is the probability that we will mistakenly reject the null hypothesis. If the p-value is less than (or equal to) the significance level, then you must reject the null hypothesis.

For a one-proportion z-test:

$$\textbf{Formula 8.1: } z = \frac{\hat{p} - p_0}{SE}$$

where

$$SE = \sqrt{\frac{p_0(1 - p_0)}{n}}$$

p_0 is the proposed population proportion
$\hat{p}$ (p-hat) is the sample proportion, x/n
n is the sample size

For a two-proportion z-test:

$$\textbf{Formula 8.2: } Z = \frac{\hat{p}_1 - \hat{p}_2 - 0}{SE}$$

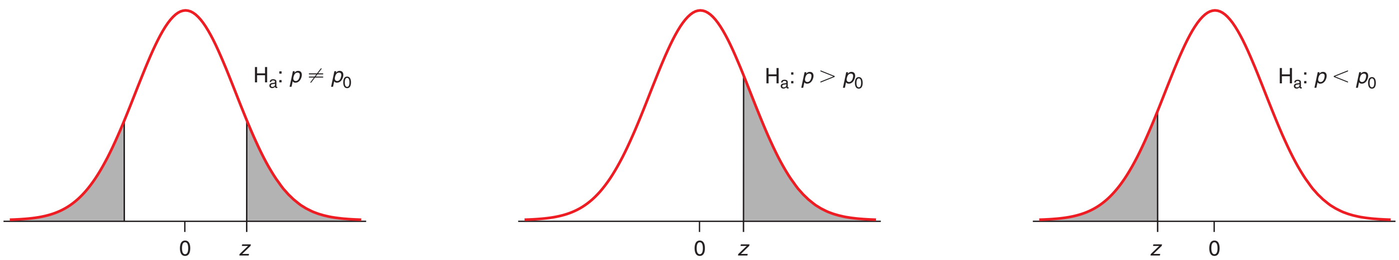
where

$$SE = \sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$\hat{p} = \frac{\text{number of successes in both samples}}{n_1 + n_2}$$

$\hat{p}_1$ is the proportion of successes in the first sample, and $\hat{p}_2$ is the proportion in the second sample

Calculating the p-value depends on which alternative hypothesis you are using. Figure 8.19 shows, from left to right, a two-tailed p-value, one-tailed (right-tailed) p-value, and a one-tailed (left-tailed) p-value.



▲ **FIGURE 8.19** Representations of possible p-values for three different alternative hypotheses. The area of the shaded regions represents the p-value.

SOURCES

Einstein-PE Investigators. 2012. Oral rivaroxaban for the treatment of symtomatic pulmonary embolism. *New England Journal of Medicine* 336, 12.87–12.97, April 5.

Elwood, P., J. Galante, J. Pickering, S. Palmer, A. Bayer, Y. Ben-Schlomo, M. Longley, and J. Gallacher. 2013. Healthy lifestyles reduce the incidence of chronic diseases and dementia: Evidence from the Caerphilly Cohort Study. *PLoS ONE* 8(10), e81877. doi: 10.137 /journal.pone.0081877

Feder, L., and L. Dugan. 2002. A test of the efficacy of court-mandated counseling for domestic violence offenders: The Broward Experiment. *Justice Quarterly* 19, 343–375.

R. Furman et al. 2014. Idelalisib and rituximab in relapsed chronic lymphocytic leukemia. *New England Journal of Medicine*, January 22, 2014, doi: 10.1056/NEJMoa1315226

Leschied, A., and A. Cunningham. 2002. *Seeking effective interventions for serious young offenders: Interim results of a four-year randomized study of multisystemic therapy in Ontario, Canada.* London, Canada: London Family Court Clinic.

Pew Forum. 2008. Declining Majority of Americans Favor Embryonic Stem Cell Research. http://www.pewforum.org/2008/07/17/declining -majority-of-americans-favor-embryonic-stem-cell-research/(accessed May 13, 2014).

Pew Forum. 2009. Support for Health Care Overhaul, but its not 1993: Stable Views of Stem Cell Research. The Pew Forum on Religion & Public Life Issues. http://www.people-press.org/2009/03/19/support -for-health-care-overhaul-but-its-not-1993/ (accessed May 13, 2014).

Pew Research Center. 2013. How Americans value public libraries in their communities. http://libraries.pewinternet.org/2013/12/11/libraries-in -communities/

Piff, P., D. Stancato, S. Cote, R. Mendoza-Denton, and D. Keltner, D. 2012. Higher social class predicts increased unethical behavior. *Proceedings of the National Academy of Sciences* 109(11), 4086–4091.

Rogers, T. and Norton, M., 2011. The Artful Dodger: Answering the Wrong Question the Right Way. *Journal of Experimental Psychology: Applied.* 17(2), 139–147.

Schwitzgebel, E., and J. Rust. 2010. Do ethicists and political philosophers vote more often than other professors? *Review of Philosophy and Psychology* 1, 189–199.

S. Shankaran et al. 2012. Childhood outcomes after hypothermia for neonatal encephalopathy. *New England Journal of Medicine* 366, 2085–2092, May 31.

SECTION EXERCISES

SECTION 8.1

8.1 Choose one of the answers given. The null hypothesis is always a statement about a ______________ (sample statistic or population parameter).

8.2 Choose one of the answers in each case. In statistical inference, measurements are made on a ______________ (sample or population), and generalizations are made to a ______________ (sample or population).

TRY **8.3 Boot Camp (Example 1)** Suppose an experiment is done with criminals released from prison in a certain state where the recidivism rate is 40%; that is, 40% of criminals return to prison within three years. One hundred random prisoners are made to attend a "boot camp" for two weeks before their release, and it is hoped that "boot camp" will have a good effect. The null hypothesis is that those attending boot camp have a recidivism rate of 40%, $p = 0.40$. Report the alternative hypothesis in words and in symbols.

8.4 Scrubs A research hospital tries a new antibiotic scrub before surgery to see whether it can lower the rate of infections of surgical sites. The old rate of infection is 4%. The null hypothesis is that the proportion of infections is 0.04, $p = 0.04$. Give the alternative hypothesis in words and symbols.

8.5 Lotto A biased lotto draws even numbers faster than odd numbers. A token is drawn 50 times, and 35 even numbers come up.

a. Pick the correct null hypothesis:

i. $\hat{p} = 0.50$ ii. $\hat{p} = 0.70$ iii. $p = 0.50$ iv. $p = 0.70$

b. Pick the correct alternative hypothesis:

i. $\hat{p} = 0.50$ ii. $\hat{p} = 0.70$ iii. $p > 0.50$ iv. $p > 0.70$

8.6 Juice A student is tested to see whether she can tell fresh juice from bottled juice. There are 40 trials (half with fresh juice and half with bottled juice), and she gets 28 right.

a. Pick the correct null hypothesis:

i. $p = 0.40$ ii. $p = 0.50$ iii. $p = 0.56$ iv. $p = 0.70$

b. Pick the correct alternative hypothesis:

i. $\hat{p} \neq 0.40$ ii. $p > 0.50$ iii. $p > 0.56$ iv. $p \neq 0.70$

8.7 Cancer Survival Rate The proportion of people who live after fighting cancer is 0.75. Suppose there is a new therapy that is used to increase the survival rate. Use the parameter p to represent the population portion of people who survive after fighting cancer. For a hypothesis test of the therapy's effectiveness, researchers use a null hypothesis of $p = 0.75$.

Pick the correct alternative hypothesis.

i. $p > 0.75$ ii. $p < 0.75$ iii. $p \neq 0.75$

8.8 Diet Recommendation A new food supplement is being tested to see whether it can reduce obesity in the population of concern. At present, the percentage of obese people is 60. The null hypothesis is that p (the proportion of the population using the food supplement that is still obese) is 0.60. Pick the correct alternative hypothesis.

i. $p \neq 0.60$ ii. $p < 0.60$ iii. $p > 0.60$

TRY **8.9 Coin Flips (Example 2)** A coin is flipped 30 times and comes up heads 18 times. You want to test the hypothesis that the coin does not come up 50% heads in the long run.

Pick the correct null hypothesis for this test.

i. H_0: $\hat{p} = 0.60$

ii. H_0: $p = 0.50$

iii. H_0: $p = 0.60$

iv. H_0: $\hat{p} = 0.50$

8.10 Coin Flips A fair coin is flipped 80 times, and it turns up heads 30 times. You want to test the hypothesis that the coin does not turn up heads one-half of the time. Pick the correct null hypothesis.

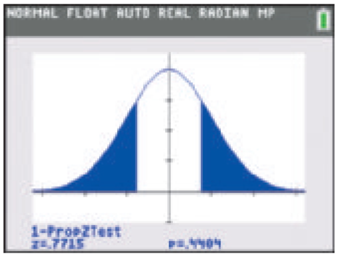
i. H_0: $p = 3/8$ ii. H_0: $p = 1/2$ iii. H_0: $\hat{p} = 3/8$ iv. H_0: $\hat{p} = 1/2$

TRY **8.11 ESP (Example 3)** We are testing someone who claims to have ESP by having that person predict whether a coin will come up heads or tails. The null hypothesis is that the person is guessing and does not have ESP, and the population proportion of success is 0.50. We test the claim with a hypothesis test, using a significance level of 0.05. Select an answer and fill in the blank.

The probability of concluding that the person has ESP when in fact she or he (does/does not) ________________ have ESP is ________________.

* **8.12 Credit Card Application** An applicant is filling out a credit card application form that has 10 multiple-choice questions (about income and asset details), each with three possible answers. The banker's null hypothesis is that the applicant is answering randomly, and the population proportion of correct information is 0.33. Suppose we do a test with a significance level of 0.05. Write a sentence describing the significance level in the context of the hypothesis test.

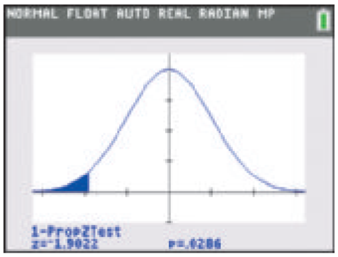
8.13 Dropouts According to *Time* magazine (June 11, 2012), the dropout rate for all college students with loans is 30%. Suppose that 65 out of 200 random college students with loans drop out.



a. Give the null and alternative hypotheses to test that the dropout rate is not 30%.

b. Report the test statistic (z) from the output given.

8.14 SUVs According to *Time* magazine (June 11, 2012), 33% of all cars sold in the United States are SUVs. Suppose a random sample of 500 recently sold cars shows that 145 are SUVs.



a. Write the null and alternative hypotheses to test that fewer than 33% of cars sold are SUVs.

b. Report the value of the test statistic (z) from the figure.

TRY **8.15 Boot Camp, Again (Example 4)** Refer to Exercise 8.3. Suppose 100 people attend boot camp and 44 of them return to prison within three years). The population recidivism rate for the whole state is 40%.

a. What is $\hat{p}$, the sample proportion of successes? (It is somewhat odd to call returning to prison a success.)

b. What is p_0, the hypothetical proportion of success under the null hypothesis?

c. What is the value of the test statistic? Explain in context.

8.16 Scrubs Refer to Exercise 8.4. Suppose a sample of 600 surgeries with the new scrub shows 18 infections. Find the value of the test statistic, z, and explain its meaning in context. The old infection rate was 4%.

TRY **8.17 Coke vs. Pepsi (Example 5)** Suppose you are testing someone to see whether she or he can tell Coke from Pepsi, and you are using 20 trials, half with Coke and half with Pepsi. The null hypothesis is that the person is guessing.

a. About how many should you expect the person to get right under the null hypothesis that the person is guessing?

b. Suppose person A gets 13 right out of 20, and person B gets 18 right out of 20. Which will have a smaller p-value, and why?

8.18 Bilderberg Group Bilderberg Group is a private conference comprising American and European political elites who make up one-third and two-thirds of the group, respectively. Suppose you are looking at conferences, each with 150 members, in the Bilderberg Group. The null hypothesis is that the probability of a European being selected into the club is 67%.

a. How many Europeans would you expect in a conference of 150 people if the null hypothesis is true?

b. Suppose Conference A contains 103 Europeans out of 150 and Conference B contains 108 Europeans out of 150. Which will have a smaller p-value and why?

TRY **8.19 Coke vs. Pepsi (Example 6)** Suppose you are testing someone to see if he or she can tell Coke from Pepsi, and you are using 20 trials, half with Coke and half with Pepsi. The null hypothesis is that the person is guessing. The alternative is one-sided: H_a: $p_0 > 0.5$. The person gets 13 right out of 20. The p-value comes out to be 0.090. Explain the meaning of the p-value.

* **8.20 Helmets** Suppose we are testing bikers to see if the rate of use of helmets has changed from a previous value of 79%. Suppose that in our random sample of 300 bikers, we see that 255 wear helmets.

a. About how many out of 300 would we expect to be wearing a helmet if the proportion who wear helmets is unchanged?

b. We observe 255 bikers out of a random sample of 300 wearing helmets. The p-value is 0.33. Explain the meaning of the p-value.

8.21 Cheating? A professor creates two versions of a 20-question multiple-choice quiz. Each question has four choices. One student got a score of 19 out of 20 *for the version of the test given to the person sitting next to her.* The professor thinks the student was copying another exam. The student admits that he hadn't studied for the test, but he says he was simply guessing on each question and just got lucky. For the professor, the null hypothesis is that $p = 0.25$, where p is the probability that the student chooses the correct answer if just guessing, and the alternative is $p > 0.25$. Would you say that the p-value for this hypothesis test will be high or low? Explain.

* **8.22 Random Answering** A 25-question multiple-choice questionnaire has four choices for each question. Suppose an applicant is an expert and knows all the correct answers. The employer carries out a hypothesis test to determine whether a job applicant was answering randomly. The null hypothesis is $p = 0.25$, where p is the probability of correct answer.

a. Which of the following describes the value of the z-test statistic that is likely to result? Explain your choice.
   i. The z-test statistic will be far from 0.
   ii. The z-test statistic will be close to 0.

b. Which of the following describes the p-value that is likely to result? Explain your choice.
   i. The p-value will be large.
   ii. The p-value will be small.

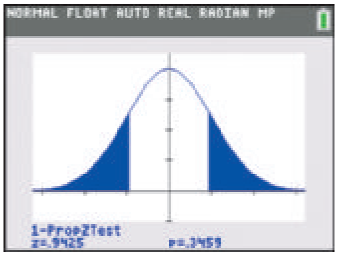
SECTION 8.2

TRY **8.23 Dreaming (Example 7)** A 2003 study of dreaming found that out of a random sample of 113 people, 92 reported dreaming in color. However, the proportion of people who reported dreaming in color that was established in the 1940s was 0.29 (Schwitzgebel 2003). Check to see whether the conditions for using a one-proportion z-test are met assuming the researcher wanted to see whether the proportion dreaming in color had changed since the 1940s.

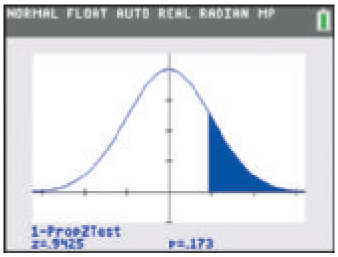
8.24 Nationality Discrimination Refugees make up about 20% of the population in a country. However, only 3% of the 1500 applications rejected by an employment agency are those of refugees. Experts might argue that if the agency hired people regardless of their nationality, the distribution of nationalities would be the same as though they had hired people at random from the country's population. Check whether the conditions for using the one-proportion z-test are met.

TRY **8.25 Marriage Obsolete (Example 8)** When asked whether marriage is becoming obsolete, 782 out of 2004 randomly selected adults answering a Pew Poll said yes. We are testing the hypothesis that the population proportion that believes marriage is becoming obsolete is *more than* 38% using a significance level of 0.05. One of the following figures is correct. Indicate which graph matches the alternative hypothesis, $p > 0.38$. Report and interpret the correct p-value.

(A)



(B)

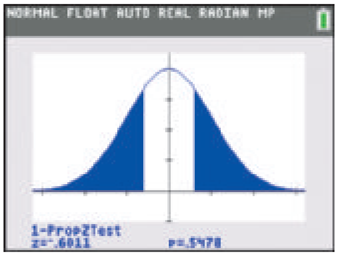


* **8.26 Anxiety Disorders** According to *Time* magazine (December 5, 2011), 18% of adult Americans suffer from an anxiety disorder. Suppose a test of 300 random college students showed that 50 suffered from an anxiety disorder.

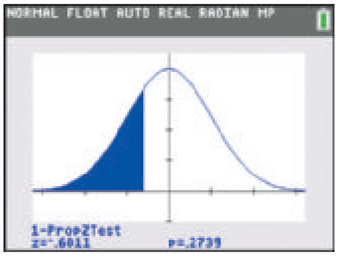
a. How many out of 300 would you expect to have an anxiety disorder if the 18% is correct?

b. Suppose you are testing the hypothesis that the population proportion of college students suffering from an anxiety disorder *is not* 0.18 at the 0.05 significance level. Choose the correct figure and interpret the p-value.

(A)

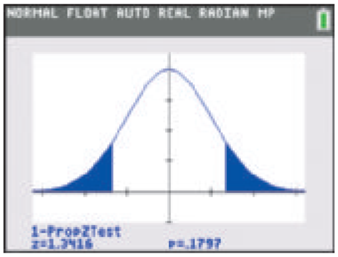


(B)

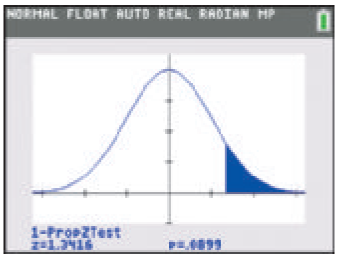


8.27 Coke vs. Pepsi A taste test is done to see whether a person can tell Coke from Pepsi. In each case, 20 random and independent trials are done (half with Pepsi and half with Coke) in which the person determines whether she or he is drinking Coke or Pepsi. One person gets 13 right out of 20 trials. Which of the following is the correct figure to test the hypothesis that the person can tell the difference? Explain your choice.

(A)

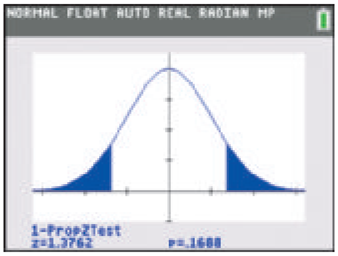


(B)

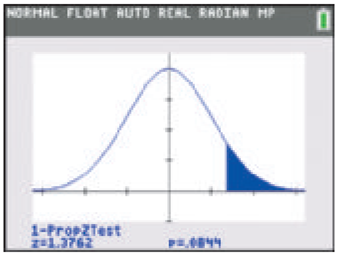


8.28 Seat Belts Suppose we are testing people to see whether the rate of use of seat belts has changed from a previous value of 88%. Suppose that in our random sample of 500 people we see that 450 have the seat belt fastened. Which of the following figures has the correct p-value for testing the hypothesis that the proportion who use seat belts has changed? Explain your choice.

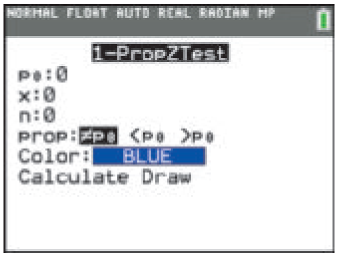
(A)



(B)



TRY **8.29 Sleep Walking (Example 9)** According to *Time* magazine (May 28, 2012), 33% of people in the United States have sleep-walked at least once in their lives. Suppose a random sample of 200 people showed that 42 reported sleepwalking. Carry out the first two steps of a hypothesis test that will test whether the proportion of people who have sleep walked is 0.33. Use a significance level of 0.05. Explain how you would fill in the required TI calculator entries for p_0, x, n, and prop shown in the figure.



8.30 Women CEOs The percentage of female CEOs in 2013 in Fortune 500 companies was 4%, according to *Time* magazine (March 18, 2013). Suppose a student did a survey of 250 randomly selected large companies (not Fortune 500 companies), and 15 of them had women CEOs.

a. About how many of the 250 should we expect to have female CEOs if the proportion of female CEOs is 4%, as it is for the Fortune 500 companies?

b. Carry out the first two steps of a hypothesis test that will test whether the population proportion of women CEOs is not 0.04.

TRY **8.31 Sleep Walking, Again (Example 10)** According to *Time* magazine (May 28, 2012), 33% of people in the U.S. have sleep-walked at least once in their lives. Suppose a random sample of 200 people showed that 42 reported sleepwalking. The first two steps were asked for in Exercise 8.29. Use the computer output provided to carry out the third and fourth steps of a hypothesis test that will test whether the proportion of people who have sleepwalked is 0.33. Use a significance level of 0.05.

Options

Hypothesis test results:
p : Proportion of successes
H_0 : p = 0.33
H_A : p ≠ 0.33

Proportion	Count	Total	Sample Prop.	Std. Err.	Z-Stat	P-value
p	42	200	0.21	0.03324906	-3.6091246	0.0003

8.32 Women CEOs, Again The percentage of female CEOs in 2013 in Fortune 500 companies was 4%, according to *Time* Magazine (March 18, 2013). Suppose a student did a survey of 250 randomly selected large companies (not Fortune 500 companies), and 15 of them had women CEOs. In Exercise 8.30 you carried out the first two steps to test the hypothesis that the population proportion is not 0.04. Now carry out the last two steps, using the computer output provided and a significance level of 0.05.

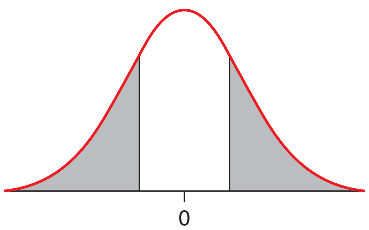
Options

Hypothesis test results:
p : Proportion of successes
H_0 : p = 0.04
H_A : p ≠ 0.04

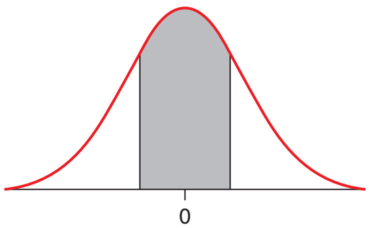
Proportion	Count	Total	Sample Prop.	Std. Err.	Z-Stat	P-value
p	15	250	0.06	0.012393547	1.6137431	0.1066

8.33 p-Values For each graph, indicate whether the shaded area could represent a p-value. Explain why or why not. If yes, state whether the area could represent the p-value for a one-sided or a two-sided alternative hypothesis.

(A)

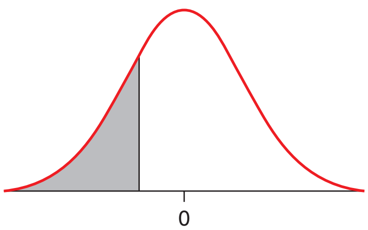


(B)

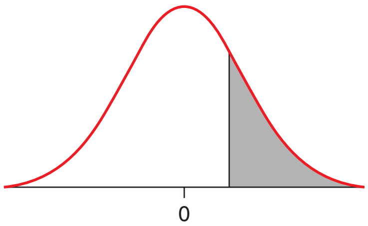


8.34 p-Values For each graph, state whether the shaded area could represent a p-value. Explain why or why not. If yes, state whether the area could represent the p-value for a one-sided or a two-sided alternative hypothesis.

(A)



(B)



8.35 Gun Control Historically, the percentage of U.S. residents who support stricter gun control laws has been 52%. A recent Gallup Poll of 1011 people showed 495 in favor of stricter gun control laws. Assume the poll was given to a random sample of people. Test the claim that the proportion of those favoring stricter gun control has changed. Perform a hypothesis test, using a significance level of 0.05. *See page 423 for guidance.* Choose one of the following conclusions:

i. The percentage is not significantly different from 52%. (A significant difference is one for which the p-value is less than or equal to 0.050.)
ii. The percentage is significantly different from 52%.

8.36 Salary Deduction A random survey showed that 1680 out of 2015 surveyed employees favored salary deduction for late attendance.

a. Test the hypothesis that more than half of the employees favor salary deduction using a significance level of 0.05. Label each step.
b. If there were a vote by the public about whether to discontinue the salary deduction, would it pass? (Base your answer on part a.)

8.37 Global Warming Historically (from about 2001 to 2005), about 58% of Americans believed that Earth's temperature was rising ("global warming"). A March 2010 Gallup Poll sought to determine whether this proportion had changed. The poll interviewed 1014 adult Americans, and 527 said they believed that global warming was real. (Assume these 1014 adults represented a simple random sample.)

a. What percentage in the sample believed global warming was real in 2010? Is this more or less than the historical 58%?
b. Test the hypothesis that the proportion of Americans who believe global warming is real has changed. Use a significance level of 0.05.
c. Choose the correct interpretation:
   i. In 2010, the percentage of Americans who believe global warming is real is not significantly different from 58%.
   ii. In 2010, the percentage of Americans who believe global warming is real has changed from the historical level of 58%.

8.38 Plane Crashes According to one source, 50% of plane crashes are due at least in part to pilot error (http://www.planecrashinfo.com). Suppose that in a random sample of 100 separate airplane accidents, 62 of them were due to pilot error (at least in part.)

a. Test the null hypothesis that the proportion of airplane accidents due to pilot error is not 0.50. Use a significance level of 0.05.
b. Choose the correct interpretation:
   i. The percentage of plane crashes due to pilot error is not significantly different from 50%.
   ii. The percentage of plane crashes due to pilot error is significantly different from 50%.

TRY **8.39 Mercury in Freshwater Fish (Example 11)** Some experts believe that 20% of all freshwater fish in the United States have such high levels of mercury that they are dangerous to eat. Suppose a fish market has 250 fish tested, and 60 of them have dangerous levels of mercury. Test the hypothesis that this sample is *not* from a population with 20% dangerous fish. Use a significance level of 0.05.

Comment on your conclusion: Are you saying that the percentage of dangerous fish is definitely 20%? Explain.

8.40 Taxes Suppose a poll is taken that shows that 281 out of 500 randomly selected, independent people believe the rich should pay more taxes than they do. Test the hypothesis that a majority (more than 50%) believe the rich should pay more taxes than they do. Use a significance level of 0.05.

8.41 Morse's Proportion of t's Samuel Morse determined that the percentage of t's in the English language in the 1800s was 9%. A random sample of 600 letters from a current newspaper contained 48 t's. Using the 0.10 level of significance, test the hypothesis that the proportion of t's in this modern newspaper is 0.09.

8.42 Morse's Proportion of a's Samuel Morse determined that the percentage of a's in the English language in the 1800s was 8%. A random sample of 600 letters from a current newspaper contained 60 a's. Using the 0.10 level of significance, test the hypothesis that the proportion of t's in this modern newspaper is 0.09.

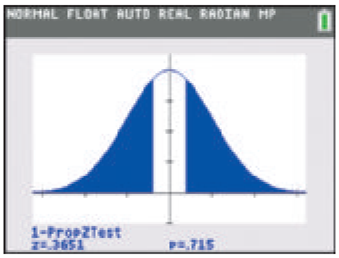
SECTION 8.3

TRY **8.43 p-Values (Example 11)** A researcher carried out a hypothesis test using a two-sided alternative hypothesis. Which of the following *z*-scores is associated with the smallest p-value? Explain.

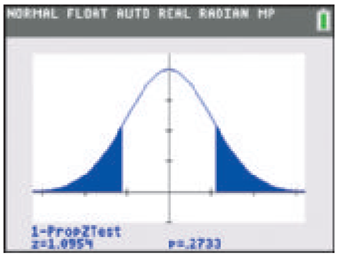
i. $z = 0.50$ ii. $z = 1.00$ iii. $z = 2.00$ iv. $z = 3.00$

8.44 Coin Flips A test is conducted in which a coin is flipped 30 times to test whether the coin is unbiased. The null hypothesis is that the coin is fair. The alternative is that the coin is not fair. One of the accompanying figures represents the p-value after getting 16 heads out of 30 flips, and the other represents the p-value after getting 18 heads out of 30 flips. Which is which, and how do you know?

(A)



(B)



TRY **8.45 Errors with Pennies (Example 12)** Suppose you are spinning pennies to test whether you get biased results. When you reject the null hypothesis when it is actually true, that is often called the first kind of error. The second kind of error is when the null is false and you fail to reject. Report the first kind of error and the second kind of error.

8.46 Errors with Cheese Suppose you are testing someone to see whether he or she can tell goat cheese from cheddar cheese. You use many bite-sized cubes selected randomly, half from goat cheese and half from cheddar cheese. The taster is blindfolded. The null hypothesis is that the taster is just guessing and should get about half right. When you reject the null hypothesis when it is actually true, that is often called the first kind of error. The second kind of error is when the null is false and you fail to reject. Report the first kind of error and the second kind of error.

8.47 Blackstone on Errors in Trials Sir William Blackstone (1723–1780) wrote influential books on common law. He made this statement: "All presumptive evidence of felony should be admitted cautiously; for the law holds it better that ten guilty persons escape, than that one innocent party suffer."

Keep in mind that the null hypothesis in criminal trials is that the defendant is not guilty. State which of these errors (in blue) is the first type of error (rejecting the null hypothesis when it is actually true) and which is the second type of error.

8.48 Alpha By establishing a small value for the significance level, are we guarding against the first type of error (rejecting the null hypothesis when it is true) or guarding against the second type of error?

TRY **8.49 Flaws (Example 13)** A person spinning a 1962 penny gets 10 heads out of 50 spins. Because she gets a p-value of 0.00002, she says she has proved the coin is biased. What is the flaw in the statement and how would you correct it?

8.50 Defects The null hypothesis on rolling a die is a multiple of 3, and the proportion of multiples of 3 is 0.33. A person rolling the die 12 times gets 8 successful rolls out of 12. The person says that he knows how to roll a die, because the one-tailed p-value from the one-proportion z-test is 0.006, and he is using a significance level of 0.10. What is wrong with his approach?

8.51 Which Method? A proponent of a new proposition on a ballot wants to know whether the proposition is likely to pass. Suppose a poll is taken, and 580 out of 1000 randomly selected people support the proposition. Should the proponent use a hypothesis test or a confidence interval to answer this question? Explain. If it is a hypothesis test, state the hypotheses and find the test statistic, p-value, and conclusion. If a confidence interval is appropriate, find the approximate 95% confidence interval. In both cases, assume that the necessary conditions have been met.

8.52 Which Method? A proponent of a new proposition on a ballot wants to know the population percentage of people who support the bill. Suppose a poll is taken, and 580 out of 1000 randomly selected people support the proposition. Should the proponent use a hypothesis test or a confidence interval to answer this question? Explain. If it is a hypothesis test, state the hypotheses and find the test statistic, p-value, and conclusion. Use a 5% significance level. If a confidence interval is appropriate, find the approximate 95% confidence interval. In both cases, assume that the necessary conditions have been met.

* **8.53 Effectiveness of Incentives** A teacher is interested in testing whether offering students some form of incentive improves their GPA. She collects data and performs a hypothesis test to test whether the probability of getting the maximum GPA is greater with an incentive than without. Her null hypothesis is that the probability of getting the maximum GPA is the same with or without an incentive. The alternative is that this probability is greater. She gets a p-value from her hypothesis test of 0.07. Which of the following is the best interpretation of p-value?

i. The p-value is the probability of getting exactly the result obtained, assuming that incentives *are* effective in this context.
ii. The p-value is the probability that incentives are *not* effective in this context.
iii. The p-value is the probability of getting a result as extreme as or more extreme than the one obtained, assuming that incentives *are* effective in this context.
iv. The p-value is the probability of getting a result as extreme as or more extreme than the one obtained, assuming that incentives are *not* effective in this context.
v. The p-value is the probability of getting exactly the result obtained, assuming that incentives *are* effective in this context.

8.54 Is it acceptable practice to look at your research results, note the direction of the difference, and then make the alternative hypothesis one-sided in order to achieve a significant difference? Explain.

8.55 If we reject the null hypothesis, can we claim to have *proved* that the null hypothesis is false? Why or why not?

8.56 If we do not reject the null hypothesis, is it valid to say that we *accept* the null hypothesis? Why or why not?

8.57 When a person stands trial for murder, the jury is instructed to assume that the defendant is innocent. Is this claim of innocence an example of a null hypothesis, or is it an example of an alternative hypothesis?

8.58 When, in a criminal court, a defendant is found "not guilty," is the court saying with certainty that he or she is innocent? Explain.

* **8.59 Arthritis** A magazine advertisement claims that wearing a magnetized bracelet will reduce arthritis pain in those who suffer from arthritis. A medical researcher tests this claim with 233 arthritis sufferers randomly assigned either to wear a magnetized bracelet or to wear a placebo bracelet. The researcher records the proportion of each group who report relief from arthritis pain after 6 weeks. After analyzing the data, he fails to reject the null hypothesis. Which of the following are valid interpretations of his findings? There may be more than one correct answer.

a. The magnetized bracelets are not effective at reducing arthritis pain.
b. There's insufficient evidence that the magnetized bracelets are effective at reducing arthritis pain.
c. The magnetized bracelets had exactly the same effect as the placebo in reducing arthritis pain.
d. There were no statistically significant differences between the magnetized bracelets and the placebos in reducing arthritis pain.

* **8.60 Sugar and Arthritis** An arthritis diet claims that the disease can be relieved by reducing sugar from the diet. To test this claim, a researcher randomly assigns arthritis patients to two groups. Both groups eat the same amount of calories, but one group eats almost no sugar and the other group includes sugar in their meal. After 3 months, the doctor tests the claim that the sugar-free diet is better than the usual diet. She records the proportion of each group that got relieved of almost 10% of their problem. She then announced that she failed to reject the null hypothesis. Which of the following are valid interpretations of her findings?

a. The sugar-free diet was less effective than the normal diet.
b. The researcher did not see enough evidence to conclude that the sugar-free diet was more effective.
c. The sugar-free diet and the normal diet were equally effective.
d. There were no significant differences in effectiveness between the sugar-free diet and normal diet.

SECTION 8.4

8.61 When comparing two sample proportions with a two-sided alternative hypothesis, all other factors being equal, will you get a smaller p-value if the sample proportions are close together or if they are far apart? Explain.

8.62 When comparing two sample proportions with a two-sided alternative hypothesis, all other factors being equal, will you get a smaller p-value with a larger sample size or a smaller sample size? Explain.

g **8.63 Treatment for CLL** Furman et al. (2014) reported on a study of patients with recurring chronic lymphocytic leukemia (CLL). The study was randomized, double-blind, and placebo-controlled. After 12 months, 101 of the 110 patients assigned to the combination treatment of idelalisib (idel) with rituximab (rit) were still alive, and 88 of the 110 patients assigned to placebo with rit were still alive. Perform a hypothesis test to test whether those who take idel have a *better* chance of surviving than those taking the placebo. Use a level of significance of 0.05. Can we conclude that idel *causes* an increased chance of survival? *See page 424 for guidance.*

8.64 Vaccine for Diarrhea A vaccine to prevent severe rotavirus gastroenteritis (diarrhea) was given to African children within the first year of life as part of a drug study. The study reported that of the 3298 children randomly assigned the vaccine, 63 got the virus. Of the 1641 children randomly assigned the placebo, 80 got the virus. (Source: Madhi et al., Effect of human rotavirus vaccine on severe diarrhea in African infants, *New England Journal of Medicine*, vol. 362: 289–298, January 28, 2010)

a. Find the sample percentage of children who caught the virus in each group. Is the sample percent lower for the vaccine group, as investigators hoped?
b. Determine whether the vaccine is effective in reducing the chance of catching the virus, using a significance level of 0.05. Steps 1 and 2 of the hypothesis-testing procedure are given. Complete the question by doing steps 3 and 4.

Step 1: H_0: $p_v = p_p$ (p_v is the proportion that got the virus among those who took the vaccine, and p_p is the proportion that got the virus among those who took the placebo.) H_a: $p_v < p_p$

Step 2: Although we don't have a random sample, we do have random assignment to groups.

$$\hat{p} = \frac{63 + 80}{3298 + 1641} = \frac{143}{4939} = 0.028953$$

$n_p \times \hat{p} = 3298 \times 0.028953 = 95.49$, which is more than 10

$n_p \times \hat{p} = 1641 \times 0.028953 = 47.51$, which is more than 10

(and the other two products are larger)

TRY **8.65 Nicotine Gum (Example 14)** A study used nicotine gum to help people quit smoking. The study was placebo-controlled, randomized, and double-blind. Each participant was interviewed after 28 days, and success was defined as being abstinent from cigarettes for 28 days. The results showed that 174 out of 1649 people using the nicotine gum succeeded, and 66 out of 1648 using the placebo succeeded. Although the sample was not random, the assignment to groups was randomized. (Source: Shiffman et al., Quitting by gradual smoking reduction using nicotine gum: A randomized controlled trial, *American Journal of Preventive Medicine*, vol. 36, issue 2, February, 2009)

a. Find the proportion of people using nicotine gum that stopped smoking and the proportion of people using the placebo that stopped smoking, and compare them. Is this what the researchers had expected?
b. Find the observed value of the test statistic, assuming that the conditions for a two-proportion z-test hold.

8.66 Hypothermia for Babies Shankaran and colleagues (2012) reported the results of a randomized trial of whole-body hypothermia (cooling) for neonatal hypoxic–ischemic encephalopathy (brain problems in babies due to lack of oxygen). Twenty-seven of the 97 infants randomly assigned to hypothermia died and 41 of the 93 infants in the control group died.

a. Find and compare the sample percentages that died for these two groups.
b. Test the hypothesis that the death rate was less for those treated with hypothermia at a 0.05 significance level.

8.67 Criminology and Therapy In London, Ontario, Canada, investigators (Leschied and Cunningham 2002) performed a randomized experiment in which 409 juvenile delinquents were randomly assigned to either multisystemic therapy (MST) or just probation (control group). Of the 211 assigned to therapy, 87 had criminal convictions within 12 months. Of the 198 in the control group, 74 had criminal convictions within 12 months. Determine whether the therapy caused significantly fewer arrests at a 0.05 significance level. Start by comparing the sample percentages.

8.68 Criminology and Counseling Feder and Dugan (2002) reported a study in which 404 domestic violence defendants were randomly assigned to counseling and probation (the experimental group) or just probation (the control group). Out of 230 people in the counseling group, 55 were arrested within 12 months. Out of 174 people assigned to probation, 42 were arrested within 12 months. Determine whether counseling lowered the arrest rate; use a 0.05 significance level. Start by comparing the percentages.

TRY **8.69 Smiling and Gender (Example 15)** In a 1997 study, people were observed for about 10 seconds in public places, such as malls and restaurants, to determine whether they smiled during the randomly chosen 10-second interval. The table shows the results for comparing males and females. (Source: M. S. Chappell, Frequency of public smiling over the life span, *Perceptual and Motor Skills*, vol. 45: 474, 1997)

	Male	Female
Smile	3269	4471
No Smile	3806	4278

a. Find and compare the sample percentages of women who were smiling and men who were smiling.
b. Treat this as though it were a random sample, and test whether there are differences in the proportion of men and the proportion of women who smile. Use a significance level of 0.05.
c. Explain why there is such a small p-value even though there is such a small difference in sample percentages.

* **8.70 Smiling and Age** Refer to the study discussed in Exercise 8.69. The accompanying table shows the results of the study for different age groups.

	Age Range				
	0–10	**11–20**	**21–40**	**41–60**	**61+**
Smile	1131	1748	1608	937	522
No Smile	1187	2020	3038	2124	1509

a. For each age group, find the percentage who were smiling.

b. Treat this as a random sample of people, and merge the groups 0–10 and 11–20 into one age group (0–20) and the groups 21–40, 41–60, and 61+ into another age group (21−65+). Then determine whether these two age groups have different proportions of people who smile in the general population, using a significance level of 0.05. Please comment on the results.

CHAPTER REVIEW EXERCISES

8.71 Choosing a Test and Naming the Populations For each of the following, state whether a one-proportion z-test or a two-proportion z-test would be appropriate, and name the populations.

a. A marketing manager asks a random sample of cricketers and a random sample of soccer players whether they support television commercials during games. The manager wants to determine whether the proportion of cricketers who support commercials is less than the proportion of soccer players who support these.

b. A survey takes a random sample to determine the proportion of students in India who support the Women's Reservation Bill.

8.72 Choosing a Test and Naming the Populations For each of the following, state whether a one-proportion z-test or a two-proportion z-test would be appropriate, and name the populations.

a. The minimum quantity of ingredient X in a product is 38% of total weight. A random sample of products manufactured in a factory is examined to see whether the rate of ingredient X in the product is significantly higher than 38%.

b. A student watches a random sample of male and female car drivers parking in a parking lot. Some drivers park the car as they drive in and some park after reversing the car. He wants to compare the proportions of male and female drivers who park the car after reversing.

8.73 Choosing a Test and Giving the Hypothesis Give the null and alternative hypothesis for each test, and state whether a one-proportion z-test or a two-proportion z-test would be appropriate.

a. You test a random sample of eighth-grade students who play daily for 3 hours and devote the same time to homework, comparing boys and girls.

b. You test a person to see whether she can tell butter on bread from margarine spread on bread. You give her 20 toast bits selected randomly (half with butter and half with margarine) and record the proportion she gets correct to test the hypothesis.

8.74 Choosing a Test and Naming the Population(s) In each case, choose whether the appropriate test is a one-proportion z-test or a two-proportion z-test. Name the population(s).

a. A person is observed to check whether he or she can predict the results of the die roll better than chance alone.

b. A researcher wants to know whether a new skin cream effectively reduces skin rashes compared to an old cream.

c. A survey agency conducts a survey in a city to find out whether the residents like dark and white chocolate equally.

d. A news agency takes a random sample of all corporate professionals to see whether more than 50% approve of the new tax regime.

e. A teacher takes a random sample of students in statistics class to find out whether girls or boys are more likely to remember the conditions for applying z-test.

8.75 Cola Taste Test A student who claims he can tell cola A from cola B is blindly tested with 20 trials. At each trial, cola A or cola B is randomly chosen and presented to the student, who must correctly identify the cola. The experiment is designed so that the student will have exactly 10 sips from each cola. He gets 6 identifications right out of 20. Can he tell cola A from cola B at the 0.05 level of significance? Explain.

8.76 Butter Taste Test A man is tested to determine whether he can tell butter from margarine. He is blindfolded and given small bites of English muffin to identify. At each trial, an English muffin with either butter or margarine is randomly chosen. The experiment is designed so that he will have exactly 15 bites with butter and 15 with margarine. He gets 14 right out of 30. Can he tell butter from margarine at the 0.05 level? Explain.

* **8.77 Biased Coin?** A study is done to see whether a coin is biased. The alternative hypothesis used is two-sided, and the obtained z-value is 2. Assuming that the sample size is sufficiently large and that the other conditions are also satisfied, use the Empirical Rule to approximate the p-value.

* **8.78 Biased Coin?** A study is done to see whether a coin is biased. The alternative hypothesis used is two-sided, and the obtained z-value is 1. Assuming that the sample size is sufficiently large and that the other conditions are also satisfied, use the Empirical Rule to approximate the p-value.

8.79 ESP A statistician studying ESP tests 500 students. Each student is asked to predict the outcome of a large number of dice rolls. For each student, a hypothesis test using a 10% significance level is performed. If the p-value for the student is less than or equal to 0.10, the researcher concludes that the student has ESP. Out of 500 students who do *not* have ESP, about how many could you expect the statistician to declare *do* have ESP?

8.80 Coin Flips Suppose you tested 50 coins by flipping each of them many times. For each coin, you perform a significance test with a significance level of 0.05 to determine whether the coin is biased. Assuming that none of the coins is biased, about how many of the 50 coins would you expect to appear biased when this procedure is applied?

8.81 Education An education board declared that 64% of *all* students who appeared for examination had passed in 2015, whereas 58% passed in 2016. Why can you not use this report for a hypothesis test?

8.82 Inference Harry was tested to see whether he could tell cotton fabric from rayon fabric by touch alone. He got 18 instances right out of 25, and the p-value was 0.174. Explain what is wrong with the following inference, and write a correct inference: "We proved that Harry cannot tell the difference between cotton and rayon fabrics by touch."

* **8.83 Gun Control Laws** The Gallup organization frequently conducts polls in which they ask the following question:

"In general, do you feel that the laws covering the sale of firearms should be made more strict, less strict, or kept as they are now?"

In February 1999, 60% of those surveyed said "more strict," and on April 26, 1999, shortly after the Columbine High School shootings, 66% of those surveyed said "more strict."

a. Assume that both polls used samples of 560 people. Determine the number of people in the sample who said "more strict" in February 1999, before the school shootings, and the number who said "more strict" in late April 1999, after the school shootings.

b. Do a test to see whether the proportion that said "more strict" is statistically significantly different in the two different surveys, using a *significance level of 0.01*.

c. Repeat the problem, assuming that the sample sizes were both 1120.

d. Comment on the effect of different sample sizes on the p-value and on the conclusion.

8.84 Weight Loss in Men Many polls have asked people whether they are trying to lose weight. A Gallup Poll in November of 2008 showed that 22% of men said they were seriously trying to lose weight. In 2006, 24% of men (with the same average weight of 194 pounds as the men polled in 2008) said they were seriously trying to lose weight. Assume that both samples contained 500 men.

a. Determine how many men in the sample from 2008 and how many in the sample from 2006 said they were seriously trying to lose weight.

b. Determine whether the difference in proportions is significant at the 0.05 level.

c. Repeat the problem with the same proportions but assuming both sample sizes are now 5000.

d. Comment on the different p-values and conclusions with different sample sizes.

8.85 Literacy in 2015 In March 2016, the UNESCO Institute for Statistics (UIS) reported that the literacy rate in Zimbabwe was 88.5% for males and 84.6% for females. Would it be appropriate to draw a two-proportion z-test to determine whether the rates for males and females were significantly different (assuming we knew the total number of males and females)? Explain.

8.86 Annual Sports Meet For a school's annual sports meet, 56% of boys enrolled for track events and 44% of boys enrolled for water sports. Also, 45% of girls enrolled for track events and 55% of girls enrolled for water sports. Would it be appropriate to do a two-proportion z-test to determine whether the proportions of boys and girls enrolling for track events were significantly different (assuming we know the number of boy and girl students)? Explain.

8.87 Work from Home In June 30, 2012, the *Ventura County Star* reported that 63% of employers allow employees to work from home sometimes (this is up from 34% in 2005). Suppose a random sample of 400 employers shows that 235 allow some work from home. Test the hypothesis that the percentage is not 63%, using a significance level of 0.10.

8.88 Flex Time In 2012 the *Ventura County Star* reported that 77% of employers allow employees to use flex time and periodically change their start and quit times (this is up from 66% in 2005). Suppose a random sample of 200 employers shows that 130 allow flex time. Test the hypothesis that the percentage is less than 77%, using a significance level of 0.10.

8.89 Wording of Polls A poll in California (done by the Public Policy Institute) asked whether the government should regulate greenhouse gases, and 751 out of 1138 likely voters said yes. However, when a different polling agency asked whether stricter environmental controls are worth the cost, 523 of 1138 likely voters said yes; the alternative was that the laws hurt the economy and cost too many jobs. (Source: *Ventura County Star*, March 21, 2013)

a. Find both sample proportions and compare them. Comment on the similarities of the two questions.

b. Determine whether the two sample proportions are significantly different at a 0.05 significance level. Comment on the effect of changing the wording of this question.

c. Using methods learned in Chapter 7, find a 95% confidence interval for the difference between the two percentages, and interpret it. Does it capture 0? What does that mean?

8.90 Gay Marriage A *Washington Post* Poll (March 18, 2013) and a Pew Poll (March 17, 2013) both claimed to ask a random sample of adults in the United States whether they supported or opposed gay marriage. In the *Washington Post* Poll, 581 supported and 360 opposed gay marriage. In the Pew Poll, 735 supported and 660 opposed gay marriage.

a. Find the percentages supporting gay marriage in these two polls and compare them.

b. Test the hypothesis that the population proportions are not equal at the 0.05 significance level.

c. Using methods learned in Chapter 7, find a 95% confidence interval for the difference between the two percentages, and interpret it. Does it capture 0? What does that show?

* **8.91 Three-Strikes Law** California's controversial "three-strikes law" requires judges to sentence anyone convicted of three felony offenses to life in prison. Supporters say that this decreases crime both because it is a strong deterrent and because career criminals are removed from the streets. Opponents argue (among other things) that people serving life sentences have nothing to lose, so violence within the prison system increases. To test the opponents' claim, researchers examined data starting from the mid-1990s from the California Department of Corrections. "Three Strikes: Yes" means the person had committed three or more felony offenses and was probably serving a life sentence. "Three Strikes: No" means the person had committed no more than two offenses. "Misconduct" includes serious offenses (such as assaulting an officer) and minor offenses (such as not standing for a count). "No Misconduct" means the offender had not committed any offenses in prison.

a. Compare the proportions of misconduct in these samples. Which proportion is higher, the proportion of misconduct for those who had three strikes or that for those who did not have three strikes? Explain.

b. Treat this as though it were a random sample, and determine whether those with three strikes tend to have more offenses than those who do not. Use a 0.05 significance level.

	Three Strikes	
	Yes	**No**
Misconduct	163	974
No Misconduct	571	2214
Totals	734	3188

8.92 Sleep Medicine for Shift Workers Shift workers, who work during the night and must sleep during the day, often become sleepy when working and have trouble sleeping during the day. In a study done at Harvard Medical School, 209 shift workers were randomly divided into two groups; one group received a new sleep medicine (modafinil, or Provigil), and the other group received a placebo. During the study, 54% of the workers taking the placebo and 29% of those taking the medicine reported accidents or near accidents commuting to and from work. Assume that 104 of the people were assigned the medicine and 105 were assigned the placebo. (Source: Czeisler et al., Modafinil for excessive sleepiness associated with shift-work sleep disorder, *New England Journal of Medicine*, vol. 353: 476–486, August 4, 2005)

a. State the null and alternative hypothesis. Is the alternative hypothesis one-sided or two-sided? Explain your choice.

b. Perform a statistical test to determine whether the difference in proportions is significant at the 0.05 level.

8.93 Guns: Two Polls In 2013 a Gallup Poll reported that about 40% of people say they have a gun in their home. In the same year, a Pew Poll reported that about 33% of people say they have a gun in their home. Assume that each used a sample size of 1000. Do these polls disagree? Assume that both polls are based on random samples with independent observations.

a. Test the hypothesis that the two population proportions are different using a significance level of 0.05, and show all four steps.

b. Using methods learned in Chapter 7, estimate the difference in the two population proportions using a 95% confidence interval, and comment on what this says about the null hypothesis in part a.

8.94 Deep Vein Thrombosis Standard anticoagulant therapy (to prevent blood clots) requires frequent laboratory monitoring to prevent internal bleeding. A new procedure using rivaroxaban (riva) was tested because it does not require frequent monitoring. A randomized trial (Einstein-PE Investigators 2012) was carried out, with standard therapy being randomly assigned to half of 4832 patients and riva randomly assigned to the other half. A bad result was recurrence of a blood clot in a vein. Fifty of the 2416 patients on standard therapy had a bad outcome, and 44 of the 2416 patients on riva had a bad outcome.

a. Test the hypothesis that the proportions of bad results are different for riva and standard therapy patients. Use a significance level of 0.05, and show all four steps.

b. Using methods learned in Chapter 7, estimate the difference between the two population proportions using a 95% confidence interval, and comment on how it can be used to evaluate the null hypothesis in part a.

8.95 A friend claims he can predict the numbers that will appear on rolling a standard die. There are six numbers and each one has an equal chance. The parameter, p, is the probability of success, and the null hypothesis is that the friend is just guessing.

a. Which is the correct null hypothesis?

i. $p = 1/3$ ii. $p > 1/3$ iii. $p = 1/6$ iv. $p > 1/6$

b. Which hypothesis best fits the friend's claim? (This is the alternative hypothesis.)

i. $p = 1/6$ ii. $p \neq 1/6$ iii. $p < 1/6$ iv. $p > 1/6$

8.96 A friend claims he can predict how a two-faced token will land. The parameter, p, is the long-run likelihood of success, and the null hypothesis is that the friend is just guessing.

a. Pick the correct null hypothesis.

i. $p \neq 1/2$ ii. $p = 1/2$ iii. $p < 1/2$ iv. $p > 1/2$

b. Which hypothesis best fits the friend's claim? (This is the alternative hypothesis.)

i. $p \neq 1/2$ ii. $p = 1/2$ iii. $p < 1/2$ iv. $p > 1/2$

8.97 Choosing Science Judging on the basis of experience, a school counselor claims that 60% of Australian students choose the science stream for higher studies. Suppose you surveyed 25 randomly selected students, and 18 of them reported having chosen the science stream. The null hypothesis is that the overall proportion of the students who have made such a choice is 60%. What value of the test statistic should you report?

8.98 Choosing Science Refer to exercise 8.97. Suppose 21 out of 25 students chose the science stream for higher studies. The null hypothesis is that the population proportion is 0.60. What value of the test statistic should you report?

* **8.99 Texting While Driving** The mother of a teenager has heard a claim that 25% of teenagers who drive and use a cell phone reported texting while driving. She thinks that this rate is too high and wants to test the hypothesis that fewer than 25% of these drivers have texted while driving. Her alternative hypothesis is that the percentage of teenagers who have texted when driving is less than 25%.

$$H_0: p = 0.25$$
$$H_a: p < 0.25$$

She polls 40 randomly selected teenagers, and 5 of them report having texted while driving, a proportion of 0.125. The p-value is 0.034. Explain the meaning of the p-value in the context of this question.

* **8.100 True/False Test** A teacher giving a true/false test wants to make sure her students do better than they would if they were simply guessing, so she forms a hypothesis to test this. Her null hypothesis is that a student will get 50% of the questions on the exam correct. The alternative hypothesis is that the student is not guessing and should get more than 50% in the long run.

$$H_0: p = 0.50$$
$$H_a: p > 0.50$$

A student gets 30 out of 50 questions, or 60%, correct. The p-value is 0.079. Explain the meaning of the p-value in the context of this question.

8.101 ESP Suppose a friend says he can predict whether a coin flip will result in heads or tails. You test him, and he gets 10 right out of 20. Do you think he can predict the coin flip (or has a way of cheating)? Or could this just be something that occurs by chance? Explain without doing any calculations.

8.102 ESP Again Suppose a friend says he can predict whether a coin flip will result in heads or tails. You test him, and he gets 20 right out of 20. Do you think he can predict the coin flip (or has a way of cheating)? Or could this just be something that is likely to occur by chance? Explain without performing any calculations.

8.103 Does Hand Washing Save Lives? In the mid-1800s, Dr. Ignaz Semmelweiss decided to make doctors wash their hands with a strong disinfectant between patients at a clinic with a death rate of 9.9%. Semmelweiss wanted to test the hypothesis that the death rate would go down after the new hand-washing procedure was used. What null and alternative hypotheses should he have used? Explain, using both words and symbols. Explain the meaning of any symbols you use.

8.104 Tax Regime Suppose you wanted to test the claim that more than half of French citizens support revising the tax regime and structure. Give the null and alternative hypotheses, and explain, using both words and symbols.

8.105 Guessing on an MCQ Test An MCQ test has 75 questions with four options each. Suppose a passing grade is 50 or more correct answers. Test the claim that a student knows more than half of the answers and is not just guessing. Assume the student gets 50 answers correct out of 75. Use a significance level of 0.05. Steps 1 and 2 of a hypothesis test procedure are given. Show steps 3 and 4, and be sure to write a clear conclusion.

Step 1: H_0: $p = 0.25$

H_a: $p > 0.25$

Step 2: Choose the one-proportion z-test. Sample size is large enough, because np_0 is $75(0.25) = 18.75$, and $n(1 - p_0) = 75(0.75) = 56.25$, and both are more than 10. Assume the sample is random and $\alpha = 0.05$.

8.106 Guessing on a Yes/No Test A yes/no test has 75 questions. A passing grade is 50 or more correct answers.

a. What is the probability that a person will guess one yes/no question correctly?

b. Test the hypothesis that a person who got 50 right out of 75 is not just guessing, using an alpha of 0.05. Steps 1 and 2 of a hypothesis testing procedure are given. Finish the question by doing steps 3 and 4.

Step 1: H_0: $p = 0.50$

H_a: $p > 0.50$

Step 2: Choose the one-proportion z-test. n times p is 75 times 0.50, which is 37.5. This is more than 10, and 75 times 0.50 is also more than 10. Assume a random sample.

GUIDED EXERCISES

g **8.35 Gun Control** Historically, the percentage of U.S. residents who support stricter gun control laws has been 52%. A recent Gallup Poll of 1011 people showed 495 in favor of stricter gun control laws. Assume the poll was given to a random sample of people.

QUESTION Test the claim that the proportion of those favoring stricter gun control has changed from 0.52. Perform a hypothesis test, using a significance level of 0.05, by following the steps.

Step 1 ► Hypothesize

H_0: The population proportion that supports gun control is 0.52, $p =$ ______.

H_a: p ______.

Step 2 ► Prepare

$\alpha = 0.05$

Choose the one-proportion z-test.

Random and independent sample: Yes

Sample size: $np_0 = 1011(0.52) =$ about 526, which is more than 10, and $n(1 - p_0) =$ about ____, which is more than ____.

Population size is more than 10 times 1011.

Step 3 ► Compute to compare

$\hat{p} =$ __

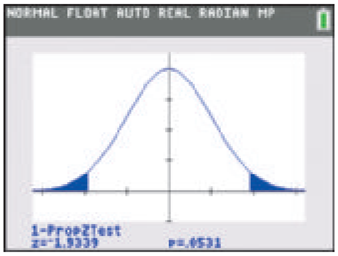
$$SE = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.52(\_\_)}{1011}} = \_\_\_\_$$

$$z = \frac{\hat{p} - p_0}{SE} = \frac{0.4896 - \_\_}{SE} = \_\_$$

p-value = __

Please report your p-value with three decimal digits.

Check your answers with the accompanying figure. Don't worry if the last digits are a bit different (this can occur due to rounding).



TI-84: One-Proportion z-Test

Step 4 ► Interpret

Reject H_0 (if the p-value is 0.05 or less) or do not reject H_0 and choose one of the following conclusions:

i. The percentage is not significantly different from 52%. (A significant difference is one for which the p-value is less than or equal to 0.05.)

ii. The percentage is significantly different from 52%.

g **8.63 Treatment for CLL** Furman et al. (2014) reported on a study of patients with recurring chronic lymphocytic leukemia (CLL). The study was randomized, double-blind, and placebo-controlled. After 12 months, 101 of the 110 patients assigned to the combination treatment of idelalisib (idel) with rituximab (rit) were still alive, and 88 of the 110 patients assigned to placebo with rit were still alive. Perform a hypothesis test to test whether those taking idel have a *better* chance of surviving than those taking the placebo. Use a level of significance of 0.05. Follow the steps below to answer the question. In order to be able to use the typical four steps without changing the numbering, we have called the first step "step 0." Can we conclude that idel *causes* an increased chance of survival?

Step 0 ▶ Find the proportion of those taking idel (and rit) who were still alive, and compare it with the proportion of those taking placebo (and rit) who were still alive. Which group appears to have done better?

Step 1 ▶ Hypothesize

Let p_{idel} be the proportion of those taking idel who are still alive, and let p_{plac} be the proportion of those taking the placebo who are still alive.

H_0: _____

H_a: $p_{idel} > p_{plac}$

Step 2 ▶ Prepare

The significance level is 0.05. Choose the two-proportion z-test. Although we don't have a random sample, we have random assignment to two independent groups. The pooled proportion of survival is

$$\hat{p} = \frac{101 + 88}{110 + 110} = \frac{189}{220} = 0.8591$$

We must check the following four products to make sure none is below 10:

$n_1 \times \hat{p} = 110(0.8591) = 94.5$

$n_1 \times (1 - \hat{p}) = 110(1 - 0.8591) = 110(0.1409) = 15.5$

$n_2 \times \hat{p} =$ _____

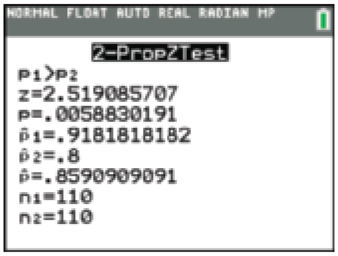
$n_2 \times (1 - \hat{p}) =$ _____

Step 3 ▶ Compute to compare

Refer to the figure.

$z =$ _____

p-value = _____



TI-84 Output for the Two-Proportion z-test

Step 4 ▶ Interpret

Reject or do not reject the null hypothesis, and choose i or ii.

i. Those taking idel did not have a significantly higher rate of survival than those taking the placebo.

ii. Those taking idel did have a significantly higher rate of survival than those taking the placebo.

Causality

Can we conclude that idel causes a better result than the placebo? Why or why not?

TechTips

General Instructions for All Technology

All technologies will use the examples that follow.

EXAMPLE A: ► Do a one-proportion z-test to determine whether you can reject the hypothesis that a coin is a fair coin if 10 heads are obtained from 30 flips of the coin. Find z and the p-value.

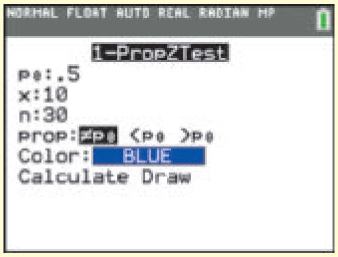
EXAMPLE B: ► Do a two-proportion z-test: Find the observed value of the test statistic and the p-value that tests whether the proportion of people who support stem cell research changed from 2002 to 2007. In both years, the researchers sampled 1500 people. In 2002, 645 people expressed support. In 2007, 765 people expressed support.

TI-84

One-Proportion *z*-Test

1. Press **STAT**, choose **TESTS**, and choose **5: 1-PropZTest**.
2. See Figure 8A.

Enter: p_0, **.5**; **x**, **10**; **n, 30**.
Leave the default $\neq p_0$.
Scroll down to **Calculate** and press **ENTER**.



▲ **FIGURE 8A** TI-84 Input for One-Proportion *z*-Test

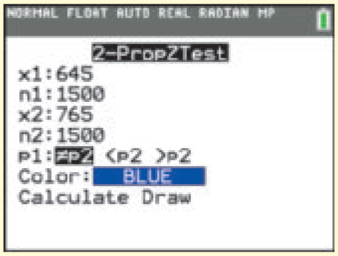
You should get a screen like Figure 8B. If you choose **Draw** instead of **Calculate**, you can see the shading of the Normal curve.

▲ **FIGURE 8B** TI-84 Output for One-Proportion *z*-Test

Two-Proportion *z*-Test

1. Press **STAT**, choose **TESTS**, and choose **6: 2-PropZTest**.
2. See Figure 8C.

Enter: **x1**, **645**; **n1**, **1500**; **x2**, **765**; **n2**, **1500**.
Leave the default **p1** $\neq$ **p2**
Scroll down to **Calculate** (or **Draw**) and press **ENTER**.



▲ **FIGURE 8C** TI-84 Input for Two-Proportion *z*-Test

You should get a screen like Figure 8D.

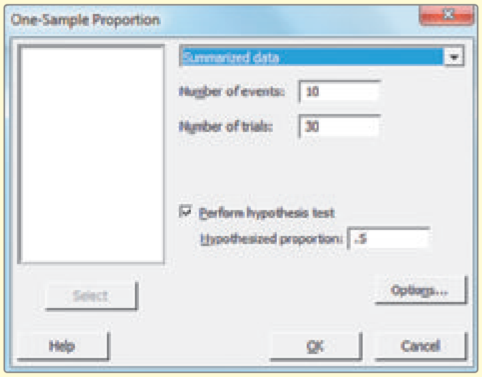
▲ **FIGURE 8D** TI-84 Output for Two-Proportion *z*-Test

Caution! Beware of p-values that appear at first glance to be larger than 1. In Figure 8D, the p-value is 1.1 times 10 to the negative fifth power (−5), or 0.000011.

MINITAB

One-Proportion *z*-Test

1. **Stat** > **Basic Statistics** > **1-Proportion**.
2. Refer to Figure 8E. Select **Summarized data.** Enter: **number of events**, **10**; **Number of trials**, **30**; check **Perform hypothesis test;** and enter **hypothesized proportion**, **.5**
3. Click **Options** and for **Method,** select **Normal approximation**. (If you wanted to change the alternative hypothesis to one-sided, you would do that also through **Options**.) Click **OK**: click **OK**.



▲ **FIGURE 8E** Minitab Input for One-Proportion *z*-Test

You should get output that looks like Figure 8F. Note that you also get a 95% confidence interval (95% CI) for the proportion.

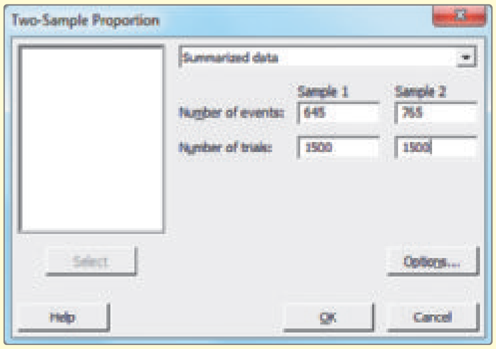
```
Test of p = 0.5 vs p ≠ 0.5

Sample   X   N  Sample p         95% CI         Z-Value  P-Value
1       10  30  0.333333  (0.164646, 0.502020)    -1.83    0.068

Using the normal approximation.
```

▲ **FIGURE 8F** Minitab Output for One-Proportion *z*-Test

Two-Proportion *z*-Test

1. **Stat** > **Basic Statistics** > **2 Proportions**
2. See Figure 8G.

   Select **Summarized data**. Enter **Sample 1 Number of events: 645**, **Number of trials: 1500; Sample 2 Number of events: 765**, **Number of trials: 1500**.
3. Click **OK**.



▲ **FIGURE 8G** Minitab Input for Two-Proportion *z*-test

Your output should look like Figure 8H. Note that it includes a 95% confidence interval for the difference between the proportions, as well as *z* and the p-value.

```
Test and CI for Two Proportions

Sample    X     N  Sample p
1       645  1500  0.430000
2       765  1500  0.510000


Difference = p (1) - p (2)
Estimate for difference:  -0.08
95% CI for difference:  (-0.115605, -0.0443955)
Test for difference = 0 (vs ≠ 0):  Z = -4.40 P-Value = 0.000

Fisher's exact test: P-Value = 0.000
```

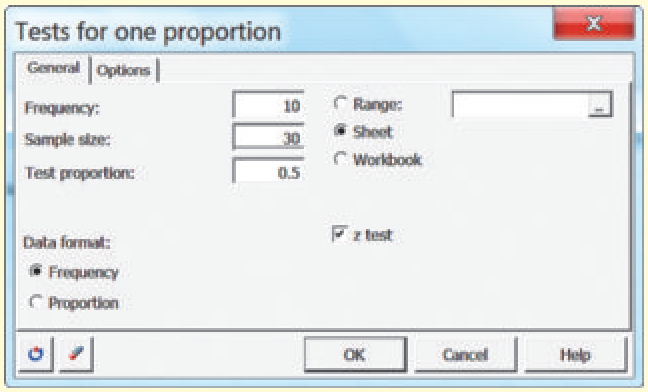
▲ **FIGURE 8H** Minitab Output for Two-Proportion *z*-Test and Interval

EXCEL

One-Proportion *z*-Test

1. Click **Add-ins**, **XLSTAT**, **Parametric tests**, **Tests for one proportion**.
2. See Figure 8I.

Enter: **Frequency**, **10**; **Sample size, 30**; **Test proportion**, **.5**.
(If you wanted a one-sided hypothesis, you would click **Options**.)
Click **OK**.



▲ **FIGURE 8I** XLSTAT input for One Proportion *z*-Test

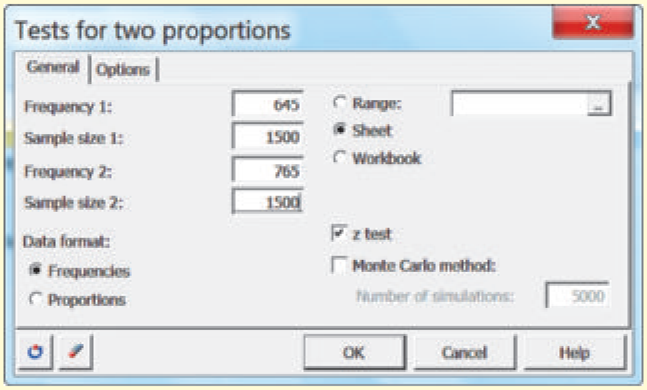
When the output appears, you may need to change the column width to see the answers. Click **Home**, and in the **Cells** group click **Format** and **AutoFit Column Width**. The relevant parts of the output are shown here.

Difference	−0.167
z (Observed value)	−1.826
p-value (Two-tailed)	0.068

Two-Proportion *z*-Test

1. Click **Add-ins**, **XLSTAT**, **Parametric tests**, **Tests for two proportions**.
2. See Figure 8J.

Enter: **Frequency 1**, **645**; **Sample size 1**, **1500**; **Frequency 2**, **765**; **Sample size 2**, **1500**.
(If you wanted a one-sided alternative, you would click **Options**.)
Click **OK**.



▲ **FIGURE 8J** XLSTAT Input for Two-Proportion *z*-Test

The relevant parts of the output are shown.

Difference	−0.080
z (Observed value)	−4.404
p-value (Two-tailed)	<0.0001

STATCRUNCH

One-Proportion *z*-Test and Confidence Interval

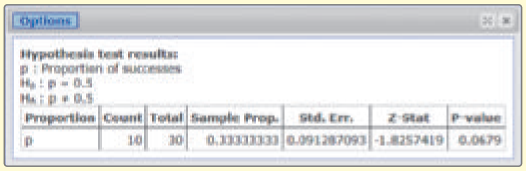
1. **Stat** > **Proportion Stats** > **One Sample** > **With Summary**
2. Enter: **# of successes**, **10**; **# of observations**, **30**.

   Select the **Hypothesis Test** or **Confidence interval** option.

   a. Leave Hypothesis test checked. Enter: **Ho: p**=, **0.5**. Leave the **Alternative 2-tailed**, which is the default.

   b. For a Confidence interval, leave the **Level**, **0.95** (the default). For **Method**, leave ***Standard-Wald***, the default.
3. Click **Compute!**

   Figure 8K shows the output for the hypothesis test.



Options

Hypothesis test results:
p : Proportion of successes
H_0 : p = 0.5
H_A : p ≠ 0.5

Proportion	Count	Total	Sample Prop.	Std. Err.	Z-Stat	P-value
p	10	30	0.33333333	0.091287093	-1.8257419	0.0679

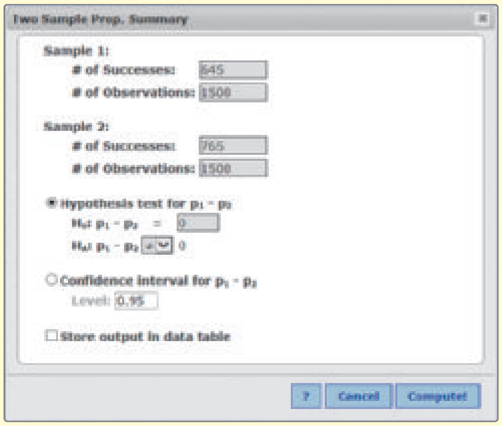
▲ **FIGURE 8K** StatCrunch Output for One-Proportion *z*-Test

Two-Proportion *z*-Test

1. **Stat** > **Proportion Stats** > **Two Sample** > **With Summary**
2. Refer to Figure 8L.

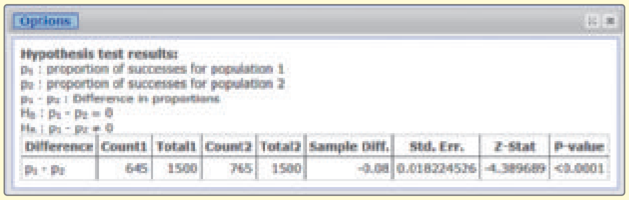
Enter: **Sample 1: # of successes**, **645**;
# of observations, **1500**:
Sample 2: # of successes, **765**;
# of observations, **1500**.

3. You may want to change the alternative hypothesis from the two-sided default (or want a confidence interval). Otherwise, click **Compute!**



▲ **FIGURE 8L** StatCrunch Input for Two-Proportion *z*-Test

Figure 8M shows the StatCrunch output.



Options

Hypothesis test results:
p_1 : proportion of successes for population 1
p_2 : proportion of successes for population 2
$p_1 - p_2$: Difference in proportions
H_0 : $p_1 - p_2 = 0$
H_A : $p_1 - p_2 \neq 0$

Difference	Count1	Total1	Count2	Total2	Sample Diff.	Std. Err.	Z-Stat	P-value
$p_1 - p_2$	645	1500	765	1500	-0.08	0.018224526	-4.389689	<0.0001

▲ **FIGURE 8M** StatCrunch Output for Two-Proportion *z*-Test

9 Inferring Population Means

THEME

The sample mean can be used to estimate the population mean. To understand how to form confidence intervals and carry out hypothesis tests, we need to understand the precision and accuracy of the sample mean as an estimate, and we need to know its probability distribution. To compare two means, we calculate the difference of the sample means, and we also need to know the accuracy, precision, and probability distribution for the difference of two samples means.

Brewing beer is a tricky business. Beer has only four main ingredients: malted barley, hops, yeast, and water. However, these four ingredients must be mixed in precise quantities at precise temperatures. Having some experience in trying to get this mix just right, in the late 1800s the Guinness brewery in Dublin, Ireland, began to hire the best and brightest science graduates to help them perfect the brewing process. One of these, hired in 1899 at the age of 23, was William Sealy Gosset (1876–1937), who had majored in chemistry and mathematics. One of Gosset's jobs was to measure the amount of yeast in a small sample of beer. On the basis of this sample, he was to estimate the mean amount of yeast in the beers produced. If this yeast amount was too high or too low, then something was wrong and the process would have to be fixed.

Naturally, uncertainty played a role. Suppose the average yeast count in his sample was too high or too low. Did this indicate that the mean yeast count in the entire factory was off? Or was his sample different just because of chance? The statistical science of the day knew how to answer this question if the number of samples was large, but Gosset worked in a context in which large samples were just too expensive and time-consuming to collect. A decision had to be based on a small sample. Gosset solved the problem, and his approach (which is called the *t*-test) is now one of the most widely used techniques in statistics.

Estimating one or more population means is still an important part of science and public policy. How do we compare the effects of different drugs on epilepsy? How do commuting times vary between cities? Does our sense of smell differ when we are sitting upright compared to lying down? To answer these questions, we need good estimates of the population means, and these estimates must be based on reliable data from small samples.

In Chapters 7 and 8 you learned two important techniques for statistical inference: the confidence interval and the hypothesis test. In those chapters, we applied statistical inference for population proportions. In this chapter, we use the same two techniques for making inferences about means of populations. We begin with inference for one population and conclude with inferring the difference between the means of two populations.

CASE STUDY

Epilepsy Drugs and Children

Epilepsy is a neurological disorder that causes seizures, which can range from mild to severe. It has been estimated (wikipedia.org) that over 50 million people around the world suffer from epilepsy. It is usually treated with drugs, and four drugs that are commonly used for this purpose are carbamazepine, lamotrigine, phenytoin, and valproate. In 2009 researchers at the *New England Journal of Medicine* reported that pregnant mothers who take valproate might risk impairing the cognitive development of their children, compared to mothers who take one of the other drugs. As evidence, on the basis of a sample of pregnant women with epilepsy, they estimated the mean IQ of three-year-old children whose mothers took one of these four drugs during their pregnancies. They gave 95% confidence intervals (CI) for the mean IQ, as shown in Table 9.1 on the next page (Meador et al. 2009).

Drug	95% CI
Carbamazepine	(95, 101)
Lamotrigine	(98, 104)
Phenytoin	(94, 104)
Valproate	(88, 97)

▲ **TABLE 9.1** 95% confidence intervals for the average IQ of children whose mothers took various epilepsy drugs during their pregnancies.

Why did these four intervals lead the researchers to recommend that pregnant women not use valproate as a "first choice" drug for epilepsy? The researchers wrote that "Although the confidence intervals for carbamazepine and phenytoin overlap with the confidence interval for valproate, the confidence intervals for the differences between carbamazepine and valproate and between phenytoin and valproate do not include zero." What does this tell us?

In this chapter we discuss how confidence intervals can be used to estimate characteristics of a population—in this case, the population of all children of women with epilepsy who took one of these drugs during pregnancy. Confidence intervals can also be used to judge between hypotheses about the means and about differences between means. The population of pregnant women with epilepsy is large, and yet if conditions are right, we can make decisions and reach an understanding about the entire population on the basis of a small sample. At the end of this chapter, we will return to this study and see if we can better understand its conclusions.

SECTION 9.1

Sample Means of Random Samples

As you learned in Chapter 7, we estimate population parameters by collecting a random sample from that population. We use the collected data to calculate a statistic, and this statistic is used to estimate the parameter. Whether we are using the statistic $\hat{p}$ to estimate the parameter p or are using $\bar{x}$ to estimate μ, if we want to know how close our estimate is to the truth, we need to know how far away that statistic is, typically, from the parameter.

Just as we did in Chapter 7 with $\hat{p}$, we now examine three characteristics of the behavior of the sample mean: its accuracy, its precision, and its probability distribution. By understanding these characteristics, we'll be able to measure how well our estimate performs and thus make better decisions.

As a reminder, Table 9.2 shows some commonly used statistics and the parameters they estimate. (This table originally appeared as Table 7.2.)

▶ **TABLE 9.2**

Statistic (based on data)		Parameter (typically unknown)	
Sample mean	$\bar{x}$	Population mean	μ (mu)
Sample standard deviation	s	Population standard deviation	σ (sigma)
Sample variance	s^2	Population variance	σ^2
Sample proportion	$\hat{p}$ (p-hat)	Population proportion	p

Details

Mu-sings
The mean of a population, represented by the Greek character μ, is pronounced "mu" as in *music*.

This chapter uses much of the vocabulary introduced in Chapters 7 and 8, but we'll remind you of important terms as we proceed. To help you visualize how a sample mean based on randomly sampled data behaves, we'll make use of the by-now-familiar technique of simulation. Our simulation is slightly artificial, because to do a simulation, we need to know the population. However, after using a simulation to understand how the sample mean behaves in this artificial situation, we will discuss what we do in the real world when we do not know very much about the population.

Accuracy and Precision of a Sample Mean

The reason why the sample mean is a useful estimator for the population mean is that the sample mean is accurate and, with a sufficiently large sample size, very precise. The accuracy of an estimator, you'll recall, is measured by the **bias**, and the precision is measured by the standard error. You will see in this simulation that

1. The sample mean is unbiased when estimating the population mean—that is, on average, the sample mean is the same as the population mean.
2. The **precision** of the sample mean depends on the variability in the population, but the more observations we collect, the more precise the sample mean becomes.

Looking Back

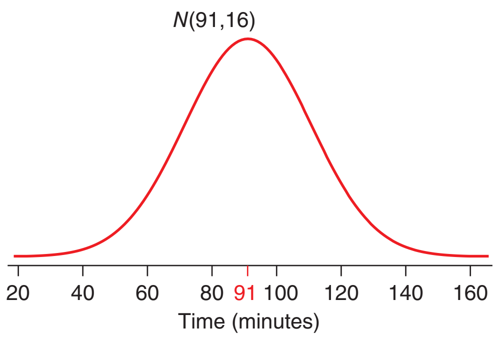
Bias and Precision
Bias is the mean distance between the sample statistic and the parameter it is estimating. Precision is measured by the standard error, which tells us how far the statistic typically deviates from its center. Review Figure 7.2 (golf balls on a putting green) to help visualize these two properties.

For our simulation, we'll use the population that consists of the finishing times of all men who ran the Cherry Blossom Ten Mile Run in 2013 (Kaplan 2009, 2014). (The finishing time is the amount of time it took to finish the race.) This race is held every spring in Washington, D.C. As in most such races, data are carefully collected on every participant. Rather than showing you the histogram of the finishing times for all 7128 runners, we're going to take advantage of the fact that the distribution closely follows a Normal model with a mean of 91 minutes and a standard deviation of 16 minutes: $N(91, 16)$. Figure 9.1 shows the distribution of this population.

The population parameters are, in symbols,

$$\mu = 91 \text{ minutes}$$
$$\sigma = 16 \text{ minutes}$$

For our simulation, we will randomly sample 30 runners and calculate the average of their finishing times. We'll then repeat this many, many times. (The exact number of repetitions we performed isn't important for this discussion.) We are interested in two questions: (1) What is the typical value of the sample mean? If it is 91 minutes—the value of the population mean—then the sample mean is unbiased. (2) Typically, how far away is a sample mean from 91? In other words, how much spread is there in the distribution of sample means? This spread helps us measure the precision of the sample mean as an estimator of the population mean.

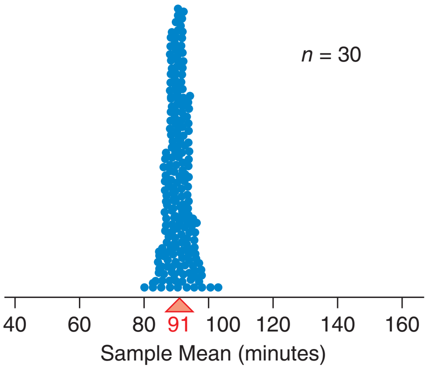


◀ **FIGURE 9.1** A Normal model for finishing times in the Cherry Blossom Ten Mile Run, with a mean of 91 minutes and a standard deviation of 16 minutes.

For example, our very first sample of 30 runners had a mean finishing time of 95.1 minutes. We plotted this sample mean, as well as the means of the many other samples we took, in Figure 9.2 on the next page. The plot is on the same scale as Figure 9.1, the picture of the population distribution, so that you can see how much narrower this distribution of sample means is.

From this dotplot of sample means, we learn that the typical value of the sample means is the same as the population mean of 91 minutes. And we see that the sample mean is a relatively precise estimator: All of the sample means are within about 10 minutes (either above or below) the true mean value of 91 minutes.

▶ **FIGURE 9.2** Each dot represents a sample mean based on 30 runners who were randomly selected from the population whose distribution is shown in Figure 9.1. Note that the spread of this distribution is much smaller than the spread of the population, but the center looks to be at about the same place: 91 minutes.



Looking Back

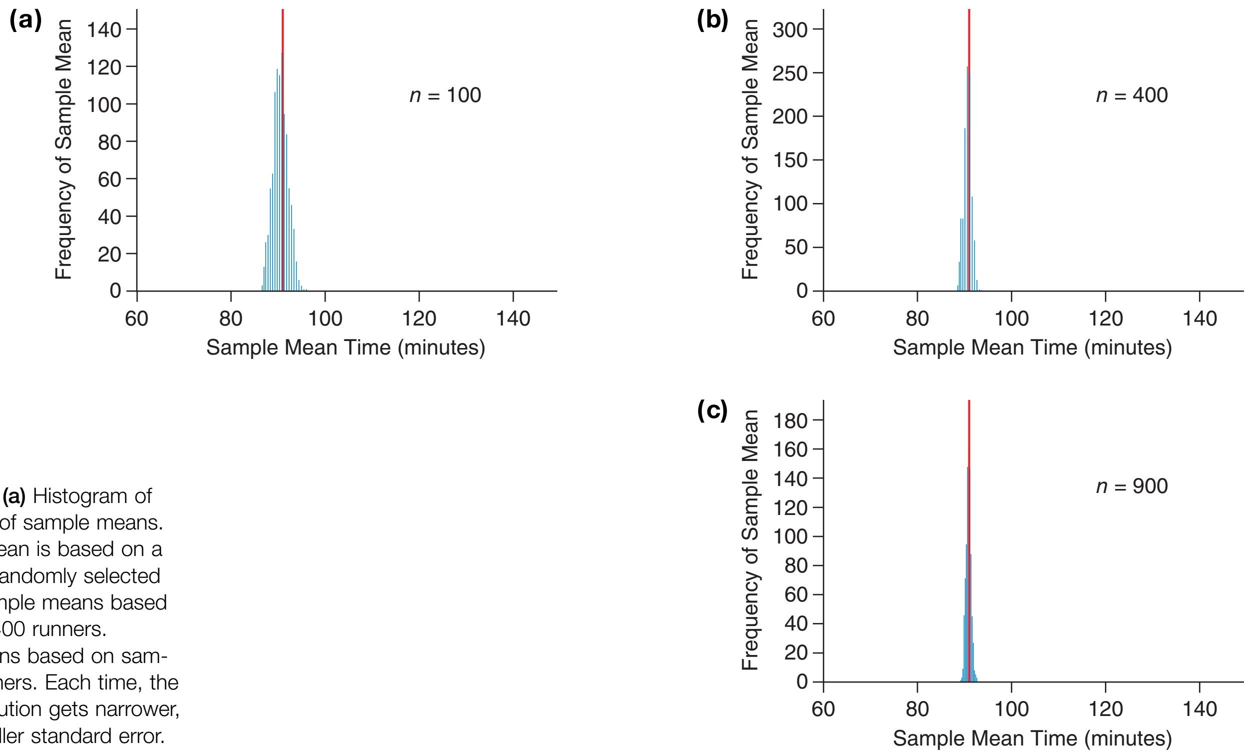
Sampling Distribution
In Chapter 7 we introduced the sampling distribution of sample proportions. The sampling distribution of sample means is the same concept: It is a distribution that gives us probabilities for sample means drawn from a population. The sampling distribution is the distribution of all possible sample means.

Figure 9.2 is a very approximate picture of the **sampling distribution** of the sample mean for samples of size 30. Recall that a sampling distribution is a probability distribution of a statistic; in this case, the statistic is the sample mean. You can think of the sampling distribution as the distribution of *all* possible sample means that would result from drawing repeated random samples of a certain size from the population.

When the mean of the sampling distribution is the same value as the population mean, we say that the statistic is an **unbiased estimator**. This appears to be the case here, because both the mean of the distribution of sample means in Figure 9.2 and the population mean are about 91 minutes.

The standard deviation of the sampling distribution is what we call the **standard error**. The standard error measures the precision of an estimator by telling us how much the statistic varies from sample to sample. For the sample mean, the standard error is smaller than the population standard deviation. We can see this because the spread for the sampling distribution is smaller than the spread of the population distribution. Soon you'll see how to calculate the standard error.

What happens to the center and spread of the sampling distribution if we increase the sample size? Let's start all over with the simulation. But this time, we take a random sample of 100 runners and calculate the mean. We then repeat this many hundreds of times. Figure 9.3 shows the results for this simulation and also for two new simulations



▶ **FIGURE 9.3** **(a)** Histogram of a large number of sample means. Each sample mean is based on a sample of 100 randomly selected runners. **(b)** Sample means based on samples of 400 runners. **(c)** Sample means based on samples of 900 runners. Each time, the sampling distribution gets narrower, reflecting a smaller standard error.

where each sample mean is based on 400 runners and then on 900 runners. The scale of the *x*-axis is the same as in Figure 9.1. Note that the spread of the distributions becomes quite small—so small, in fact, that we can't get a good look at the shape of the distributions.

What Have We Demonstrated with These Simulations?

Because the sampling distributions are always centered at the population mean, we have demonstrated that the sample mean is an unbiased estimator of the population mean. We saw this for only one type of population distribution: the Normal distribution. But in fact, this is the case for any population distribution.

We have demonstrated that the standard deviation of the sampling distribution, which is called the **standard error** of the sample mean, is smaller when based on a larger sample size. This is true for any population distribution.

We can be more precise. If the symbol μ represents the mean of the population and if σ represents the standard deviation of the population, then

1. The mean of the sampling distribution is also μ (which tells us that the sample mean is unbiased when estimating the population mean).
2. The standard error is $\frac{\sigma}{\sqrt{n}}$ (which tells us that the standard error depends on the population distribution and is smaller for larger samples).

Looking Back

Sample Proportions from Random Samples
Compare the properties of the sample mean to those of the sample proportion, $\hat{p}$, given in Chapter 7. The sample proportion is also an unbiased estimator (for estimating the population proportion, p). It has standard error $\sqrt{\frac{p(1-p)}{n}}$.

KEY POINT For all populations, the sample mean, if based on a random sample, is unbiased when estimating the population mean. The standard error of the sample mean is $\frac{\sigma}{\sqrt{n}}$, so the sample mean is more precise for larger sample sizes.

EXAMPLE 1 iTunes Music Library Statistics

A student's iTunes music library has a very large number of songs. The mean length of the songs is 243 seconds, and the standard deviation is 93 seconds. The distribution of song lengths is right-skewed. Using his digital music player, this student will create a playlist that consists of 25 randomly selected songs.

QUESTIONS

a. Is the mean value of 243 seconds an example of a parameter or a statistic? Explain.

b. What should the student expect the average song length to be for his playlist?

c. What is the standard error for the mean song length of 25 randomly selected songs?

SOLUTIONS

a. The mean of 243 is an example of a parameter, because it is the mean of the population that consists of all of the songs in the student's library.

b. The sample mean length can vary, but it is typically the same as the population mean: 243 seconds.

c. The standard error is $\frac{\sigma}{\sqrt{n}} = \frac{93}{\sqrt{25}} = \frac{93}{5} = 18.6$ seconds.

TRY THIS! Exercise 9.9

SECTION 9.2

The Central Limit Theorem for Sample Means

In the last simulation, all of the approximate sampling distributions (Figures 9.2 and 9.3) looked very close to Normal. This probably doesn't surprise you, because the population distribution was Normal.

What might surprise you is that the sampling distribution of the mean is always Normal (or at least approximately Normal), regardless of the shape of the population distribution. (If the sample size is small, however, the approximation can be pretty lousy.) This is the conclusion of the Central Limit Theorem, an important mathematical theorem that tells us that as long as the sample size is large, we can use the Normal distribution to perform statistical inference, regardless of the population the data are sampled from.

Looking Back

CLT for Proportions
In Chapter 7, you saw that the Central Limit Theorem applies to sample proportions. Here you'll see that it also applies to sample means.

The **Central Limit Theorem (CLT)** assures us that no matter what the shape of the population distribution, if a sample is selected such that the following conditions are met, then the distribution of sample means follows an approximately Normal distribution. The mean of this distribution is the same as the population mean. The standard deviation (also called the standard error) of this distribution is the population standard deviation divided by the square root of the sample size. As a rule of thumb, sample sizes of 25 or more may be considered "large."

When determining whether you can apply the Central Limit Theorem to analyze data, there are three conditions to consider:

Condition 1: *Random Sample and Independence.* Each observation is collected randomly from the population, and observations are independent of each other. The sample can be collected either with or without replacement.

Condition 2: *Large Sample*. Either the population distribution is Normal or the sample size is large.

Condition 3: *Big Population*. If the sample is collected without replacement (as is done in a SRS), then the population must be at least 10 times larger than the sample size.

KEY POINT

The sampling distribution of $\bar{x}$ is approximately $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, where μ is the mean of the population and σ is the standard deviation of the population. The larger the sample size, n, the better the approximation. If the population is Normal to begin with, then the sampling distribution is exactly a Normal distribution, regardless of the sample size.

EXAMPLE 2 The Sample Mean as Estimator

Why is the sample mean of several runners a better estimator of the "typical" race time than is a single observation? Consider our Cherry Blossom race, and suppose that we did not have access to the entire population, but only to a random sample. (In fact, many websites for similar races allow us to see only a few observations at a time.) First, we will select a single runner at random. How close do you think his finishing time is likely to be to 91 minutes, the mean finishing time for all runners? In other words, is a randomly selected runner likely to be "typical"? Next, we will select 9 runners at random and calculate their average finishing times. How close do you think this sample average is likely to be to 91 minutes? To answer, we will use the facts that the population distribution of finishing times is close to Normal, that the population mean is 91, and that the population standard deviation is 16 minutes.

QUESTIONS

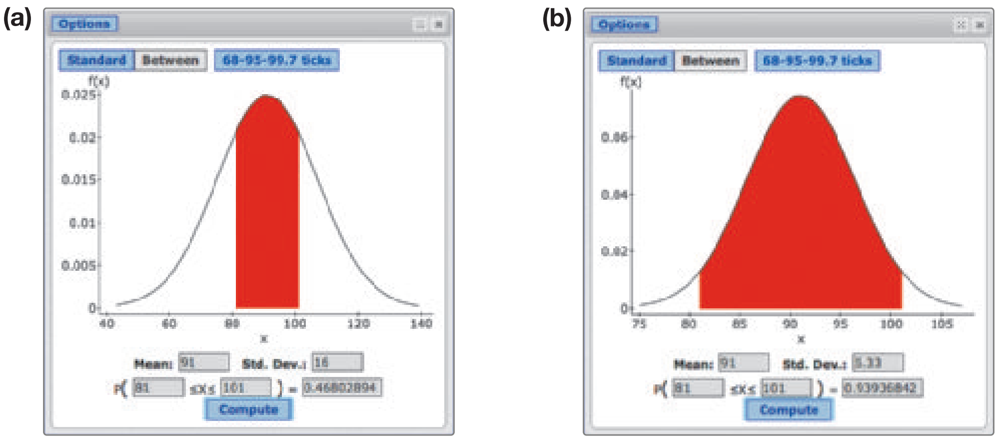
a. What is the probability that the finishing time of a single randomly selected runner is within 10 minutes of the population mean?

b. What is the probability that the sample mean finishing time of 9 runners is within 10 minutes of the population mean?

c. Compare the two probabilities. What does this tell us about the benefits of a large sample size for estimating the population mean? (Note that because the population distribution is Normal, we need only check that condition 1, Random and Independent Sample, is met to carry out our calculations.)

SOLUTIONS

a. We can use a probability calculator to find the probability that a single runner's finishing time will be within 10 minutes of the mean. This is the same as finding the probability that this time will be between $91 - 10 = 81$ minutes and $91 + 10 = 101$ minutes. Figure 9.4a illustrates StatCrunch output that shows this probability to be 0.47.



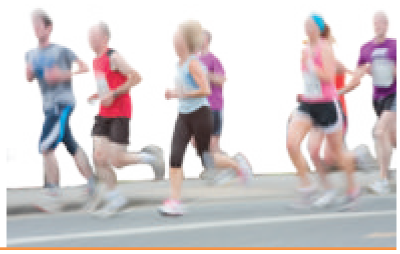
◀ **FIGURE 9.4** StatCrunch calculator showing the probability that **(a)** a single runner's time will be within 10 minutes of the mean and **(b)** the average of 9 runners' times will be within 10 minutes of the mean.

b. With a sample size of $n = 9$, the standard error of $\bar{x}$ is $\dfrac{\sigma}{\sqrt{n}} = \dfrac{16}{\sqrt{9}} = 5.33$ minutes.

Because the mean of the sampling distribution of x is the same as the population mean of 91 minutes, we use the $N(91, 5.33)$ distribution to find the probability that $\bar{x}$ is between 81 and 101 minutes.

This calculation is shown in Figure 9.4b. The probability is about 0.94.

c. The probability that we are close to the population mean is much greater with a sample size of 9 than with a sample size of 1, because the standard error is smaller. The larger the sample size, the smaller the standard error. This is why larger sample sizes are better for estimating the population mean; the sample mean is more likely to be "close" to the population mean for larger samples.



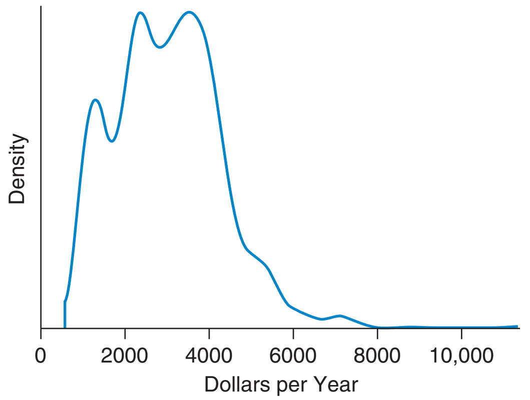
TRY THIS! Exercise 9.11

Example 2 used the fact that the population distribution was Normal. The beauty of the Central Limit Theorem, though, is that as long as the three conditions are met, the shape of the population distribution doesn't matter.

Visualizing Distributions of Sample Means

The sketch in Figure 9.5 shows the distribution of in-state tuition and fees for all two-year colleges in the United States for the 2012–2013 academic year (Integrated Postsecondary Education Data System, U.S. Dept. of Education). Note that the distribution looks nothing at all like a Normal distribution. It is skewed and multimodal. (The mode around \$1000 is due largely to the cost of two-year colleges in California.)

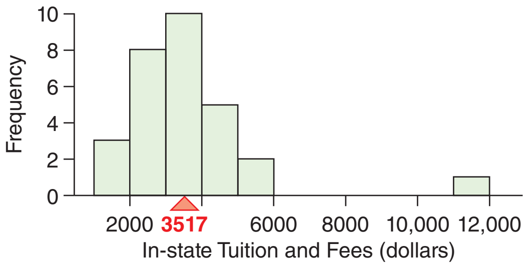
▶ **FIGURE 9.5** Distribution of annual tuitions and fees at all two-year colleges in the United States for the 2012–2013 academic year.



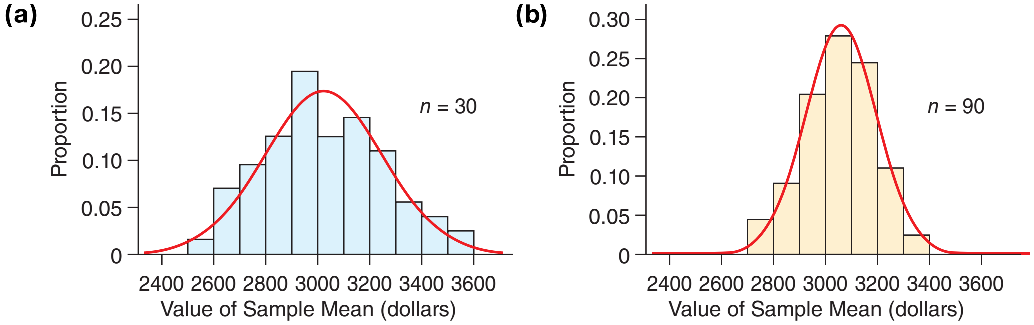
This distribution represents the distribution of a population, because it includes *all* two-year colleges. The mean of this population—the "typical" tuition of all two-year colleges—is \$3030.

Using this distribution, we now show the results of a simulation that should be starting to feel familiar. First, we take a random sample of 30 colleges. The distribution of this sample is shown in Figure 9.6. We find the mean tuition of the 30 colleges in the sample and record this figure; for example, the sample mean for the sample shown in Figure 9.6 is about \$3517.

▶ **FIGURE 9.6** Distribution of a sample of 30 colleges taken from the population of all colleges. The mean of this sample, \$3517, is indicated.



We repeat this activity (that is, we sample another 30 colleges from the population of all colleges and record the mean tuition of the sample) 200 times. When we are finished, we have 200 sample mean tuitions, each sample mean based on a sample of 30 colleges. Figure 9.7a shows this distribution. Figure 9.7b shows the distribution of averages when, instead of sampling 30 colleges, we triple the number and sample 90 colleges. What differences do you see among the population distribution (Figure 9.5), the distribution of one sample (Figure 9.6), the sampling distribution when the sample size is 30 (Figure 9.7a), and the sampling distribution when the sample size is 90 (Figure 9.7b)?



◀ **FIGURE 9.7** **(a)** Distribution of sample means, where each sample mean is based on a sample size of $n = 30$ college tuitions and is drawn from the population shown in Figure 9.5. This is (approximately) the sampling distribution of $\bar{x}$ when $n = 30$. A Normal curve is superimposed. **(b)** The approximate sampling distribution of $\bar{x}$ when $n = 90$.

Both of the sampling distributions in Figures 9.7a and 9.7b show us the values and relative frequencies for $\bar{x}$, but they are based on different sample sizes. We see that even though the *population* distribution has an unusual shape (Figure 9.5), the sampling distributions for $\bar{x}$ are fairly symmetric and unimodal. Although the Normal curve that is superimposed doesn't match the histogram very closely when $n = 30$, the match is pretty good for $n = 90$.

This is exactly what the CLT predicts. When the sample size is large enough, we can use the Normal distribution to find approximate probabilities for the values we might see for $\bar{x}$ when we take a random sample from the population.

The more observations in your sample, the better an approximation the Normal distribution provides. Generally, the CLT provides a useful approximation of the true probabilities if the sample size is 25 or more. But this is just a rule of thumb. Be aware that you might need larger sample sizes in some situations. Unlike in Chapter 7, where we worked with sample proportions, we can't provide a hard-and-fast rule for sample size. For nearly all examples in this text, though not always in real life, 25 is large enough.

Looking Back

Distribution of a Sample vs. Sampling Distribution
Remember that these are two different concepts. The *distribution of a sample*, from Chapter 3, is the distribution of one single sample of data (Figure 9.6). The *sampling distribution*, on the other hand, is the probability distribution of an estimator or statistic such as the sample mean (Figures 9.7a and 9.7b).

Applying the Central Limit Theorem

The Central Limit Theorem helps us find probabilities for sample means when those means are based on a random sample from a population. Example 3 demonstrates how we can answer probability questions about the sample mean even if we can't answer probability questions about individual outcomes.

EXAMPLE 3 Pulse Rates Are Not Normal

According to one very large study done in the United States, the mean resting pulse rate of adult women is about 74 beats per minute (bpm), and the standard deviation of this population is 13 bpm (NHANES). The distribution of resting pulse rates is known to be skewed right.

QUESTIONS

a. Suppose we take a random sample of 36 women from this population. What is the approximate probability that the average pulse rate of this sample will be below 71 or above 77 bpm? (In other words, what is the probability that it will be more than 3 bpm away from the population mean of 74 bpm?)

b. Can you find the probability that a single adult woman will have a resting pulse rate more than 3 bpm away from the mean value of 74?

SOLUTION

a. It doesn't matter that the population distribution is not Normal. Because the sample size of 36 women is relatively large, the distribution of sample means will be approximately (though not exactly) Normal.

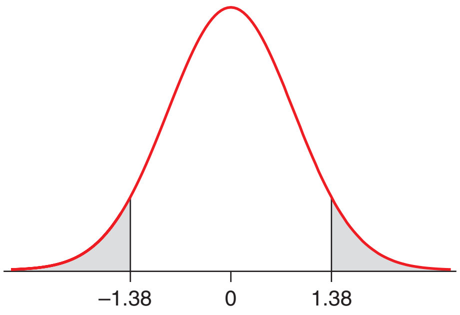
The mean of this Normal distribution will be the same as the population mean: $\mu = 74$ bpm. The standard deviation of this distribution is the standard error:

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{13}{\sqrt{36}} = \frac{13}{6} = 2.167$$

To use the Normal table to find probabilities requires that the values of 71 bpm and 77 bpm be converted to standard units:

$$z = \frac{\bar{x} - \mu}{SE} = \frac{71 - 74}{2.167} = \frac{-3}{2.167} = -1.38$$

Figure 9.8 shows that the area that corresponds to the probability that the sample mean pulse rate will be more than 1.38 standard errors away from the population mean pulse rate. This probability is calculated to be about 17%.



▶ **FIGURE 9.8** Area of the Normal curve outside of z-scores of -1.38 and 1.38.

Caution

CLT Not Universal
The CLT does not apply to all statistics you run across. It does not apply to the sample median, for example. No matter how large the sample size, you cannot use the Normal distribution to find a probability for the median value. It also does not apply to the sample standard deviation.

CONCLUSIONS

a. The approximate probability that the average pulse of 36 adult women will be more than 3 bpm away from 74 bpm is about 17%.

b. We cannot find the probability for a single woman because we do not know the probability distribution. We know only that it is "right-skewed," which is not enough information to find actual probabilities.

TRY THIS! Exercise 9.13

Many Distributions

It's natural at this point to feel that you have seen a confusingly large number of types of distributions, but it's important that you keep them straight. The *population distribution* is the distribution of values from the population. Figure 9.5 (two-year college tuitions) is an example of a population distribution because it shows the distribution of *all* two-year colleges. Figure 9.1 (runners' times for all competitors in a race) is another example of a population distribution. For some populations, we don't know precisely what this distribution is. Sometimes we assume (or know) it is Normal, sometimes we know it is skewed in one direction or the other, and sometimes we know almost nothing.

We then take a random sample of n observations. We can make a histogram of these data. This histogram gives us a picture of the *distribution of the sample*. If the sample size is large, and if the sample is random, then the sample will be representative of the population, and the distribution of the sample will look similar to (but not the same as!) the population distribution. Figure 9.6 is an example of the distribution of a sample of size $n = 30$ taken from the population of two-year college tuitions.

The *sampling distribution* is more abstract. If we take a random sample of data and find the sample mean (the center of the distribution of the sample), and then repeat this many, many times, we will get an idea what the sampling distribution looks like. Figures 9.7a and 9.7b are examples of approximate sampling distributions for the sample mean, based on samples from the two-year college tuition data. Note that these do not share the same shape as the population or the sample; they are both approximately Normal.

EXAMPLE 4 Identify the Distribution

Figure 9.9 shows three distributions. One distribution is a population. The other two distributions are (approximate) sampling distributions of sample means randomly sampled from that population. One sampling distribution is based on sample means of size 10, and the other is based on sample means of size 25.

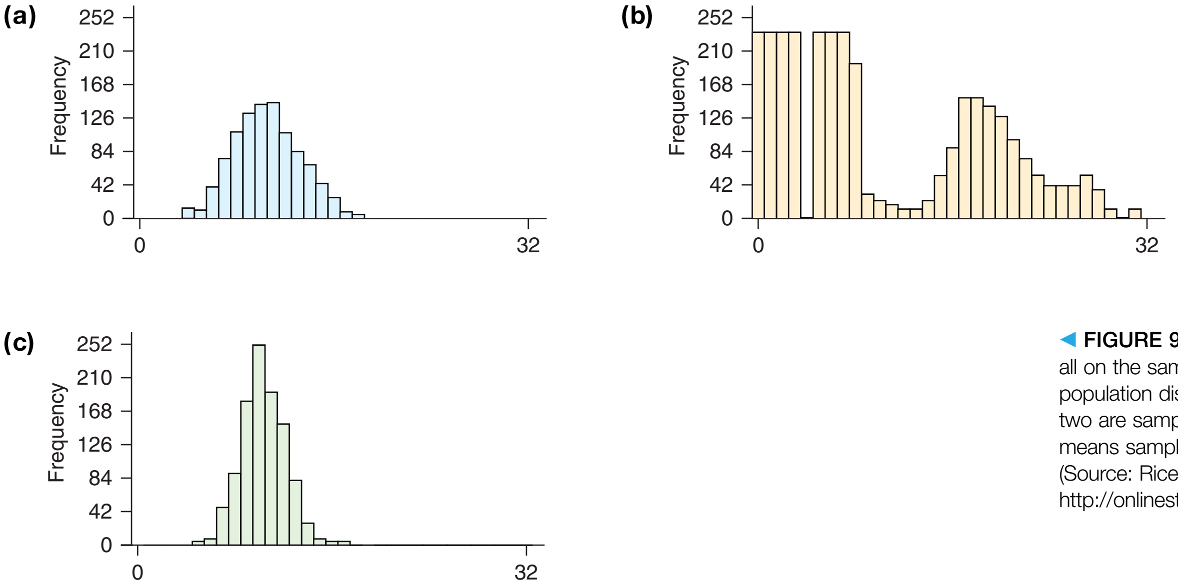


FIGURE 9.9 Three distributions, all on the same scale. One is a population distribution, and the other two are sampling distributions for means sampled from the population. (Source: Rice Virtual Lab in Statistics, http://onlinestatbook.com/)

QUESTION Which graph (a, b, or c) is the population distribution? Which shows the sampling distribution for the mean with $n = 10$? Which with $n = 25$?

SOLUTION The Central Limit Theorem tells us that sampling distributions for means are approximately Normal. This implies that Figure 9.9b is not a sampling distribution, so it must be the population distribution from which the samples were taken. We know that the sample mean is more precise for larger samples, and because Figure 9.9a has the larger standard error (is wider), it must be the graph associated with $n = 10$. This means that Figure 9.9c is the sampling distribution of means with $n = 25$.

TRY THIS! Exercise 9.15

SNAPSHOT THE SAMPLE MEAN ($\bar{x}$)

WHAT IS IT? ► The arithmetic average of a sample of data.

WHAT DOES IT DO? ► Estimates the mean value of a population, μ. The mean is used as a measure of what is "typical" for a population.

HOW DOES IT DO IT? ► If the sample was a random sample, then the sample mean is unbiased, and we can make the precision of the estimator as good as we want by taking a large enough sample size.

HOW IS IT USED? ► If the sample size is large enough (or the population is Normal), we can use the Normal distribution to find the probability that the sample mean will take on a value in any given range. This lets us know how wrong our estimate could be.

The *t*-Distribution

The hypothesis tests and confidence intervals that we will use for estimating and testing the mean are based on a statistic called the ***t*-statistic**:

$$t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)}$$

Looking Back

Sample Standard Deviation
In Chapter 3 we gave the formula for the sample standard deviation:

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}}$$

The *t*-statistic is very similar to a *z*-score for the sample mean. In the numerator, we subtract the population mean from the sample mean. Then we divide not by the standard error but, rather, by an *estimate* of the standard error.

It would be nice if we could divide by the true standard error. But in real life, we almost never know the value of σ, the population standard deviation. So instead, we replace it with an estimate: the sample standard deviation, s. This gives us an estimate of the standard error:

$$SE_{\text{EST}} = \frac{s}{\sqrt{n}}$$

Compare the *t*-statistic to the *z*-statistic, and you will see that we simply replaced σ in the *z*-statistic with s.

$$z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$$

Details

Degrees of Freedom
Degrees of freedom are related to the sample size: Generally, the larger the sample size, the larger the degrees of freedom. When estimating a single mean, as we are doing here, the number of degrees of freedom is equal to the sample size minus one.

$$\text{df} = n - 1$$

The *t*-statistic does *not* follow the Normal distribution. One reason for this is that the denominator changes with every sample. For this reason, the *t*-statistic is more variable than the *z*-statistic (whose denominator is always the same.) Instead, if the three conditions for using the Central Limit Theorem hold, the *t*-statistic follows a distribution called—surprise!—the ***t*-distribution**. This was Gosset's great discovery at the Guinness brewery. When small sample sizes were used to make inferences about the mean, even if the population was Normal, the Normal distribution just didn't fit the results that well. Gosset discovered a new distribution, which he called the *t*-distribution, that turned out to be a better model than the Normal for the sampling distribution of $\bar{x}$ when σ is not known.

The *t*-distribution shares many characteristics with the $N(0, 1)$ distribution. Both are symmetric, are unimodal, and might be described as "bell-shaped." However, the

t-distribution has thicker tails. This means that in a t-distribution, it is more likely that we will see extreme values (values far from 0) than it is in a standard Normal distribution.

The t-distribution's shape depends on only one parameter, called the **degrees of freedom (df)**. The number of degrees of freedom is (usually) an integer: 1, 2, 3, and so on. If df is small, then the t-distribution has very thick tails. As the degrees of freedom get larger, the tails get thinner. Ultimately, when df is infinitely large, the t-distribution is exactly the same as the $N(0, 1)$ distribution.

Figure 9.10 shows t-distributions with 1, 10, and 40 degrees of freedom. In each case, the t-distribution is shown with a $N(0, 1)$ curve so that you can compare them. (We compare to the $N(0, 1)$ because it is familiar and because, as you can see, the t-distribution and the Normal distribution are very similar.) The t-distribution is the one whose tails are "higher" at the extremes. Note that by the time the degrees of freedom reaches 40 (Figure 9.10c), the t-distribution and the $N(0, 1)$ distribution are too close to tell apart (on this scale.)

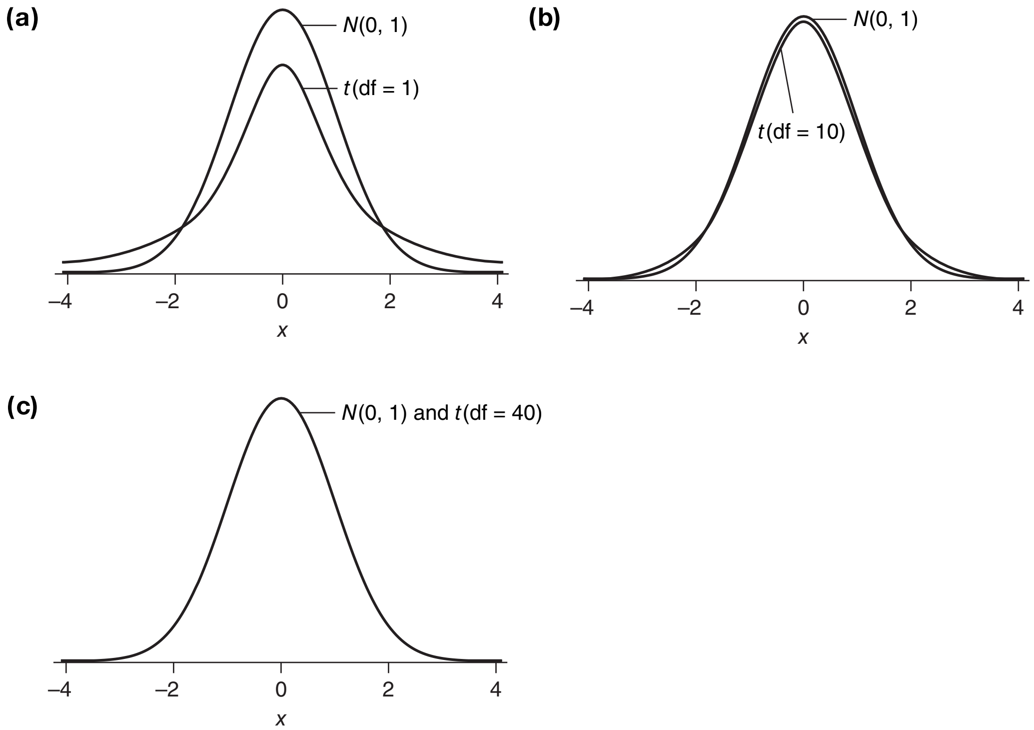


FIGURE 9.10 **(a)** A t-distribution with 1 degree of freedom, along with a $N(0, 1)$ distribution. The t-distribution has much thicker tails. **(b)** The degrees of freedom are now equal to 10, and the tails are only slightly thicker in the t-distribution. **(c)** The degrees of freedom are now equal to 40, and the two distributions are visually indistinguishable.

SECTION 9.3

Answering Questions about the Mean of a Population

Do you commute to work? How long does it take you to get there? Is this amount of time typical for others in your state? Which state has the greatest commuting times? This information is important not just to those of us who must fight traffic every day, but also to business leaders and politicians who make decisions about quality of living and the cost of doing business. The U.S. Census performs periodic surveys that determine, among other things, commuting times around the country. In 2012, the state of Maryland had the greatest mean commuting time, which was 31.9 minutes. South Dakota was lowest at 16.7 minutes.

These means are estimates of the mean commuting time for all residents in these states who work outside of their homes. We can learn the true mean commuting time only by asking all residents, which is clearly too time-consuming to do very often. Instead, the U.S. Census takes a random sample of U.S. residents to estimate these values.

In this section we present two techniques for answering questions about the population mean. Confidence intervals are used for estimating parameter values. Hypothesis tests are used for deciding whether a parameter's value is one thing or another. These are the same methods that were introduced in Chapter 7 (confidence intervals) and Chapter 8 (hypothesis tests) for population proportions, but here you'll see how they are modified to work with means.

Estimation with Confidence Intervals

Confidence intervals are a technique for communicating an estimate of the mean, along with a measure of the uncertainty in our estimate. The job of a confidence interval is to provide us with a range of plausible values that, according to the data, are highly plausible values for the unknown population mean. For instance, the range of plausible values for the mean commuting time for all South Dakota residents is 16.3 to 17.1 minutes.

Not all confidence intervals do an equally good job; the "job performance" of a confidence interval is therefore measured with something called the **confidence level**. The higher this level, the better the confidence interval performs. The confidence level for mean South Dakota commuting times is 90%, which means we can be confident that this interval contains the true mean.

Sometimes, you will be in a situation in which you will know only the sample mean and sample standard deviation. In these situations, you can use a calculator to find the confidence interval. However, if you have access to the actual data, you are much better off using statistical software to do all the calculations for you. We will show you how to respond to both situations.

No matter which situation you are in, you will need to judge whether a confidence interval is appropriate for the situation, and you will need to interpret the confidence interval. Therefore, we will discuss these essential skills before demonstrating the calculations.

When Are Confidence Intervals Useful? A confidence interval is a useful answer to the following questions: "What's the typical value for a variable in this large group of objects or people? And how far away from the truth might this estimate of the typical value be?" You should provide a confidence interval whenever you are estimating the value of a population parameter on the basis of a random sample from that population. For example, judging on the basis of a random sample of 30 adults, what's the typical body temperature of all healthy adults? On the basis of a survey of a random sample of Maryland residents, what's the typical commuting time for all Maryland residents? A confidence interval is useful for answering questions such as these because it communicates the uncertainty in our estimate and provides a range of plausible values.

A confidence interval is not appropriate if there is no uncertainty in your estimate. This would be the case if your "sample" were actually the entire population. For example, it is not necessary to find a confidence interval for the mean score on your class's statistics exam. The population is your class, and all of the scores are known. Thus the population mean is known, and there is no need to estimate it.

Checking Conditions In order for us to measure the correct confidence level, these conditions must hold:

Condition 1: *Random Sample and Independence*. The data must be collected randomly, and each observation must be independent of the others.

Condition 2: *Large Sample*. Either the population must be Normally distributed or the sample size must be fairly large (at least 25).

Condition 3: *Big Population*. If the sample is collected without replacement, then the population must be at least 10 times larger than the sample size.

If these conditions do not hold, then we cannot measure the job performance of the interval; the confidence level may be incorrect. This means that we may advertise a 95% confidence level when, in fact, the true performance is much worse than this.

To check the first condition, you must know how the data were collected. This is not always possible, so rather than checking these conditions, you must simply assume that they hold. If they do not, your interval will not be valid.

The requirement for independence means that measurement of one object in the sample does not affect any other. Essentially, if we know the value of any one observation, this knowledge should tell us nothing about the values of other observations. This condition might be violated if, say, we randomly sampled several schools and gave all of the students math exams. The individual math scores would not be independent, because we would expect that students within the same school might have similar scores.

The second condition is due to the Central Limit Theorem. If the population distribution is Normal (or very close to it), then we have nothing to worry about. But if it is non-Normal, then we need a large enough sample size so that the sampling distribution of sample means is approximately Normal. For many applications, a sample size of 25 is large enough, but for extremely skewed distributions, you might need an even larger sample size.

Throughout this chapter, we will assume that the population is large enough to satisfy the third condition, unless stated otherwise.

EXAMPLE 5 Is the Cost of College Rising?

Many cities and states are finding it more difficult to offer low-cost college educations. Did the mean cost of attending two-year colleges increase in the United States from the 2009–2010 academic year to 2012–2013? In 2009–2010, the mean cost of all two-year colleges was \$2517. A random sample of 35 two-year colleges in the United States found that the average tuition charged in 2012–2013 was \$2919, with a standard deviation of \$1079. Figure 9.11 provides the Minitab output, which shows that a 90% confidence interval for the mean cost of attending two-year colleges in 2012–2013 was \$2611 to \$3227.

One-Sample T

N	Mean	StDev	SE Mean	90% CI
35	2919	1079	182	(2611, 3227)

◀ **FIGURE 9.11** Minitab output for a 90% confidence interval of mean in-state tuition and fees of all two-year colleges in the United States during the 2012–2013 academic year.

QUESTIONS

a. Describe the population. Is the number \$2919 an example of a parameter or a statistic?

b. Verify that the conditions for a valid confidence interval are met.

SOLUTIONS

a. The population consists of all two-year college tuitions (for in-state residents) in the academic year 2012–2013. (There are over 1000 two-year colleges in the United States.) The number \$2919 is the mean of a sample of only 35 colleges. Because it is the mean of a *sample* (and not of a population), it is a statistic.

b. The first condition is that the data represent a random sample of independent observations. We are told the sample was collected randomly, so we assume this is true. Independence also holds, because knowledge about any one school's tuition tells us nothing about other schools in the sample. The second condition requires that the population be roughly Normally distributed or the sample size be equal to or larger than 25. We do not know the distribution of the population, but because the sample size is large enough (greater than 25), this condition is satisfied.

TRY THIS! Exercise 9.17

Interpreting Confidence Intervals To understand confidence intervals, you must know how to interpret a confidence interval and how to interpret a confidence level.

A confidence *interval* can be interpreted as a range of plausible values for the population parameter. In other words, in the case of population means, we can be confident that if we were to someday learn the true value of the population mean, it would be within the range of values given by our confidence interval. For example, the U.S. Census estimates that the mean commuting time for South Dakota residents is 16.3 minutes to 17.1 minutes, with a 90% confidence level. We interpret this to mean that we can be fairly confident that the true mean commute for *all* South Dakota residents is between 16.3 and 17.1 minutes. Yes, we could be wrong. The mean might be less than 16.3 minutes, or it might be more than 17.1 minutes. However, we would be rather surprised to find this was the case; we are highly confident that the mean is within this interval.

KEY POINT A confidence interval can be interpreted as a range of plausible values for the population parameter.

EXAMPLE 6 Evidence for Changing College Costs

Based on a random sample of 35 two-year colleges, a 90% confidence interval for the mean in-state tuition at two-year colleges for the 2012–2013 academic year is \$2611 to \$3227. When we examined the data for *all* two-year colleges in 2008–2009, we learned that the population mean in-state tuitions in 2008–2009 was \$2517.

QUESTION Does the confidence interval for mean tuitions in 2012–2013 provide evidence that the mean tuition has changed since the 2008–2009 academic year?

SOLUTION Yes, it does. Although, on the basis of this random sample of 35 colleges, we cannot know with certainty the population mean of all tuitions in 2012–2013, we are highly confident that it is in the range of \$2611 to \$3227. This range does *not* include the 2008–2009 value of \$2517, so there is evidence that the mean is higher than it was in 2008–2009.

TRY THIS! Exercise 9.19

Measuring Performance with the Confidence Level The confidence *level*, which in the case of both the intervals for mean commuting times and for mean tuition costs was 90%, tells us about the method used to find the interval. A value for the level of 90% tells us that the U.S. Census used a method that works in 90% of all samples. In other words, if we were to take many same-sized samples of commuters, and for each sample calculate a 90% confidence interval, then 90% of those intervals would contain the population mean.

The confidence level does *not* tell us whether the interval (16.3 to 17.1) contains or does not contain the population mean. The "90%" just tells us that the method that produced this interval is a pretty good method.

Details

Making Mistakes
Because the U.S. Census provides 90% confidence intervals for the mean commuting times in all 50 states, we can expect about 5 of these intervals (roughly 10% of them) to be wrong.

Suppose you decided to purchase a digital music player online. You have your choice of several manufacturers, and they are rated in terms of their performance level. One manufacturer has a 90% performance level, which means that 90% of the players it produces are good ones, and 10% are defective. Some other manufacturers have lower levels: 80%, 60%, and worse. From whom do you buy? You choose to buy from the manufacturer with the 90% level, because you can be very confident that the player it sends you will be good. Of course, once the player arrives at your home, the confidence level isn't too useful. Your player either works or does not work; there's no 90% about it.

Confidence levels work the same way. We prefer confidence intervals that have 90% or higher confidence levels, because then we know that the process that produced these levels is a good process, and therefore, we are confident in any decisions or conclusions we reach. But the level doesn't tell us whether this one particular interval sitting in front of us is good or bad. In fact, we shall never know that, unless we someday gain access to the entire population.

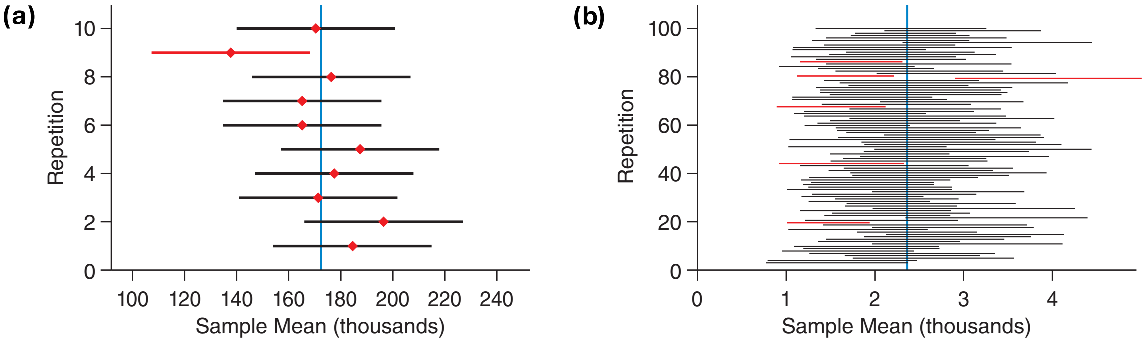
KEY POINT The confidence level is a measure of how well the method used to produce the confidence interval performs. We can interpret the confidence level to mean that if we were to take many random samples of the same size from the same population, and for each random sample calculate a confidence interval, then the confidence level is the proportion of intervals that "work"—the proportion that contain the population parameter.

! Caution

Confidence Levels Are Not Probabilities

A confidence level, such as 90%, is not a probability. Saying we are 90% confident the mean is between 21.1 minutes and 21.3 minutes does *not* mean that there is a 90% chance that the mean is between these two values. It either is, or isn't. There's no probability about it.

Figure 9.12 illustrates this interpretation of confidence levels. From the population of all U.S. movies that made over 100 million dollars (adjusted for inflation; http://www.thenumbers.com/), we took a random sample (with replacement) of 30 movies and calculated the mean revenue in this sample (in millions of dollars). Because the samples were random, each sample produced a different sample mean. For each sample we also calculated a 95% confidence interval. We repeated this process 100 times, and each time we made a plot of the confidence interval. Figure 9.12a shows the results from the first 10 samples of 30 randomly selected movies. Nine of the ten intervals were "good"—intervals that contained the true population mean of $172 million. Figure 9.11b shows what happened after we collected 100 different 95% confidence intervals. With a 95% confidence interval, we would expect about 95% of the intervals to be good and 5% to be bad. And in fact, six intervals (shown in red) were bad.



▲ **FIGURE 9.12** **(a)** Ten different 95% confidence intervals, each based on a separate random sample of 30 movies. The population mean of $172 million is shown with a vertical bar. Nine of the ten intervals are good because they include this population mean. **(b)** One hundred confidence intervals, each based on a random sample of 30 movies. Because we are using a 95% confidence level, we expect about 95% of the intervals to be good. In fact, 94 of the 100 turned out to be good, this time. The red intervals are "bad" intervals that do not contain the population mean.

EXAMPLE 7 iPad Batteries

A consumer group wishes to test Apple's claim that the iPad has a 10-hour battery life. With a random sample of 5 iPads, and running them under identical conditions, the group finds a 95% confidence interval for the mean battery life of an iPad to be 9.5 hours to 12.5 hours. One of the following statements is a correct interpretation of the confidence level. The other is a correct interpretation of the confidence interval.

(i) We are very confident that the mean battery life of all iPads is between 9.5 and 12.5 hours.

(ii) In about 95% of all samples of 5 iPads, the resulting confidence interval will contain the mean battery life of all iPads.

QUESTION Which of these statements is a valid interpretation of a confidence interval? Which of these statements is a valid interpretation of a confidence level?

SOLUTION Statement (i) interprets a confidence *interval* (9.5, 12.5). Statement (ii) tells us the meaning of the 95% confidence *level*.

TRY THIS! Exercise 9.23

Calculating the Confidence Interval You will be calculating confidence intervals in two situations. In the first situation, you'll have only summary statistics from the sample: the sample mean, the standard deviation, and the sample size. In this situation, you can often find the confidence interval using a calculator, although a computer will generally give a more accurate interval. In the second situation, you'll have the actual data. In this case, you should definitely use a computer. In both situations, it is very important to know the general structure of the formula in order to understand how confidence intervals are interpreted.

Confidence intervals for means have the same basic structure as they did for proportions:

$$\text{Estimator} \pm \text{margin of error}$$

As in Chapter 7, the margin of error has the structure

$$\text{Margin of error} = (\text{multiplier}) \times SE$$

The standard error (SE) is $SE = \frac{\sigma}{\sqrt{n}}$. Because we usually do not know the standard deviation of the population and hence the *standard error*, we replace SE with its estimate. This leads to a formula similar in structure, but slightly different in details, from the one you learned for proportions.

> **Looking Back**
> **CI Structure**
> Formula 7.2 gives the structure of a confidence interval for the population proportion. The details of the margin of error differ, but the structure is the same.

Formula 9.1: One-Sample *t*-Interval

$$\bar{x} \pm m$$

$$\text{where} \qquad m = t^* SE_{\text{EST}} \text{ and } SE_{\text{EST}} = \frac{s}{\sqrt{n}}$$

The multiplier t^* is a constant that is used to fine-tune the margin of error so that it has the level of confidence we want. This multiplier is found using a t-distribution with $n - 1$ degrees of freedom. (The degrees of freedom determine the shape of the t-distribution.) SE_{EST} is the estimated standard error.

To compute a confidence interval for the mean, you first need to choose the level of confidence. After that, you need either the original data or these four pieces of information:

1. The sample average, $\bar{x}$, which you calculate from the data.
2. The sample standard deviation, s, which you calculate from the data.
3. The sample size, n, which you know from looking at the data.
4. The multiplier, t^*, which you look up in a table (or use technology) and which is determined by your confidence level and the sample size n. The value of t^* tells us how wide the margin of error is, in terms of standard errors. For example, if t^* is 2, then our margin of error is two standard errors wide.

The structure of the confidence interval for the mean is the same as for the proportion: the estimated value plus or minus the margin of error. One difference is that for the proportion, the multiplier in the margin of error depends only on our desired confidence level. If we want a 95% confidence level, we always use 1.96 as the multiplier value z^* (and we sometimes round to 2). If we want a 90% confidence level, we always use 1.64 for z^*. For means, however, the multiplier is also determined by the sample size.

The reason for this is that when we are finding a confidence interval for a proportion, our confidence level is determined by the Normal distribution. But when we are working with the mean, the level is determined by the t-distribution based on $n - 1$ degrees of freedom. The correct values can be found in Table 4 in Appendix A, or you can use technology. Table 4 is organized such that each row represents possible values of t^* for each degree of freedom. The columns contain the values of t^* for a given confidence level. For example, for a 95% confidence level and a sample size of $n = 30$, we use $t^* = 2.045$. We find this in the table by looking in the row with $\text{df} = n - 1 = 30 - 1 = 29$ and using the column for a 95% confidence level. Refer to Table 9.3, which is from the table in Appendix A.

Example 8 shows how to use Table 4 to find the multiplier, a technique that is useful if you do not have access to a statistical calculator.

EXAMPLE 8 Finding the Multiplier t^*

A study to test the life of iPad batteries reports that in a random sample of 30 iPads, the mean battery life was 9.7 hours and the standard deviation was 1.2 hours. The raw data were not made available to the public.

QUESTION Using Table 9.3, which is from the table in Appendix A, find t^* for a 90% confidence interval when $n = 30$.

DF	Confidence Level			
	90%	95%	98%	99%
28	1.701	2.048	2.467	2.763
29	<u>1.699</u>	2.045	2.462	2.756
30	1.697	2.042	2.457	2.750
34	1.691	2.032	2.441	2.728

◀ **TABLE 9.3** Critical values of t.

SOLUTION We find the number of degrees of freedom from rd the sample size:

$$\text{df} = n - 1 = 30 - 1 = 29$$

And so we find, from Table 9.3, $t^* = 1.699$ (shown underlined)

TRY THIS! Exercise 9.25

Looking Back

Why Not 100%?
In Chapter 7, you learned that one reason why a 95% confidence level is popular is that increasing the confidence level only a small amount beyond 95% requires a much larger margin of error.

It is best to use technology to find the multiplier, because most tables stop at 35 or 40 degrees of freedom. For a 95% confidence level, if you do not have access to technology and the sample size is bigger than 40, it is usually safe to use $t^* = 1.96$—the same multiplier that we used for confidence intervals for sample proportions (for 95% confidence). The precise value, if we used a computer, is 2.02, but this is only 0.06 unit away from 1.96, so the result is probably not going to be affected in a big way.

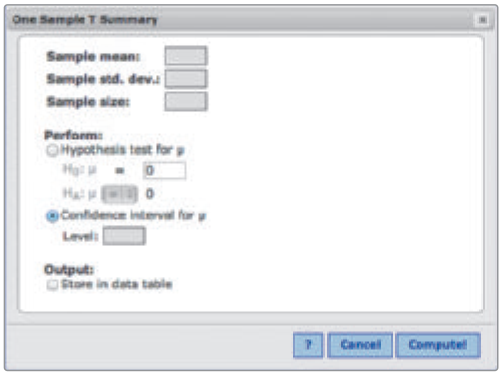
Example 9 illustrates the use of statistical software to find a confidence interval when only summary statistics are provided.

EXAMPLE 9 Pizza Size

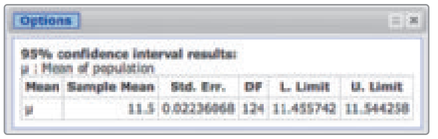
Eagle Boys, an Australian chain of pizza stores, published data on the size of its pizzas to convince the public of their value. Using a random sample of 125 pizzas, the store found that the mean diameter was 11.5 inches with a standard deviation of 0.25 inch. Figure 9.13a shows a screenshot from StatCrunch for calculating a confidence interval using summary statistics. Figure 9.13b shows the results (Dunn 2012).

QUESTION For each field in StatCrunch, provide the value or setting required to calculate a 95% confidence interval for the mean diameter of all pizzas produced by this store. State and interpret the confidence interval.

(a)



(b)



▶ **FIGURE 9.13** **(a)** StatCrunch fields for calculating a confidence interval for the population mean when only summary statistics are available. **(b)** Result of the calculation.

SOLUTION

Sample mean: 11.5
Sample standard deviation: 0.25
Sample size: 125
Confidence interval level for μ: 0.95

The 95% confidence level for the mean pizza diameter is 11.46 to 11.54 inches. We are 95% confident that the mean diameter of all pizzas produced by this company is between 11.46 and 11.54 inches.

TRY THIS! Exercise 9.27

Example 10 shows that in order to have a higher level of confidence, we need a larger margin of error. This larger margin of error means that the confidence interval is wider, so our estimate is less precise.

EXAMPLE 10 College Tuition Costs

A random sample of 35 two-year colleges in 2012–2013 had a mean tuition (for in-state students) of \$2918, with a standard deviation of \$1079.

QUESTION Find a 90% confidence interval and a 95% confidence interval for the mean in-state tuition of all two-year colleges in 2008–2009. Interpret the intervals. First, verify that the necessary conditions hold.

SOLUTION This time, we will show how to find the confidence interval using the formula, which is what you must do if you do not have a statistical calculator or statistical software.

We are told that the sample is random and that the sample size is larger than 25, so the necessary conditions hold.

We are given the desired confidence level, the standard deviation, and the sample mean, $\bar{x}$, so the next step is to calculate the estimated standard error.

$$\bar{x} = 2918$$

$$SE_{\text{EST}} = \frac{1079}{\sqrt{35}} = 182.3843$$

$$\bar{x} \pm m \text{ is}$$

$$\bar{x} \pm t^* SE_{\text{EST}}$$

We find the appropriate values of t^* (from Table 9.3):

$$t^* \text{ (for 90\%)} = 1.691$$

$$t^* \text{ (for 95\%)} = 2.032$$

For the 90% confidence interval,

$$\bar{x} \pm t^* SE_{\text{EST}} \text{ becomes } 2918 \pm (1.691 \times 182.3843), \text{ or } 2918 \pm 308.4119$$

$$\text{Lower limit: } 2918 - 308.4119 = 2609.59$$

$$\text{Upper limit: } 2918 + 308.4119 = 3226.41$$

A 90% confidence interval for the mean tuition of all two-year colleges in the 2012–2013 academic year is (\$2610, \$3226).

For the 95% confidence interval,

$$\bar{x} \pm t^* SE_{\text{EST}} \text{ becomes } 2918 \pm (2.032 \times 182.3843) \text{ or } 2918 \pm 370.6049$$

$$\text{Lower Limit: } 2918 - 370.6049 = 2547.395$$

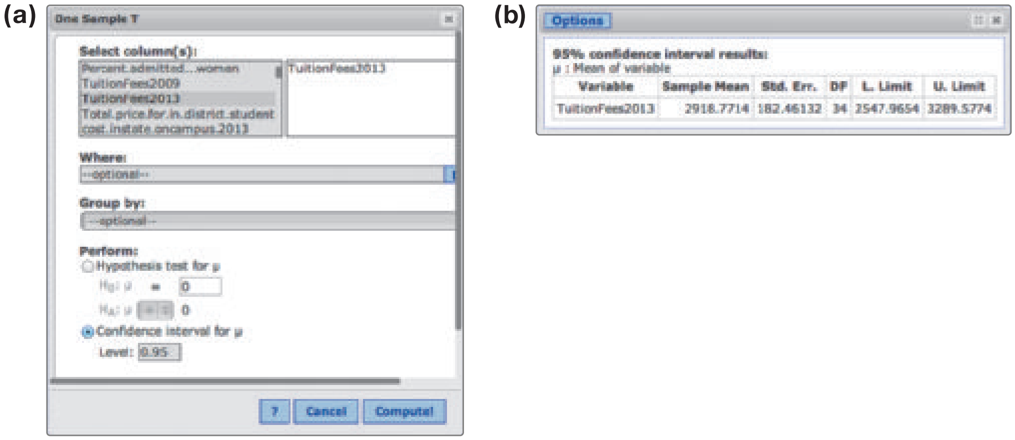
$$\text{Upper Limit: } 2918 + 370.6049 = 3288.605$$

CONCLUSION The 90% confidence interval is (\$2610, \$3226). The 95% confidence interval is (\$2547, \$3289), which is wider. We are 90% confident that the mean tuition (that is, the typical tuition) of all two-year colleges is between \$2610 and \$3226. We are 95% confident that the mean tuition of all two-year colleges is between \$2547 and \$3289.

TRY THIS! Exercise 9.29

If you have access to the original data (and not just to the summary statistics, as we were given in Example 10), then it is always best to use a computer to find the confidence interval for you. Figure 9.14a shows the information required by StatCrunch to calculate a 95% confidence interval for the mean tuition of two-year

Tech



▶ **FIGURE 9.14** Screenshot showing **(a)** required StatCrunch input and **(b)** the resulting output for a 95% confidence interval for the mean in-state tuition at two-year colleges in the 2012-2013 academic year.

colleges in the 2012–2013 academic year using the data themselves (which you'll recall are drawn from a random sample of 35 two-year colleges.) The output in Figure 9.14b shows the estimated mean (\$2918.7714), the standard error (\$182.46132), the degrees of freedom (34), the lower limit of the confidence interval (\$2547.9654), and the upper limit (\$3289.5774).

Reporting and Reading Confidence Intervals There are two ways of reporting confidence intervals. Professional statisticians tend to report (lower boundary, upper boundary). This is what we've done so far in this chapter. Thus in Example 10 we reported the 95% confidence interval for the mean of two-year college tuitions in 2012–2013 as (\$2547, \$3289).

In the press, however, and in some scholarly publications, you'll also see confidence intervals reported as

$$\text{Estimate} \pm \text{margin of error}$$

For the two-year college tuitions, we calculated the margin of error to be \$370.605 for 95% confidence. Thus we could also report the confidence interval as

$$\$2918 \pm \$371$$

This form is useful because it shows our estimate for the mean (\$2918) as well as our uncertainty (the mean could plausibly be \$371 lower or \$371 more).

You're welcome to choose whichever you think best, although you should be familiar with both forms.

Understanding Confidence Intervals As shown in Chapter 7 and illustrated in Example 10, a wider interval results in a higher level of confidence. Imagine that the population mean is a tennis ball and the confidence interval is a tennis racquet. Which would make you more confident of hitting the ball: using a (small) ping-pong paddle-sized racquet or using a (larger) tennis racquet? The larger racquet fills more space, so you should feel more confident that you'll connect with the ball. Using a wider confidence interval gives us a higher level of confidence that we'll "connect" with the true population mean.

Wider intervals are not always desirable, however, because they mean that we have less precision. For example, we could have a 100% confidence interval for the mean tuition at two-year colleges: \$0 to infinity dollars. But this interval is so imprecise that it is useless. The 95% confidence interval offers less than 100% confidence, but it is much more precise.

Because the margin of error depends on the standard error, and because the standard error depends on the sample size, we can make our interval more precise by collecting more data. A larger sample size provides a smaller standard error, and this means a smaller margin of error at the same level of confidence.

EXAMPLE 11 Confidence in Growing Enrollment

The IPEDS website provides much data on colleges and universities in the United States. We took a random sample of 35 two-year colleges from this data set and calculated two confidence intervals for the mean total enrollment of all two-year colleges in the United States. Both intervals are based on the same sample. One of the intervals is 4948 students to 16170 students. The other is 3815 students to 17303 students. IPEDS stands for Integrated Postsecondary Education Data System.

QUESTIONS

a. Which interval has the higher confidence level, and why?

b. Suppose we take a larger sample size (say, 40 colleges). Assuming that the estimated mean and standard deviation remain the same (which in fact will be approximately the case), what will be the effect on the confidence level?

c. What will be the effect of taking a larger sample on the width of the interval?

SOLUTION

a. The interval (3815, 17303) has the higher level of confidence because it is the wider interval. This interval has width $17303 - 3815 = 13{,}488$ students and is wider than the other interval, which is 11,222 students wide.

b. The confidence level of both intervals will be unchanged because we are still using the same multiplier to calculate the intervals, and the multiplier determines the confidence level.

c. If we take a larger sample, the standard error for our estimator will be smaller. This means the margin of error will be smaller, so both intervals will be narrower.

NOW TRY! Exercise 9.31

SECTION 9.4

Hypothesis Testing for Means

In Chapter 8 we laid out the foundations of hypothesis testing. Here, you'll see that the same four steps can be used to test hypotheses about means of populations. These four steps are

Step 1: Hypothesize.
State your hypotheses about the population parameter.

Step 2: Prepare.
Get ready to test: Choose and state a significance level. Choose a test statistic appropriate for the hypotheses. State and check conditions required for the computations, and state any assumptions that must be made.

Step 3: Compute to compare.
Compute the observed value of the test statistic in order to compare the null hypothesis value to our observed value. Find the p-value to measure your level of surprise.

Step 4: Interpret.
Do you reject or not reject the null hypothesis? What does this mean?

As an example of testing a mean, consider this "study" one of the authors did. McDonald's advertises that its ice cream cones have a mean weight of 3.2 ounces ($\mu = 3.2$). A human server starts and stops the machine that dispenses the ice cream, so we might expect some variation in the amount. Some cones might weight slightly more, some cones slightly less. If we weighed all of the McDonald's ice cream cones at a particular store, would the mean be 3.2 ounces, as the company claims?

One of the authors collected a sample of five ice cream cones (all in the name of science) and weighed them on a postage scale. The weights were (in ounces)

$$4.2, 3.6, 3.9, 3.4, \text{ and } 3.3$$

We summarize these data as

$$\bar{x} = 3.68 \text{ ounces}, s = 0.3701 \text{ ounce}$$

Do these observations support the claim that the mean weight is 3.2 ounces? Or is the mean a different value? We'll apply the four steps of the hypothesis test to make a decision.

Step 1: Hypothesize

Hypotheses come in pairs and are always statements about the population. In this case, the population consists of all ice cream cones that have been, will be, or could be dispensed from a particular McDonald's. In this chapter, our hypotheses are about the mean values of populations.

The null hypothesis is the status-quo position, which is the claim that McDonald's makes. An individual cone might weigh a little more than 3.2 ounces, or a little less, but after looking at a great many cones, we would find that McDonald's is right and the mean weight is 3.2 ounces.

We state the null hypothesis as

$$H_0: \mu = 3.2$$

Recall that the null hypothesis always contains an equals sign.

The alternative hypothesis, on the other hand, says that the mean weight is different from 3.2 ounces:

$$H_a: \mu \neq 3.2$$

This is an example of a two-sided hypothesis. We will reject the null hypothesis if the average of our sample cones is very big (suggesting that the population mean is greater than 3.2) or very small (suggesting that the population mean is less than 3.2). It is also possible to have one-sided hypotheses, as you will see later in this chapter.

KEY POINT Hypotheses are always statements about population parameters. For the test you are about to learn, this parameter is always μ, the mean of the population.

Details

What Value for α?
For most situations, using a significance level of 0.05 is a good choice and is recommended by many scientific journals. Values of 0.01 and 0.10 are also commonly used.

Step 2: Prepare

The first step is to set the significance level α (alpha), as we discussed in Chapter 8. The significance level is a performance measure that helps us evaluate the quality of our test procedure. It is the probability of making the mistake of rejecting the null hypothesis when, in fact, the null hypothesis is true. In this case, this is the probability that we will say McDonald's cones do not weigh an average of 3.2 ounces when, in fact, they really do. We will use $\alpha = 0.05$.

The test statistic, called the one-sample *t*-test, is very similar in structure to the test for one proportion and is based—not surprisingly, given the name of the test—on the *t*-statistic introduced in Section 9.2. The idea is simple: Compare the observed value of the sample mean, $\bar{x}$, to the value claimed by the null hypothesis, μ_0.

Formula 9.2: Test Statistic for the One-Sample *t*-Test

$$t = \frac{\bar{x} - \mu_0}{SE_{\text{EST}}}, \qquad \text{where} \qquad SE_{\text{EST}} = \frac{s}{\sqrt{n}}$$

If conditions hold, the test statistic follows a t-distribution with df $= n - 1$

Looking Back

The *z*-Test
Compare this to the z-test statistic for one proportion in Chapter 8, which has a very similar structure:

$$z = \frac{\hat{p} - p_0}{SE}$$

This test statistic works because it compares the value of the parameter that the null hypothesis says is true, μ_0, to the estimate of that value that we actually observed in our data. If the estimate is close to the null hypothesis value, then the t-statistic is close to 0. But if the estimate is far from the null hypothesis value, then the t-statistic is far from 0. The farther t is from 0, the worse things look for the null hypothesis.

Anyone can make a decision, but only a statistician can measure the probability that the decision is right or wrong. To do this, we need to know the sampling distribution of our test statistic.

The sampling distribution will follow the t-distribution under these conditions:

Condition 1: *Random Sample and Independence.* The data must be a random sample from a population, and observations must be independent of one another.

Condition 2: *Large Sample.* The population distribution must be Normal, or the sample size must be large. For most situations, 25 is large enough.

Now let's apply this to our ice cream problem. The population for testing the mean ice cream cone weight is somewhat abstract, because a constant stream of ice cream cones is being produced by McDonald's. However, it seems logical that if some cones weigh slightly more than 3.2 ounces and some weigh slightly less, then this distribution of weights should be symmetric and not too different from a Normal distribution. Because our population distribution is Normal, the fact that we have a small sample size, $n = 5$, is not a problem here.

The ice cream cone weights are independent of each other because we were careful, when weighing, to recalibrate the scale, and each cone was obtained on a different day. The cones were not, strictly speaking, randomly sampled, although because the cones were collected on different days and at different times, we will assume that they behave as though they come from a random sample. (But if we're wrong, our conclusions could be *very* wrong!)

Step 3: Compute to compare

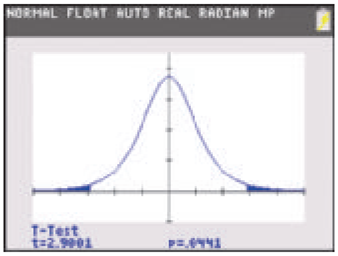
The conditions of our data tell us that our test statistic should follow a t-distribution with $n - 1$ degrees of freedom. Therefore, we proceed to do the calculations necessary to compare our observed sample mean to the hypothesized value of the population mean, and to measure our surprise.

To find the observed value of our t-statistic, we need to find the sample mean and the standard deviation of our sample. These values are given above, but you can easily calculate them from the data.

$$SE_{\text{EST}} = \frac{0.3701}{\sqrt{5}} = 0.1655$$

$$t = \frac{3.68 - 3.2}{0.1655} = \frac{0.48}{0.1655} = 2.90$$

The observed sample mean was 2.90—almost 3 standard errors above the value expected by the null hypothesis.



▲ **FIGURE 9.15** The tail areas above 2.90 and below −2.90 are shown as the small shaded areas on both sides. The p-value is 0.0441, the probability that if $\mu = 3.2$, a test statistic will be bigger than 2.90 or smaller than −2.90. The distribution shown is a t-distribution with 4 degrees of freedom.

KEY POINT The t-statistic measures how far (how many standard errors) away our observed mean, $\bar{x}$, lies from the hypothesized value, μ_0. Values far from 0 tend to discredit the null hypothesis.

How unusual is such a value, according to the null hypothesis? The p-value tells us exactly that—the probability of our getting a t-statistic as extreme as or more extreme than what we observed, if in fact the mean is 3.2 ounces.

Because our alternative hypothesis says we should be on the lookout for t-statistic values that are much bigger or smaller than 0, we must find the probability in both tails of the t-distribution. The p-value is shown in the small shaded tails of Figure 9.15. Our sample size is $n = 5$, so our degrees of freedom are $n - 1 = 5 - 1 = 4$.

The p-value of 0.044 tells us that if the typical cone really weighs 3.2 ounces, our observations are somewhat unusual. We should be surprised.

Step 4: Interpret

The last step is to compare the p-value to the significance level and decide whether to reject the null hypothesis. If we follow a rule that says we will reject whenever the p-value is less than or equal to the significance level, then we know that the probability of mistakenly rejecting the null hypothesis will be the value of α.

Our p-value (0.044) is less than the significance level we chose (0.05), so we should reject the null hypothesis and conclude that at this particular McDonald's, cones do *not* weigh, on average, 3.2 ounces.

This result makes some sense from a public relations standpoint. If the mean were 3.2 ounces, about half of the customers would be getting cones that weighed too little. By setting the mean weight a little higher than what is advertised, McDonald's can give everyone more than they thought they were getting.

Looking Back

p-Values

In Chapter 8 you learned that the p-value is the probability that when the null hypothesis is true, we will get a test statistic as extreme as or more extreme than what we actually saw. (What is meant by "extreme" depends on the alternative hypothesis.) The p-value measures our surprise at the outcome.

One- and Two-sided Alternative Hypotheses

The alternative hypothesis in the ice cream cone test was two-sided. As you learned in Chapter 8, alternative hypotheses can also be one-sided. The exact form of the alternative hypothesis depends on the research question. In turn, the form of the alternative hypothesis tells us how to find the p-value. Two-sided hypotheses will require two-tailed p-value calculations, and one-sided hypotheses will require one-tailed p-value calculations.

You will always use one of the following three pairs of hypotheses for the one-sample t-test:

Two-Sided	One-Sided (Left)	One-Sided (Right)
$H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$	$H_0: \mu = \mu_0$ $H_a: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_a: \mu > \mu_0$

Looking Back

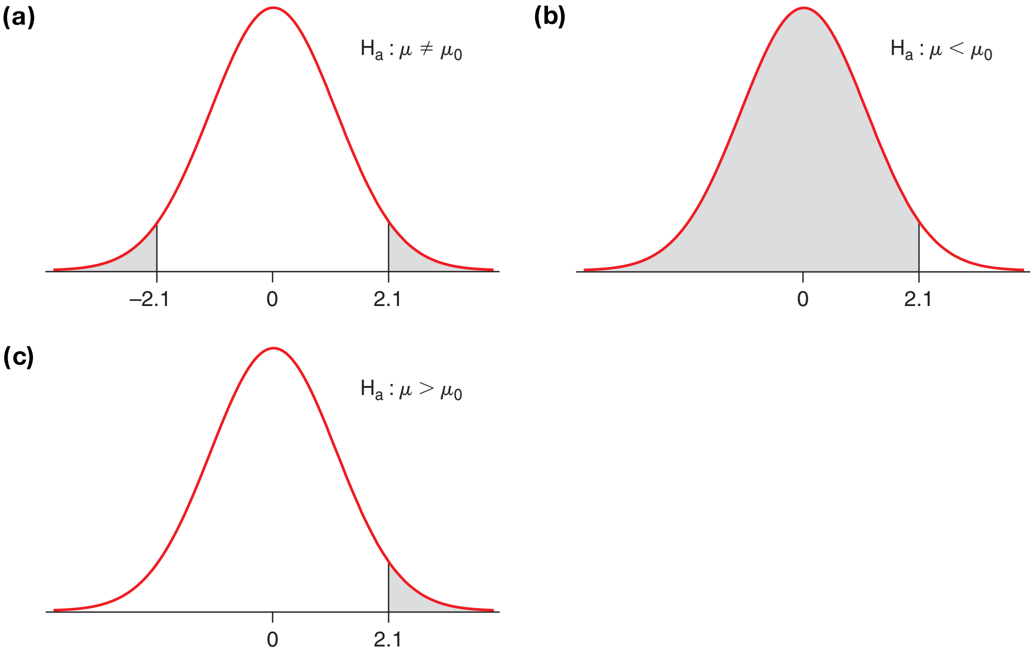
What Does "as extreme as or more extreme than" Mean?

See Chapter 8 for a detailed discussion of how the p-value depends on the alternative hypothesis.

You choose the pair of hypotheses on the basis of your research question. For the ice cream cone example, we asked if the mean weight is *different* from the value advertised, so we used a two-sided alternative hypothesis. Had we instead wanted to know whether the mean weight was less than 3.2 ounces, we would have used a one-sided (left) hypothesis.

Your choice of alternative hypothesis determines how you calculate the p-value. Figure 9.16 shows how to find the p-value for each alternative hypothesis, all using the same t-statistic value of $t = 2.1$ and the same sample size of $n = 30$.

Note that the p-value is always an "extreme" probability; it's always the probability of the tails (even if the tail is pretty big, as it is in Figure 9.16b).



◀ **FIGURE 9.16** The distributions are t-distributions with $n - 1 = 29$ degrees of freedom. The shaded region in each graph represents a p-value when $t = 2.1$ for **(a)** a two-sided alternative hypothesis (two-tailed probability), **(b)** a one-sided (left) hypothesis (left-tailed probability), and **(c)** a one-sided (right) hypothesis (right-tailed probability).

EXAMPLE 12 College Costs

In Example 5 we asked whether the mean cost of two-year colleges had increased since the 2008–2009 academic year. In 2008–2009, the mean cost of all two-year colleges (for in-state tuition and fees) was $2517. We will now examine the same question using a hypothesis test. Our data are a random sample of tuition prices at 35 two-year colleges during the 2012–2013 academic year. Figure 9.17a shows summary statistics for this sample, and Figure 9.17b shows the distribution of the sample.

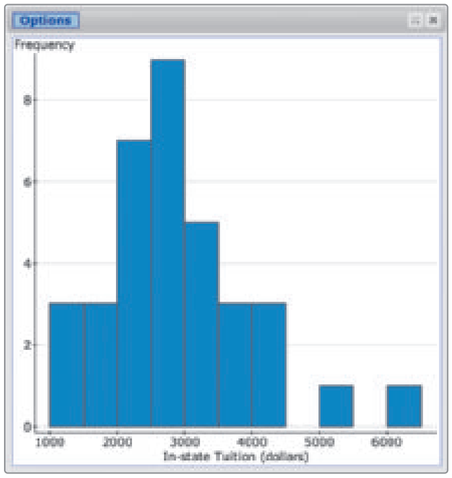
(a)

Options

Summary statistics:

Column	n	Mean	Std. dev.	Median	Min	Max	Q3	Q1	IQR
TuitionFees2013	35	2918.7714	1079.4557	2760	1140	6120	3455	2234	1221

(b)



▲ **FIGURE 9.17** Summary statistics for tuitions at 35 two-year colleges. **(a)** The sample mean tuition for these 35 colleges was $2919, give or take $1079. **(b)** The distribution is relatively symmetric, with two outliers.

QUESTION Test the hypothesis that the mean tuition cost at two-year colleges has increased since the 2009–2010 academic year.

SOLUTION First, we note that if we had the data for *all* two-year colleges in 2012–2013 (as we did for the 2008–2009 academic year), we would not need to do a hypothesis test. We could simply compute the population mean and see whether it was greater than the population mean in 2008–2009. But because we have only a sample of all of the two-year colleges, we must use a hypothesis test.

Step 1: Hypothesize
Let μ represent the mean tuition at all two-year colleges in 2012–2013. We are asked to compare the population mean in 2012–2013 with the reported population mean of \$2517 from 2008–2009.

$$H_0 : \mu = 2517$$
$$H_a : \mu > 2517$$

The null hypothesis says the mean is the same as in 2008–2009. The alternative hypothesis says the mean is higher now than it was then. This differs from the ice cream cone example, where we only wanted to know whether or not the mean weight was 3.2. Here we care about the direction: Is the new mean higher than the old mean?

Step 2: Prepare
We will test using a 5% significance level.

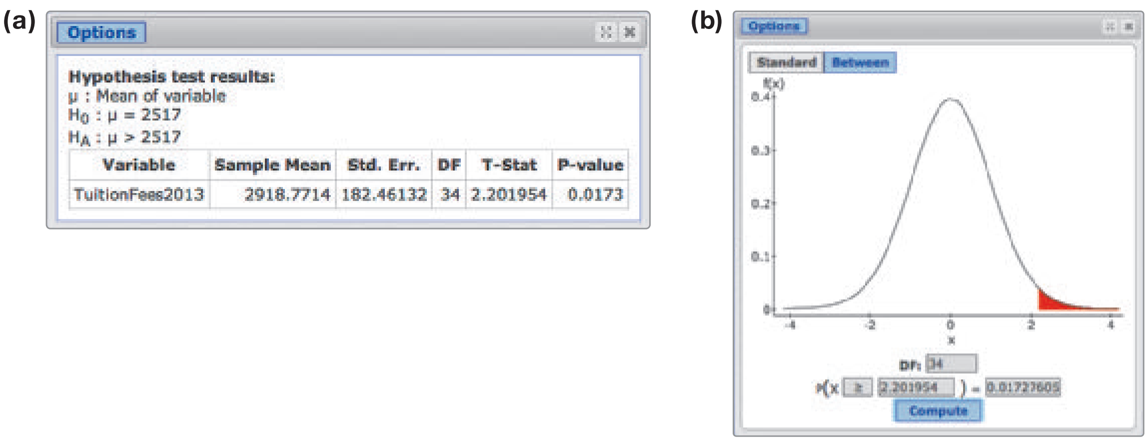
We need to check the conditions to see whether the *t*-statistic will follow a *t*-distribution (at least approximately).

Condition 1: *Random Sample and Independence*. The colleges in this study were selected randomly from the population of all two-year colleges. Our sampling procedure ensured that observations were independent of each other.

Condition 2: *Large Sample*. The distribution of the sample is not necessarily Normal (although it is not too different from Normal). But because the sample size is larger than 25, this sampling distribution will be approximately a *t*-distribution with $n - 1 = 34$ degrees of freedom.

Step 3: Compute to compare
This step is usually done using technology, and StatCrunch output is shown in Figure 9.18a. We will use Formula 9.2 to show how it is applied in this situation.

$$SE_{EST} = \frac{1079.4557}{\sqrt{35}} = 182.4613$$
$$t = \frac{\bar{x} - \mu_0}{SE_{EST}} = \frac{401.7714}{182.4613} = 2.20$$



▲ **FIGURE 9.18** StatCrunch was used to perform the calculations to enable us to compare the null hypothesis value of 2517 to our observed sample mean. **(a)** The *t*-statistic value is 2.20 with a p-value of 0.017. **(b)** The p-value is a right-tailed probability, using a *t*-distribution with 34 degrees of freedom.

This tells us that our observed mean was 2.20 standard errors above where we would expect it to be if the null hypothesis were true.

Our p-value is a right-tailed probability because the alternative hypothesis cares only whether we see values greater than expected. (Remember, we are using a one-sided hypothesis.) Using technology, we find the p-value to be 0.0173, which is illustrated in Figure 9.18b.

Step 4: Interpret
The p-value is smaller than 0.05, so we conclude that the mean tuition cost *is* higher in 2012–2013 than it was in 2009–2010.

TRY THIS! Exercise 9.37

SNAPSHOT ONE SAMPLE *t*-TEST

WHAT IS IT? ▶ $t = \dfrac{\bar{x} - \mu_0}{SE_{\text{EST}}}$, where $SE_{\text{EST}} = \dfrac{s}{\sqrt{n}}$

WHAT DOES IT DO? ▶ It tests hypotheses about a mean of a single population.

HOW DOES IT DO IT? ▶ If the estimated mean differs from the hypothesized value, then the test statistic will be far from 0. Thus, values of the *t*-statistic that are unexpectedly far from 0 (in one direction or the other) discredit the null hypothesis.

HOW IS IT USED? ▶ When proposing values about a single population mean. The observed value of the test statistic is compared to a *t*-distribution with $n - 1$ degrees of freedom to calculate the p-value. If the p-value is small, you should be surprised by the outcome and reject the null hypothesis.

SECTION 9.5

Comparing Two Population Means

Do people comprehend better when they read on paper rather than on a computer screen? Do men spend less time doing laundry than women? If so, how much less? These questions can be answered, in part, by comparing the means of two populations. Although we could construct separate confidence intervals to estimate each mean, we can construct more precise estimates by focusing on the difference between the two means.

Just as in Section 7.5, when you learned about interpreting the differences between two proportions, when we say we are looking at the difference between two means, we mean that we are going to subtract one mean from the other. For instance, we might subtract the mean amount of time Americans spend watching TV every day from the mean amount of time they spend exercising. From subtraction, we learn that

If the result is positive, the first mean is greater than the second.
If the result is negative, the first mean is less than the second.
If the result is 0, the means are equal.

Caution

Paired (Dependent) vs. Independent Samples
One indication that you have paired samples is that each observation in one group is coupled with *one particular observation* in the other group. In this case, the two groups will have the same sample size (assuming no observations are missing).

When comparing two populations, it is important to pay attention to whether the data sampled from the populations are two **independent samples** or are, in fact, one sample of related pairs (paired samples). With **paired (dependent) samples**, if you know the value that a subject has in one group, then you know something about the other group, too. In such a case, you have somewhat less information than you might have if the samples were independent. We begin with some examples to help you see which is which.

Usually, dependence occurs when the objects in your sample are all measured twice (as is common in "before and after" comparisons), or when the objects are related somehow (for example, if you are comparing twins, siblings, or spouses), or when the experimenters have deliberately matched subjects in the groups to have similar characteristics.

EXAMPLE 13 Independent or Dependent Samples?

Here are four descriptions of research studies.

a. People chosen in a random sample were asked how many minutes they had spent the day before watching television, and how many minutes they had spent exercising. Researchers want to know how different the mean amounts of times are for these two activities.

b. Men and women each had their sense of smell measured. Researchers want to know whether, typically, men and women differ in their ability to sense smells.

c. Researchers randomly assigned overweight people to one of two diets: Weight Watchers and Atkins. Researchers want to know whether the mean weight loss on Weight Watchers was different from that on Atkins.

d. The numbers of years of education for husbands and wives are compared to see whether the means are different.

QUESTION For each study, state whether it involves two independent samples or paired (that is, dependent) samples.

SOLUTIONS

a. This study is based on a single sample of people who are measured twice. One population consists of people watching TV. The second population consists of the same people exercising. These are *paired* (dependent) samples.

b. The two populations consist of men in one and women in the other. As long as the people are not related, knowledge about a measurement of a man could not tell us anything about any of the women. These are *independent* samples.

c. The two populations are people on the Weight Watchers diet and people on the Atkins diet. We are told that the two samples consist of different people; subjects are randomly assigned to one diet but not to the other. These are *independent* samples.

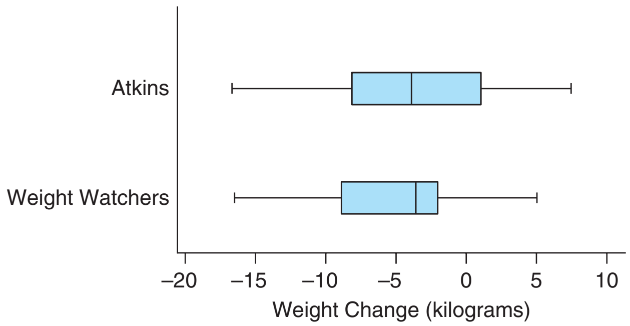
d. The populations are matched. Each husband is coupled with one particular wife, so the samples are *paired* (or *dependent)*.

TRY THIS! Exercise 9.53

As you shall see, we analyze paired data differently from data that come from two independent samples. Paired data are turned into "difference" scores: We simply subtract one value in each pair from the other. We now have just a single variable, and we can analyze it using the one-sample techniques discussed in Sections 9.3 and 9.4.

Estimating the Difference of Means with Confidence Intervals (Independent Samples)

The Weight Watchers diet is a very traditional, low-calorie diet. The Atkins diet, on the other hand, limits the amount of carbohydrates. Which diet is more effective? Researchers compared these two diets (as well as two others) by randomly assigning overweight subjects to the two diet groups. The boxplots in Figure 9.19 show summary statistics for the samples' weight losses after one year. The mean weight loss of the Weight Watchers dieters was 4.6 kilograms (about 10 pounds), and the mean loss of the Atkins dieters was 3.9 kilograms (Dansinger et al. 2005).



◀ **FIGURE 9.19** Weight change (kilograms) for people randomly assigned to the Atkins diet or the Weight Watchers diet. The medians are close, but because of the skew in the distributions, the sample means are slightly less close.

Looking Back

Boxplots
In Section 3.5, we explained that the middle line in the boxplot represents the median values, and the box includes the middle 50% of values.

To guarantee a particular confidence level—for example, 95%—requires that certain conditions hold:

Condition 1: *Random Samples and Independence.* Both samples are randomly taken from their populations, or subjects are randomly assigned to one of the two groups, and each observation is independent of any other.

Condition 2: *Independent Samples.* The two samples are independent of each other (not paired).

Condition 3: *Large Samples.* The populations are approximately Normal, or the sample size in each sample is 25 or more. (In special cases, you might need even larger sample sizes.)

If these conditions hold, we can use the following procedure to find an interval with a 95% confidence level.

The formula for a confidence interval comparing two means, when the data are from independent samples, is the same structure as before:

$$\text{(Estimate)} \pm \text{margin of error}$$

which is

$$\text{(Estimate of difference)} \pm t^*(SE_{\text{estimate of difference}})$$

We estimate the difference with

$$\text{(Mean of first sample)} - \text{(mean of second sample)}$$

It doesn't matter which sample you use as the "first" and which as the "second." But it *is* important that you remember which is which.

The standard error of this estimator depends on the sample sizes of both samples and also on the standard deviations of both samples:

$$SE_{\text{EST}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

We can put these together into a confidence interval:

Formula 9.3: Two-Sample *t*-Interval

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

The multiplier t^* is based on an approximate t-distribution. If a computer is not available, you can conservatively calculate the degrees of freedom for the t^* multiplier as the smaller of $n_1 - 1$ and $n_2 - 1$, but a computer provides a more accurate value.

Choosing the value of t^* (the critical value of t) by hand to get your desired level of confidence is tricky. For reasons requiring some pretty advanced mathematics to explain, the sampling distribution is not a t-distribution, but only approximately a t-distribution. To make matters worse, to get the approximation to be good requires using a rather complex formula to find the degrees of freedom. If you must do these calculations by hand, we recommend taking a "fast and easy" (but also safe and conservative) approach instead. For t^*, use a t-distribution with degrees of freedom equal to the smaller of $n_1 - 1$ and $n_2 - 1$. That is, use the smaller of the two samples, and subtract 1. For a 95% confidence level, if both samples contain 40 or more observations, you can use 1.96 for the multiplier.

Interpreting Confidence Intervals of Differences

The most important thing to look for is whether or not the interval includes 0. If it does *not,* then we have evidence that the two means are different from each other. In this case, check to see whether the interval contains all positive values. If so, then we are confident that the first mean is greater than the second mean. If the interval contains all negative values, we are confident that the first mean is less than the second mean. (Remember, you get to choose which mean is "first" and which is "second" when you do the subtraction.)

For instance, do men spend less time doing laundry than women? Let's use women as the first group and men as the second. Based on a random sample of about 12,000 people carried out by the U.S. Bureau of Labor Statistics, a 95% confidence interval for the difference in the mean amount of time spent doing laundry for women compared to men is (11.1 to 13.7) minutes per day. Because this interval does not include 0, we are confident that the mean times spent doing laundry are not the same. Because the interval contains all positive values, we're confident that women spend between 11.1 and 13.7 minutes *more* per day doing laundry than do men, on average.

You should also pay attention to how great or how small the difference between the two means could be. This is particularly important if the interval includes 0. In this case, the interval will contain both negative and positive values.

Example 14 shows you how to calculate and interpret a confidence interval for the difference of two means.

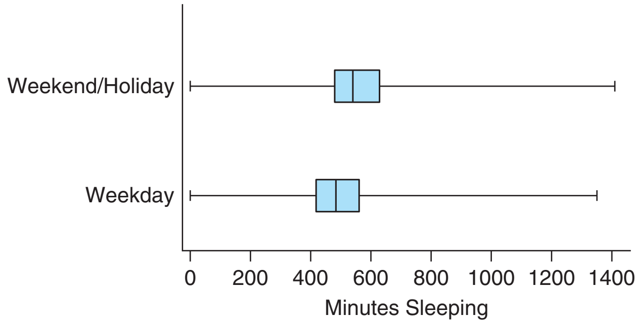
EXAMPLE 14 Sleeping In

Do people in the United States sleep more on holidays and weekends than on weekdays? The Bureau of Labor Statistics carries out a "time use" survey, in which randomly chosen people are asked to record every activity they do on a randomly chosen day of the year. For instance, you might be chosen to take part in the survey on Tuesday, April 18, while someone else will be chosen to take part on Sunday, December 5.

Because we have two separate groups of people reporting their amount of sleep—one group that reported only on weekends and holidays, and another group that reported only on weekdays—these data are two independent samples. The summary statistics follow.

Weekday: $\bar{x} = 499.7$ minutes (about 8.3 hours), $s = 126.9$ minutes, $n = 6007$
Weekend/Holiday: $\bar{x} = 555.9$ minutes (about 9.3 hours), $s = 140.9$ minutes, $n = 6436$

Boxplots are shown in Figure 9.20.

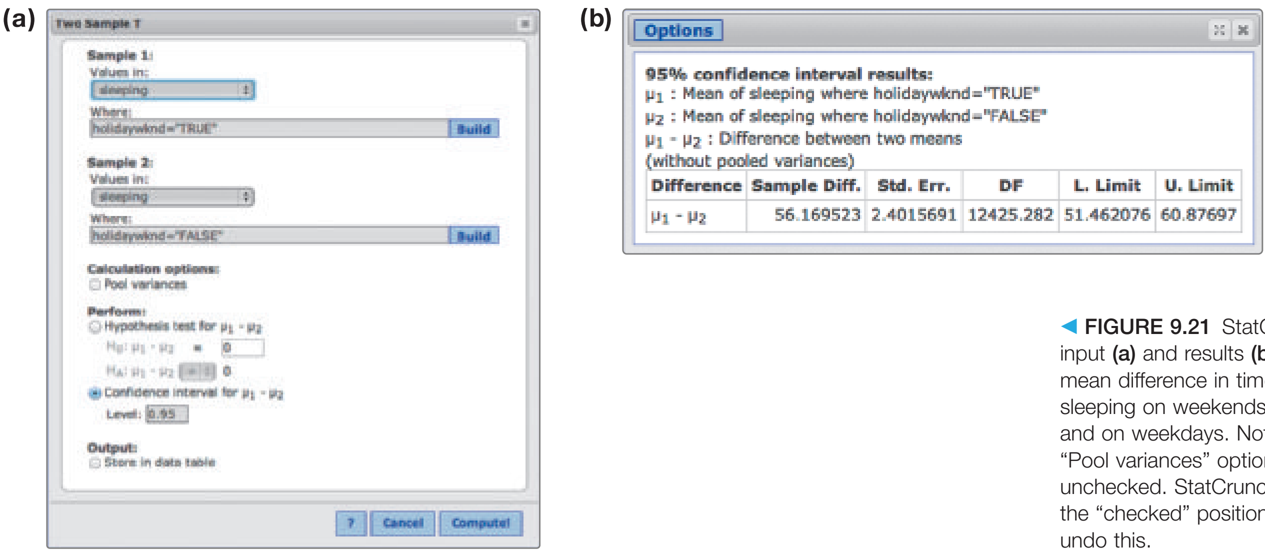


◀ **FIGURE 9.20** Distribution of minutes spent sleeping on weekends and holidays, and on weekdays. The median sleeping time is slightly greater on weekends and holidays.

QUESTION Verify that the necessary conditions hold to guarantee that our 95% confidence level will indeed be 95%. Find a 95% confidence interval for the mean difference in time spent sleeping on a weekend/holiday compared to time spent sleeping on a weekday. Interpret this interval.

SOLUTION We are told that people are selected randomly, and it also seems reasonable that the amount of sleep they get is independent. The question explained that the samples were themselves independent of each other (because different people appear in each sample.) The boxplots suggest that the distributions are roughly symmetric; it is difficult to know for sure with boxplots. However, with such large sample sizes, we do not have to worry and can proceed.

When we have the raw data at hand, the best approach (and, for a sample as large as this, the only approach) is to use a computer. Figure 9.21a shows the input required to get StatCrunch to compute the interval for us. Figure 9.21b shows the output.



◀ **FIGURE 9.21** StatCrunch input **(a)** and results **(b)** for the mean difference in time spent sleeping on weekends/holidays and on weekdays. Note that the "Pool variances" option should be unchecked. StatCrunch defaults to the "checked" position, so you must undo this.

The 95% confidence interval is (51.5 to 60.9) minutes. We are 95% confident that the true mean difference in amount of time spent sleeping on weekends/holidays and amount of time spent sleeping on weekdays is between 51.5 minutes and 60.9 minutes. The interval contains all positive values, so we are confident that people typically sleep longer on weekends and holidays than they do on weekdays. The difference may be as little as 51.5 minutes or as much as 60.9 minutes. Generally, it looks like people tend to sleep about an hour later when they do not have to work or go to school.

Notice that the degrees of freedom are not a whole number: 12425.282. Statistical software packages apply a fairly complex formula to approximate the degrees of freedom. If you are doing this "by hand," we recommend that you use the simpler method of taking the smallest sample size minus 1, as demonstrated below, but for sample sizes this large, it makes very little difference.

TRY THIS! Exercise 9.55

Tech

If we do not have the original data and are provided only summary statistics, then we might use a calculator to find the interval for the difference between the mean amount of time spent sleeping for the people who reported on weekends and holidays, and the mean amount of time spend sleeping for the people who reported on weekdays.

How would this work for finding a confidence interval for the difference in mean sleeping times in Example 14? First, we must find the multiplier t^*. The sample sizes of both groups are quite large. Our rule of thumb for finding an approximate number of degrees of freedom for this test is to use the smaller of the two sample sizes and subtract 1. The smaller sample size is 6007, so we will use 6006 as the degrees of freedom. Because this is much greater than 40, we simply use 1.96 as a conservative approximation to t^*. (If the degrees of freedom are greater than 40, then for t^* use the values for the Normal distribution: 1.96 for 95% and 1.64 for 90%.)

We'll use the Weekend/Holiday group as our sample 1, and the Weekday group as sample 2.

Estimate of difference: $555.9 - 499.7 = 56.2$ minutes

$$m = t^*\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = t^*\sqrt{\frac{140.9^2}{6436} + \frac{126.9^2}{6007}} = t^*2.401137$$

So using $t^* = 1.96$, we have $m = 1.96 \times 2.401137 = 4.706229$ minutes. Therefore, a 95% confidence interval is

$$56.2 \pm 4.706229, \text{ or about } (51.5, 60.9) \text{ minutes.}$$

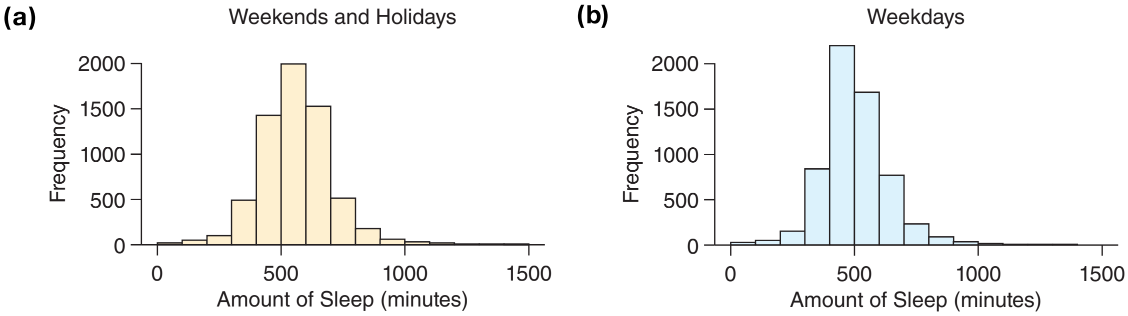
This is very close to what the software found.

Testing Hypotheses about Two Means

Hypothesis tests to compare two means from independent samples follow the same structure we discussed in Chapter 8, although the details change slightly because we now are comparing means rather than proportions. We show this structure by revisiting our comparison of sleeping times between weekends and weekdays. In Example 14 you found a confidence interval for the mean difference. Here we approach the same data with a hypothesis test.

In Example 14 we used boxplots to investigate the shape of the distribution of amount of sleep. Here, we examine histograms (Figure 9.22), which show a more detailed picture of the distributions. Both distributions have roughly the same amount of spread, and the histograms seem fairly symmetric.

We call weekend and holiday sleepers "population 1" and weekday sleepers "population 2." Then the symbol μ_1 represents the mean amount of sleep for all people in the United States on weekends and holidays, and μ_2 represents the mean amount of sleep for all people in the United States on weekdays.



▲ **FIGURE 9.22** Amount of sleep reported on **(a)** weekends and holidays and **(b)** weekdays. Note that 500 minutes is about 8 hours and 18 minutes.

Step 1: Hypothesize

H_0: $\mu_1 = \mu_2$ (typical amount of sleep is the same on weekends/holidays as weekdays)

H_a: $\mu_1 \neq \mu_2$ (typical amount of sleep time is different on weekends/holidays than on weekdays.)

Step 2: Prepare

The conditions for testing two means are not very different from those for testing one mean and are identical to those for finding confidence intervals of the difference of two means.

Condition 1: *Random Samples and Independent Observations*. Observations are taken at random from two populations, producing two samples, or all subjects are randomly assigned to one of the two groups. Observations within a sample are independent of one another, which means that knowledge of one value tells us nothing about other observed values in that sample.

Condition 2: *Independent Samples.* The samples are independent of each other. Knowledge about a value in one sample does not tell us anything about any value in the other sample.

Condition 3: *Large Sample*. Both populations are approximately Normal, or both sample sizes are 25 or more. (In extreme situations, larger sample sizes may be required.)

Condition 1 holds because we are told that the people were selected randomly and independently. And because the two groups consist of different people, condition 2 holds. The distributions of the samples both look reasonably symmetric, but with such large sample sizes (over 6000 per group), condition 3 also holds.

Another step in our preparation is to choose a significance level. It is common to use $\alpha = 0.05$, and we will do so for this example.

Step 3: Compute to compare

The test statistic used to test this hypothesis is based on the difference between the sample means. Basically, the test statistic measures how far away the observed difference in sample means is from the hypothesized difference in population means. Yes, you guessed it: The distance is measured in terms of standard errors.

$$t = \frac{\text{(difference in sample means } - \text{ what null hypothesis says the difference is)}}{SE_{\text{EST}}}$$

Using the test statistic is made easier by the fact that the null hypothesis almost always says that the difference is 0.

Details

Null Hypotheses for Two Means

Mathematically, we can easily adjust our test statistic if the null hypothesis claims that the difference in means is some value other than 0. But in a great many scientific, business, and legal settings, the null hypothesis value will be 0.

$$\text{Difference in sample means: } \bar{x}_1 - \bar{x}_2 = 555.9 - 499.7 = 56.2$$

(The summary statistics are given in Example 14.)

$$SE_{\text{EST}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{140.9^2}{6436} + \frac{126.9^2}{6007}} = 2.401137$$

$$t = \frac{56.2}{2.401137} = 23.4$$

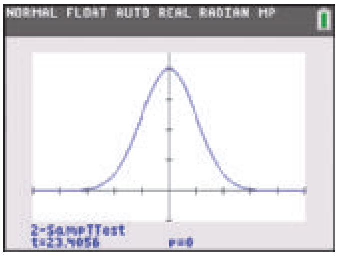
Formula 9.4: Two-Sample *t*-Test

$$t = \frac{\bar{x}_1 - \bar{x}_2 - 0}{SE_{\text{EST}}}, \text{ where } SE_{\text{EST}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

If all the conditions are met, the test statistic follows an approximate t-distribution, where the degrees of freedom are conservatively estimated to be the smaller of $n_1 - 1$ and $n_2 - 1$.

If there is no difference in the mean amounts of sleep in the United States, then the sample means should be nearly equal, and their difference should be close to 0. Our t-statistic tells us that the difference in means is 23.4 standard errors away from where the null hypothesis expects it to be.

Intuitively, you should understand that this t-statistic is extremely large and therefore casts quite a bit of doubt on our null hypothesis. But let's measure how surprising this is. To do so, we need to know the sampling distribution of the test statistic t, because we measure our surprise by finding the probability that if the null hypothesis were true, we would see a value as extreme as or more extreme than the value we observed. In other words, we need to find the p-value.



▲ **FIGURE 9.23** The TI-84 output shows us that the p-value is nearly 0, because it is extremely unlikely the t statistic will be more than 23.4 standard errors away from 0 when the null hypothesis is true.

If the conditions listed in the Prepare step hold, then t follows, approximately, a t-distribution with minimum $(n_1 - 1, n_2 - 1)$ degrees of freedom. This approximation can be made even better by adjusting the degrees of freedom, but this adjustment is, for most cases, too complex for a "by hand" calculation. For this reason, we recommend using technology for two-sample hypothesis tests, because you will get more accurate p-values.

The smallest of the sample sizes is 6007, so our conservative number of degrees of freedom is $6007 - 1 = 6006$.

Our alternative hypothesis is two-sided, and it says that the true difference is either much bigger than 0 or much smaller than 0, so we use a two-tailed area under the t-distribution. Figure 9.23 shows this calculation using a statistical calculator. As you might have guessed, the value of 23.4 is so extreme that the area is invisible.

The p-value is nearly 0.

Caution

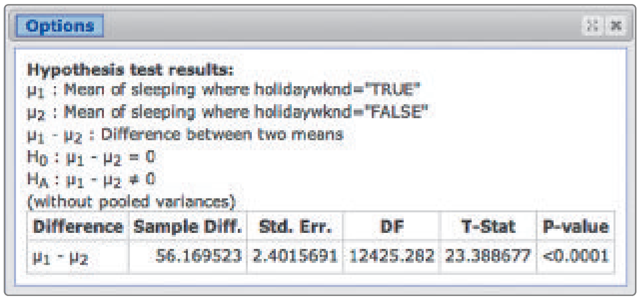
Don't Accept!

Remember from Chapter 8 that we do not "accept" the null hypothesis. It is possible that the sample size is too small (the test has low power) to detect the real difference that exists. Instead, we say that there is not enough evidence for us to reject the null.

Step 4: Interpret

Again, we compare the p-value to the significance level, α. If the p-value is less than or equal to α, we reject the null hypothesis. In this example, the p-value is nearly 0, so it is certainly much less than 0.05. If people do tend to sleep the same amount of time on weekends as on weekdays, then this outcome is extremely surprising. Nearly impossible, in fact. Therefore, we reject the null hypothesis and conclude that people do tend to sleep a different amount on weekends and holidays than on weekdays.

The previous analysis was done using only the summary statistics provided. If you have the raw data, then you should use computer software to do the analysis. You will get more accurate values and save yourself lots of time. Figure 9.24 shows StatCrunch output for testing whether the mean sleeping time on weekends and holidays is different from the mean sleeping time on weekdays.



Options

Hypothesis test results:
μ_1 : Mean of sleeping where holidaywknd="TRUE"
μ_2 : Mean of sleeping where holidaywknd="FALSE"
$\mu_1 - \mu_2$: Difference between two means
$H_0 : \mu_1 - \mu_2 = 0$
$H_A : \mu_1 - \mu_2 \neq 0$
(without pooled variances)

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	56.169523	2.4015691	12425.282	23.388677	<0.0001

◀ **FIGURE 9.24** StatCrunch output to test whether people typically sleep a different amount of time on weekends and holidays than they do on weekdays.

! Caution

Don't Pool
When using software to do a two-sample *t*-test, make sure it does the unpooled version. You might have to tell the software explicitly. The unpooled version is more accurate in more situations than the pooled version.

Into the Pool Some software packages, and some textbooks too, provide for another version of this *t*-test called the "pooled two-sample *t*-test." We have presented the unpooled version (you can see this in the StatCrunch output above the table, where it says "without pooled variances"). The unpooled version is preferred over the other version because the pooled version works only in special circumstances (when the population standard deviations are equal). The unpooled version works reasonably well in all situations, as long as the listed conditions hold.

SNAPSHOT TWO SAMPLE *t*-TEST (FROM INDEPENDENT SAMPLES)

WHAT IS IT? ▶ A procedure for deciding whether two means, estimated from independent samples, are different. The test statistic used is

$$t = \frac{\bar{x}_1 - \bar{x}_2 - 0}{SE_{EST}}, \text{ where } SE_{EST} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

WHAT DOES IT DO? ▶ Provides us with a decision on whether to reject the null hypothesis that the two means are the same and lets us do so knowing the probability that we are making a mistake.

HOW DOES IT DO IT? ▶ Compares the observed difference in sample means to 0, the value we expect if the population means are equal.

HOW IS IT USED? ▶ The observed value of the test statistic can be compared to a *t*-distribution.

Hypotheses: Choosing Sides So far, we've presented the hypotheses with one mean on the left side and one on the right, like this: H_0: $\mu_1 = \mu_2$. But you will generally see hypotheses written as differences: H_0: $\mu_1 - \mu_2 = 0$. These two hypotheses mean exactly the same thing, because if the means are equal, then subtracting one from the other will result in 0.

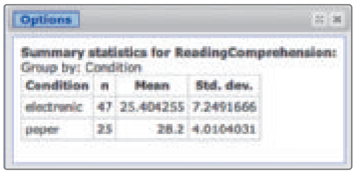
As with the other hypothesis tests you've seen, there are three different alternative hypotheses we can use. Here are your choices:

Two-Sided	One-Sided (Left)	One-Sided (Right)
H_0: $\mu_1 - \mu_2 = 0$ H_a: $\mu_1 - \mu_2 \neq 0$	H_0: $\mu_1 - \mu_2 = 0$ H_a: $\mu_1 - \mu_2 < 0$	H_0: $\mu_1 - \mu_2 = 0$ H_a: $\mu_1 - \mu_2 > 0$

EXAMPLE 15 Reading Electronics

More and more often, people are reading on computer screens or other electronic "e-readers." Do we read differently when we read on a computer screen than we do when we read material on ordinary paper? Researchers in Norway carried out a study to determine whether children (of high school age) read differently when reading material from a pdf on a computer screen than when reading a printed copy. Specifically, they measured whether reading comprehension differed between the two types of material.

To carry out this study, 72 tenth grade students were randomly assigned to one of two groups, which we'll call "electronic" and "paper." All students were asked to read two texts, both of roughly equal length. However, the students in the "electronic" group read the texts on computer screens, and the "paper" group read them on paper. The texts were formatted so that they appeared the same both on the computer screen and on paper. After reading, all students took the same reading comprehension test (Mangen et al. 2013).



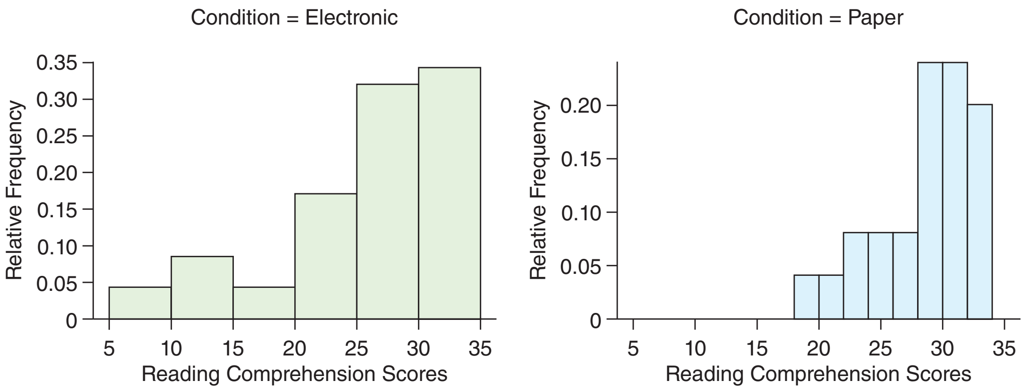
Options

Summary statistics for ReadingComprehension:
Group by: Condition

Condition	n	Mean	Std. dev.
electronic	47	25.404255	7.2491666
paper	25	28.2	4.0104031

▲ **FIGURE 9.25** Summary statistics produced by StatCrunch.

Figure 9.25 provides summary statistics for the reading comprehension scores for the two groups, and Figure 9.26 shows their histograms. The typical reading comprehension score is larger for the students who read on paper, which indicates that they typically had a greater level of understanding of what they had read.



► **FIGURE 9.26** The distribution of reading comprehension scores for tenth grade students. One group read the texts on a computer screen, the other on paper.

QUESTION Carry out the four steps of a hypothesis test to test whether students who read on paper have a different level of comprehension than those who read on computer screens. Use a significance level of 5%. If a required condition (in step 2) doesn't hold, explain the consequences and state any assumptions you must make in order to continue with the test. In step 3, refer to Figure 9.27, which displays an excerpt of StatCrunch output for this hypothesis test.

SOLUTION

Step 1: Hypothesize

We will let μ_1 represent the mean reading comprehension score of all tenth grade students in Norway who might read these texts in *paper* format, and we'll let μ_2 represent the mean reading comprehension scores of all tenth grade students in Norway who might read these texts on a *computer*.

$$H_0: \mu_1 = \mu_2$$
$$H_a: \mu_1 \neq \mu_2$$

(You might also write these as $H_0: \mu_1 - \mu_2 = 0$ and $H_a: \mu_1 - \mu_2 \neq 0$.)

Step 2: Prepare

We've been asked to use $\alpha = 0.05$, so all that remains of step 2 is to check whether the necessary conditions are met.

We do not have a random sample of tenth grade Norwegians. This means we cannot generalize our results to all students of this age from Norway. However, because random assignment was used, we can conclude that, if we find a statistically significant difference, it is due to the reading condition and not to a (possibly unidentified) confounding factor.

The second condition holds because we are told that different children are in the different groups. Finally, although the distributions are left-skewed, the sample sizes are large, so we can use the *t*-distribution. (But note that the paper group is just barely large enough, $n = 25$. It is possible that our p-value might be more approximate than we would like.)

Step 3: Compute to compare

Referring to Figure 9.27, we see that the value of the *t*-statistic is 2.11, which tells us that the observed difference in means was 2.1 standard errors above the value we would expect if the null hypothesis were true. The p-value is 0.0388.

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	2.7957447	1.3271877	69.844774	2.1065179	0.0388

◀ **FIGURE 9.27** Excerpt from statistical software to test whether mean reading comprehension for students reading on paper (group 1) was different from reading comprehension for students reading on a computer (group 2).

Step 4: Interpret

Because the p-value of 0.0388 is less than our significance level of 0.05, we reject the null hypothesis. For these students, mean reading comprehension for those reading on paper was different from reading comprehension for those reading on a computer screen.

TRY THIS! Exercise 9.57

The researchers themselves did a different analysis from the one we have presented. Their analysis took into account additional factors that we chose to leave out so that we could focus on the structure of the hypothesis test. The researchers' conclusion was that reading comprehension was higher for students reading on paper than for students reading on the computer.

CI for the Mean of a Difference: Dependent Samples

With paired samples, we turn two samples into one. We do this by finding the difference in each pair. For example, researchers wanted to know whether our sensitivity to smells is different when we were sitting up compared to when we are lying down (Lundström et al. 2006). They devised a measure of the ability of a person to detect smells, and then they measured a sample of people twice: once when they were lying down and once when they were sitting upright.

How much might this smell-sensitivity score change, if at all? One way to answer this question is with a confidence interval for the mean difference in scores. These data differ from the examples you've seen with two independent samples in that even though there are two groups (lying down, sitting upright), the data for both groups come *from the same people.* For this reason, these are *dependent,* or *paired,* samples.

When we are dealing with paired samples, our approach is to transform the original data from two variables (lying down and sitting upright, or, if you prefer, group 1 and group 2), into a single variable that contains the difference between the scores in group 1 and group 2. As before, it doesn't matter which is group 1 and which is group 2, but we do have to remember our choice. Once that is done, we have a single sample of difference scores, and we apply our one-sample confidence interval from Formula 9.1.

For the "smell" study, we create the new difference variable by subtracting each person's score when lying down from her or his score when sitting.

The first few lines of the original data are shown in Table 9.4a.

▶ **TABLE 9.4a** Smelling ability for the first four people, sitting and lying down.

Subject Number	Sex	Sitting	Lying Down
1	Woman	13.5	13.25
2	Woman	13.5	13
3	Woman	12.75	11.5
4	Man	12.5	12.5

We create a new variable, call it *Difference*, and define it to be the difference between smelling ability sitting upright and smelling ability lying down. We show this new variable in Table 9.4b.

▶ **TABLE 9.4b** Difference between smelling ability while sitting upright and smelling ability while lying down.

Subject Number	Sex	Sitting	Lying Down	Difference
1	Woman	13.5	13.25	0.25
2	Woman	13.5	13	0.50
3	Woman	12.75	11.5	1.25
4	Man	12.5	12.5	0

Here are summary statistics for the *Sitting*, *Lying,* and *Difference* variables:

Variable	*n*	Sample Mean	Sample Standard Deviation
Sitting	36	11.47	3.26
Lying	36	10.60	3.06
Difference	36	0.87	2.39

After verifying that the necessary conditions hold (they do), we can find a 95% CI for the mean difference using Formula 9.1. Because we have 36 observations, the degrees of freedom for t^* are $36 - 1 = 35$. The table in Appendix A tells us that $t^* = 2.03$.

Thus a 95% confidence interval for the mean difference in smelling scores is $\bar{x} \pm t^*\frac{s}{\sqrt{n}}$, or $0.87 \pm 2.03\frac{2.39}{\sqrt{35}}$, which works out (after rounding) to (0.05, 1.69).

Because these values are all positive, we conclude that the mean for the first group (sitting upright) is higher than the mean for the second group (lying down). For this reason, we are confident that smelling sensitivity is greater when sitting upright than when lying down. This difference could be fairly small, 0.05 unit, or as large as 1.7 units.

Many statistical software packages allow you to compute the confidence interval for two means in a paired sample by selecting a "paired two-sample" option. Example 16 shows how statistical software can be used to find a confidence interval for the difference of two means when the data are from paired samples.

EXAMPLE 16 Dieting

Americans who want to lose weight can choose from many different diets. In one study (Dansinger et al. 2005), researchers compared results from four different diets. In this example, though, we look only at a small part of these data, and examine only the 40 subjects who were randomly assigned to the Weight Watchers diet. These subjects were measured twice: at the start of the study and then 2 months later. The data consist of two variables: weight (in kilograms) at 0 months and weight at 2 months. One question we can ask is, How much weight will the typical person lose on the Weight Watchers diet after 2 months?

Figure 9.28 shows two different 90% confidence intervals for these data. Figure 9.28a shows a confidence interval for independent samples, and Figure 9.28b shows a confidence interval for paired samples. Note that these data are not a random sample, but random assignment was used.

(a)

Options

90% confidence interval results:
μ_1 : Mean of WEIGHT_2
μ_2 : Mean of WEIGHT_0
$\mu_1 - \mu_2$: Difference between two means
(without pooled variances)

Difference	Sample Diff.	Std. Err.	DF	L. Limit	U. Limit
$\mu_1 - \mu_2$	-3.628125	1.8981057	317.76806	-6.4294271	-0.82682293

(b)

Options

90% confidence interval results:
$\mu_D = \mu_1 - \mu_2$: Mean of the difference between WEIGHT_2 and WEIGHT_0 where "DIET"="Weight Watchers"

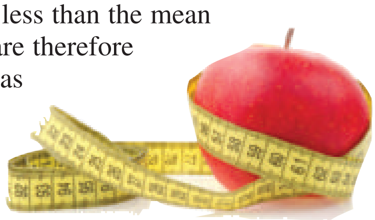
Difference	Sample Diff.	Std. Err.	DF	L. Limit	U. Limit
WEIGHT_2 - WEIGHT_0	-3.465	0.60560999	39	-4.4853772	-2.4446228

▲ **FIGURE 9.28** StatCrunch output showing 90% confidence intervals **(a)** as if the data were independent samples and **(b)** as if they were paired samples.

QUESTION State and interpret the correct interval.

SOLUTION Because the same subjects are in both samples (because each subject was measured twice), the data are paired samples. For this reason, Figure 9.28b shows the correct output.

The 90% confidence interval for the mean difference in weight is –4.5 kg to –2.4 kg. The fact that the interval contains only negative values means that the mean weight in the first measurement (at 2 months) is less than the mean weight in the second measurement (at 0 months). We are therefore confident that the typical subject lost weight: as much as 4.5 kg (about 10 pounds) or as little as 2.4 kg (about 5 pounds).



TRY THIS! Exercise 9.67

Test of Two Means: Dependent Samples

In Example 16, we asked a question about amounts: *How much* weight did the typical Weight Watchers dieter lose? Sometimes researchers aren't as interested in "How much?" as in answering the question "Did anything change at all?"

Questions such as this can be answered with hypothesis tests about paired data. In this case, as with confidence intervals for paired data, we convert the two variables into a difference variable, and our hypotheses are now not about the individual groups, but about the difference.

To illustrate, let's consider another subgroup of the dieters' data. The diet program known as The Zone promises that you'll lose weight, burn fat, and not feel hungry. The diet requires that you eat 30% protein, 30% fat, and 40% carbs, and it also imposes restrictions on the times at which you eat your meals and snacks. Can people lose weight on The Zone?

In words, our null hypothesis is that after 2 months of dieting, the mean weight of people on The Zone diet is the same as the mean weight before dieting. Our alternative hypothesis is that the mean weight is *less* after 2 months of dieting (a one-sided hypothesis.)

For the data we are analyzing, subjects were randomly assigned to one of four diets, although we will consider only those on The Zone. For each subject, weight was measured at 0 months and at 2 months. Because the same subjects appear in both groups, the data are *paired*. Instead of considering weight at 0 months and weight at

2 months as separate variables, we will calculate the *change* in weight and name this variable *Difference*. For each subject,

$$\text{Difference} = (\text{weight at 2 months}) - (\text{weight at 0 months})$$

Our hypotheses are now about just one mean, the mean of *Difference*:

$$\text{H}_0\text{: } \mu_{\text{difference}} = 0 \quad (\text{or } \mu_{\text{2months}} = \mu_{\text{0months}})$$

$$\text{H}_a\text{: } \mu_{\text{difference}} < 0 \quad (\text{or } \mu_{\text{2months}} < \mu_{\text{0months}})$$

Our test statistic is the same as for the one-sample t-test:

$$t = \frac{\bar{x}_{\text{difference}} - 0}{SE_{\text{difference}}}, \quad \text{where } SE_{\text{difference}} = \frac{s_{\text{difference}}}{\sqrt{n}}$$

We find $\bar{x}$ by averaging the difference variable: $\bar{x} = -3.795$ kilograms.

We find $s_{\text{difference}}$ by finding the standard deviation of the difference variable: $s =$ 3.5903 kg.

There were 40 subjects, so

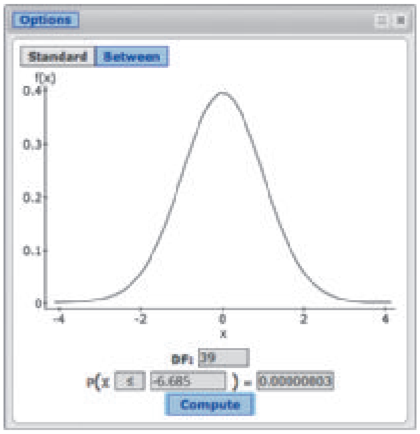
$$SE = \frac{3.5903}{\sqrt{40}} = 0.5677$$

and then

$$t = \frac{-3.795}{0.5677} = -6.685$$

To find the p-value, we use a t-distribution (assuming the conditions for a one-sample t-test hold) with $n - 1$ degrees of freedom, where n is the number of data pairs. Because there are 40 pairs, one for each subject, the degrees of freedom are 39. The alternative hypothesis is left-sided, so we find the area to the left of -6.685. The test statistic tells us that the mean difference was 6.685 standard errors below where the null hypothesis expected it to be. We should expect the p-value to be very small.

In fact, the area to the left of -6.685 is too small to see in the statistical calculator in Figure 9.29. But the calculator does verify that the p-value is nearly 0.



▶ **FIGURE 9.29** StatCrunch statistical calculator output, showing that the probability below −6.685 in a t-distribution with 39 degrees of freedom is extremely small.

Because the p-value is so small, we reject the null hypothesis and conclude that the typical subject in the study really did lose weight on The Zone diet.

EXAMPLE 17 More Rising College Costs

Earlier we questioned whether costs at two-year colleges increased by comparing a sample of schools in 2012–2013 with a population mean from the past. However, the data set that gave us the random sample of 35 schools also provides those same schools' tuition costs in 2008–2009. This means we can examine how each of these colleges changed its price. Here is the summary information:

Tuition & Fees	n	Sample Mean	Sample SD
2008–2009	35	\$2414	\$ 941
2012–2013	35	\$2919	\$1079

The observed value of the test statistic from a paired t-test was $t = 7.44$. The corresponding p-value is less than 0.001.

QUESTION Carry out a hypothesis test at a 5% significance level to test whether tuition and fees have increased at all two-year colleges from 2008–2009 through 2012–2013.

SOLUTION

Step 1: Hypothesize
Let μ_1 represent the mean tuition of all two-year colleges in 2012–2013, and let μ_2 represent the mean at all two-year colleges in 2008–2009. Then

$$H_0: \mu_{\text{difference}} = 0, \text{ where } \mu_{\text{difference}} = \mu_1 - \mu_2$$

$$H_a: \mu_{\text{difference}} > 0$$

Step 2: Prepare
Because these are a random sample, the first condition is satisfied. The same colleges appear in both groups, so we have paired data. The number of pairs (colleges) is greater than 35, so the Large Sample condition is satisfied. We can proceed.

Step 3: Calculate to Compare
The calculations were provided for us: $t = 7.44$ and the p-value is very small.

Step 4: Interpret
Because the p-value is less than the significance level, we reject the null hypothesis and conclude that the typical cost of two-year colleges has increased.

NOW TRY! Exercise 9.69

SNAPSHOT PAIRED t-TEST (DEPENDENT SAMPLES)

WHAT IS IT? ▶ A procedure for deciding whether two dependent (paired) samples have different means. Each pair is converted to a difference. The test statistic is the same as for the one-sample t-test, except that the null hypothesis value is 0:

$$t = \frac{\bar{x}_{\text{difference}} - 0}{SE_{\text{difference}}}$$

WHAT DOES IT DO? ▶ Lets us make decisions about whether the means are different, while knowing the probability that we are making a mistake.

HOW DOES IT DO IT? ▶ The test statistic compares the observed average difference, $\bar{x}_{\text{difference}}$, with the average difference we would expect if the means were the same: 0. Values far from 0 discredit the null hypothesis.

HOW IS IT USED? ▶ If the required conditions hold, the value of the observed test statistic can be compared to a t-distribution with $n - 1$ degrees of freedom.

SECTION 9.6

Overview of Analyzing Means

We hope you've been noticing a lot of repetition. The hypothesis test for two means is very similar to the test for one mean, and the hypothesis test for paired data is really a special case of the one-sample *t*-test. Also, the hypothesis tests use almost the same calculations as the confidence intervals, and they impose the same conditions, arranged slightly differently.

All the test statistics (for one proportion, for one mean, for two means, and for two proportions) have this structure:

$$\text{Test statistic} = \frac{(\text{estimated value}) - (\text{null hypothesis value})}{SE}$$

All the confidence intervals have this form:

$$\text{Estimated value} \pm (\text{multiplier})\, SE_{\text{EST}}$$

Not all confidence intervals used in statistics have this structure, but most that you will encounter do.

The method for computing a p-value is the same for all tests, although different distributions are used for different situations. The important point is to pay attention to the alternative hypothesis, which tells you whether you are finding a two-tailed or a one-tailed (and *which* tail) p-value.

Don't Accept the Null Hypothesis

If the p-value is larger than the significance level, then we do not reject the null hypothesis. But this doesn't mean we "accept" it. In other words, this doesn't mean we think the null hypothesis is true.

In Example 14 and the discussion that followed, we concluded that people tend to sleep longer on weekends and holidays than on weekdays—roughly an hour more on average. We reached this conclusion on the basis of a random sample of over 12,000 people. But let's consider what might have happened if we had taken a random sample of only 30 people from each group.

The following summary statistics are based on a random sample of just 30 people who reported their sleeping amounts on weekdays and another 30 people who reported their sleeping amounts on weekends and holidays.

$$\text{Weekday: } \bar{x} = 532 \text{ minutes, s} = 138 \text{ minutes}$$

$$\text{Weekend/holiday: } \bar{x} = 585 \text{ minutes, s} = 155 \text{ minutes}$$

If we now test the hypothesis that mean hours of sleep are different on weekends and holidays, we will get different results from our previous calculations. Now we find that our test statistic is $t = 1.40$ and the p-value is 0.167. If we believe the mean hours of sleep in the population are really the same, then the *t*-statistic value was not a surprise to us. And so we do not reject the null hypothesis.

But even though we do not reject the null, we do not necessarily believe it is true. The variation in amount of sleep is quite large: more than two hours. When our sample size is small, it is unlikely that we'll be able to tell whether two population means are truly different, because there is so much variability in our test statistic.

This is one reason why we never "accept" the null hypothesis. With a larger sample size, our test statistic will be more precise, and if the population means are truly different, we will have a better chance of correctly rejecting the null. In fact, when we did this test with more than 6000 people in each group, we determined that the means are different.

Confidence Intervals and Hypothesis Tests

If the alternative hypothesis is two-sided, then a confidence interval can be used instead of a hypothesis test. In fact, these two choices will always reach the same conclusion:

Choice 1: Perform the two-sided hypothesis test with significance level α.

Choice 2: Find a $(1 - \alpha) \times 100\%$ confidence interval (using methods given above). Reject the null hypothesis if the value for the null hypothesis does *not* appear in the interval.

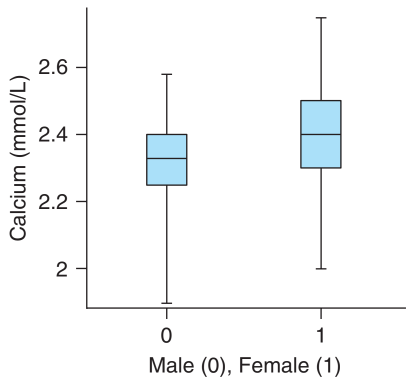
Confidence Level	Equivalent α (Two-Sided)
99%	0.01
95%	0.05
90%	0.10

▲ **TABLE 9.5** Equivalences between confidence intervals and tests with two-sided alternative hypotheses.

KEY POINT A 95% confidence interval is equivalent to a test with a two-sided alternative with a significance level of 0.05. Table 9.5 shows some other equivalences. All are true only for *two-sided* alternative hypotheses.

EXAMPLE 18 Calcium Levels in the Elderly

The boxplots in Figure 9.30 show the results of a study to determine whether calcium levels differ substantially between senior men and senior women (all older than 65 years). Calcium is associated with strong bones, and people with low calcium levels are believed to be more susceptible to bone fractures. The researchers carried out a hypothesis test to see whether the mean calcium levels for men and women were the same. Figure 9.31 shows the results. Calcium levels (the variable *cammol*) are measured in millimoles per liter (mmol/L).



◀ **FIGURE 9.30** Boxplots of calcium levels (mmol/L) for males and females.

Two sample T statistics with data

Options

Hypothesis test results:
μ_1 : mean of cammol where female=1
μ_2 : mean of cammol where female=0
$\mu_1 - \mu_2$: mean difference
$H_0 : \mu_1 - \mu_2 = 0$
$H_A : \mu_1 - \mu_2 \neq 0$
(without pooled variances)

Difference	Sample Mean	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	0.07570534	0.019790715	168.45009	3.825296	0.0002

◀ **FIGURE 9.31** StatCrunch output for testing whether mean calcium levels in men and women differ. The difference estimated is mean of females minus mean of males.

QUESTIONS

a. Assuming that all conditions necessary for carrying out *t*-tests and finding confidence intervals hold, what conclusion should the researchers make on the basis of this output? Use a significance level of 0.05.

b. Suppose the researchers calculate a confidence interval for the difference of the two means. Will this interval include the value 0? If not, will it include all negative values or all positive values? Explain.

SOLUTIONS

a. The p-value, 0.0002, is less than 0.05, so the researchers should reject the null hypothesis and conclude that men and women have different calcium levels.

b. Because we rejected the null hypothesis, we know that the confidence interval cannot include the value 0. If it did, then 0 would be a plausible difference between the means, and our hypothesis test concluded that 0 is not plausible. The estimated difference between the two means is, from the output, 0.0757. Because this value is positive, and because the interval cannot include 0, all values in the interval must be positive, showing a higher mean level of calcium in women than in men.

TRY THIS! Exercise 9.75

Hypothesis Test or Confidence Interval?

If you can use either a confidence interval or a hypothesis test, how do you choose? First of all, remember that these two techniques produce the same results only when you have a two-sided alternative hypothesis, so you need to make the choice only when you have a two-sided alternative.

These two approaches answer slightly different questions. The confidence interval answers the questions "What's the estimated value? And how much uncertainty do you have in this estimate?" Hypothesis tests are designed to answer the question "Is the parameter's value one thing, or another?"

For many situations, the confidence interval provides much more information than the hypothesis test. It not only tells us whether or not we should reject the null hypothesis but also gives us a plausible range for the population value. The hypothesis test, on the other hand, simply tells us whether to reject or not (although it does give us the p-value, which helps us see just how unusual our result is if the null hypothesis is true).

For example, in our hypothesis test of whether people tend to sleep a different amount on weekend and holidays than on weekdays, we rejected the null hypothesis and concluded that, on average, they did sleep different amounts. But that is all we can say with the hypothesis test. We cannot say, with the same significance level, whether they slept more on weekends or on weekdays, and we cannot say how much more. However, by finding that the 95% confidence interval was about 51 minutes to 61 minutes, we know now that, typically, people sleep about an hour longer on weekends and holidays than on weekdays. (Because group 1 was "weekend/holiday," the positive values mean that this mean was greater than the mean for the "weekday" group.) Because confidence intervals provide so much more information than hypothesis tests, there is a growing trend in scientific journals to require researchers to provide confidence intervals either in place of, or in support of, hypothesis tests.

CASE STUDY REVISITED

The researchers reported 95% confidence intervals for the mean IQ of three-year-old children whose mothers took one of four drugs for epilepsy, as shown in Table 9.6.

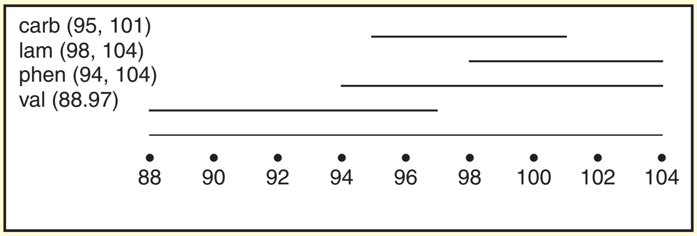
Drug	95% CI
Carbamazepine	(95, 101)
Lamotrigine	(98, 104)
Phenytoin	(94, 104)
Valproate	(88, 97)

▲ **TABLE 9.6** Confidence intervals for IQs.

The researchers could not observe all women on these drugs, so they based their observations on a random sample. If we think of these women as being sampled randomly, then the confidence intervals represent a range of plausible values for the mean IQ for the population of all three-year-olds whose mothers took these drugs.

It is helpful to display these confidence intervals graphically (as the researchers do in their paper), as shown in Figure 9.32.

▶ **FIGURE 9.32** Four confidence intervals for mean IQs of children from mothers taking different drugs for epilepsy.



From the figure, we see that the confidence interval for valproate does not overlap with that for lamotrigine. This suggests to us visually that it is *not* plausible that the mean IQ for children whose mothers took these drugs could be the same. The confidence interval for valproate has little overlap with the others, which makes us wonder how much different the mean IQs for the children of valproate users are from those for the children of the other mothers.

For this reason, we need to focus our attention on the differences between the means, not on the individual values for the means. If we do this, as the researchers did, we will find the confidence intervals shown in Table 9.7.

▶ **TABLE 9.7** Confidence intervals for differences in mean IQs.

Difference	95% CI
Carbamazepine – valproate	(0.6, 12.0)
Lamotrigine – valproate	(3.1, 14.6)
Phenytoin – valproate	(0.2, 14.0)

The first interval tells us that the difference between the mean IQ of those under carbamazepine and that of those under valproate could be as small as 0.6 IQ point or as large as 12.0 points. None of these intervals contains 0. This tells us that if we do a hypothesis test to determine whether the means are the same, we will have to reject the null hypothesis and conclude that the means are different.

EXPLORING STATISTICS
CLASS ACTIVITY

Pulse Rates

GOALS	MATERIALS
Learn to use a confidence interval (and/or a hypothesis test) to compare the means of two populations.	• A clock on the wall with a second hand or a watch with a second hand (for the instructor) • A computer or calculator

ACTIVITY

You will take your pulse rate before and after an activity to measure the effect of the activity on your pulse rate and/or to compare pulse rates between groups.

Try to find your pulse; it is usually easiest to find in the neck on one side or the other.

After everyone has found his or her pulse, your instructor will say, "Start counting," and you will count beats until the instructor says, "Stop counting." If your instructor uses a 30-second interval, double the count to get beats per minute.

Option A: Breathe in and out ten times, taking slow and deep breaths. Now measure your pulse rate again.

Option B: Stand up and sit down five times and then measure your pulse rate again.

Your instructor will collect these data and display the values (before and after each activity) for the class.

BEFORE THE ACTIVITY

1. Try finding your pulse (in your neck) to see how to do it.
2. Do you think that either activity (breathing slowly or standing up and sitting down) will change your heart rate? If so, by how much, and will it raise or lower it?
3. How would you measure the typical pulse rate of the class before and after the activity? How would you measure the change in pulse rate after each activity?
4. Do you think men and women have different mean heart rates before the activities?
5. Do you think the change in pulses will be different for men and women? (*Note:* If the class includes only one gender, your instructor may ask you to compare athletes to nonathletes or the taller half to the shorter half.)

AFTER THE ACTIVITY

1. State a pair of hypotheses (in words) for testing whether the breathing activity changes the mean pulse rate of the class. Do the same for the standing and sitting activity.
2. State a pair of hypotheses (in words) for whether men and women have the same mean resting pulse rate.
3. Calculate a 95% confidence interval for the change in pulse rates after the activity. What does this confidence interval tell us about the effect of the activity on the mean heart rate? Suppose you did a two-sided hypothesis test. On the basis of the confidence intervals, can you tell what the conclusions of the hypothesis test will be?

CHAPTER REVIEW

KEY TERMS

You may want to review the following terms, which were introduced in Chapters 7 and 8.

Chapter 7: statistic, estimator, bias, precision, sampling distribution, standard error, confidence interval, confidence level, margin of error

Chapter 8: null hypothesis, alternative hypothesis, significance level, test statistic, p-value, one-sided hypothesis, two-sided hypothesis

bias, *431*
precision, *431*
sampling distribution, *432*
unbiased estimator, *432*
standard error, *432*
Central Limit Theorem (CLT), *434*
t-statistic, *440*
t-distribution, *440*
degrees of freedom (df), *441*
confidence intervals, *442*
confidence level, *442*
independent samples, *458*
paired (dependent) samples, *458*

LEARNING OBJECTIVES

After reading this chapter and doing the assigned homework problems, you should

- Understand when the Central Limit Theorem for sample means applies and know how to use it to find approximate probabilities for sample means.
- Know how to test hypotheses concerning a population mean and concerning the comparison of two population means.
- Understand how to find, interpret, and use confidence intervals for a single population mean and for the difference of two population means.
- Understand the meaning of the p-value and of significance levels.
- Understand how to use a confidence interval to carry out a two-sided hypothesis test for a population mean or for a difference of two population means.

SUMMARY

The sample mean gives us an unbiased estimator of the population mean, provided that the observations are sampled randomly from a population and are independent of each other. The precision of this estimator, measured by the standard error (the standard deviation of the sampling distribution), improves as the sample size increases. If the population distribution is Normal, then the sampling distribution is Normal also. Otherwise, according to the Central Limit Theorem, the sampling distribution is approximately Normal, although for small sample sizes the approximation can be very bad. If the population distribution is not Normal, we recommend that you use a sample size of 25 or more.

Although we did not go into the (fairly complex) mathematics, our ability to measure confidence levels (which tell us how well confidence intervals perform), p-values, and significance levels (α) depends on the Central Limit Theorem (CLT). If the conditions for applying the CLT are not satisfied, then our reported values for these performance measures may be wrong.

Confidence intervals are used to provide estimates of parameters, along with a measure of our uncertainty in that estimate. The confidence intervals in this chapter differ only a little from those for proportions (Chapter 7). All are of the form

$$\text{Estimate} \pm \text{margin of error}$$

One thing that is different is that now you must decide whether your two samples are independent or paired before performing your analysis.

Another difference between confidence intervals for means and those for proportions is that the multiplier in the margin of error is based on the *t*-distribution, not on the Normal distribution.

By this point you have learned several different hypothesis tests, including the *z*-test for one-sample proportion and for two-sample proportions; and the t-test for one-sample mean, for two means from independent samples, and for two means from dependent samples. (You may also find pooled and unpooled versions of the two-mean independent-samples *t*-test, but you should always use the unpooled version.) It is important to learn which test to choose for the data you wish to analyze.

Hypothesis tests follow the structure described in Chapter 8, and, just as for confidence intervals, you must decide whether you have independent samples or paired samples. Tests of two means based on independent samples are based on the difference between the means. The null hypothesis is (almost) always that the difference is 0. The alternative hypothesis depends on the research question.

To find the p-value for a test of two means, use the *t*-distribution. To use the *t*-distribution, you must know the degrees of freedom (df), and this depends on whether you are doing a test for one mean ($df = n - 1$); two means from independent samples (use your computer, or, if working by hand, use the df for the smaller of $n_1 - 1$, $n_2 - 1$); or two means from paired data (number of pairs -1).

Formulas

Samples are selected randomly from each population and are independent. Population distributions are Normal, or if not, sample sizes need to be 25 or bigger for each sample.

Formula 9.1: One-Sample Confidence Interval for Mean

$$\bar{x} \pm m$$

where $m = t^*SE_{\text{EST}}$ and $SE_{\text{EST}} = \dfrac{s}{\sqrt{n}}$

The multiplier t^* is a constant that is used to fine-tune the margin of error so that it has the level of confidence we want. It is chosen on the basis of a *t*-distribution with $n - 1$ degrees of freedom. SE_{EST} is the estimated standard error.

Paired: $\bar{x}_{\text{difference}} \pm m$, where $m = t^*SE_{\text{EST}}$ and

$$SE_{\text{EST}} = \frac{s_{\text{difference}}}{\sqrt{n}}$$

(where $\bar{x}_{\text{difference}}$ is the average difference, $s_{\text{difference}}$ is the standard deviation of the differences, and n is the number of data pairs)

Formula 9.2: The One-Sample *t*-Test for Mean

$$t = \frac{\bar{x} - \mu}{SE_{\text{EST}}}, \text{ where } SE_{\text{EST}} = \frac{s}{\sqrt{n}} \text{ and, if conditions hold,}$$

t follows a t-distribution with df $= n - 1$

$$\textbf{Paired: } t = \frac{\bar{x}_{\text{difference}} - 0}{SE_{\text{difference}}}, \text{ where } SE_{\text{difference}} = \frac{s_{\text{difference}}}{\sqrt{n}}$$

(where $\bar{x}_{\text{difference}}$ is the average difference, $s_{\text{difference}}$ is the standard deviation of the differences, and n is the number of data pairs)

If conditions hold, t follows a t-distribution with degrees of freedom df $= n - 1$ (where n is the number of data pairs)

Formula 9.3: Two-Sample Confidence Interval

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

If conditions hold, t^* is based on a t-distribution. If no computer is available, the degrees of freedom are conservatively estimated as the smaller of $n_1 - 1$ and $n_2 - 1$.

Formula 9.4: Two-Sample *t*-Test (Unpooled)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{SE_{\text{EST}}}, \text{ where } SE_{\text{EST}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

If conditions hold, t follows an approximate t-distribution. If no computer is available, the degrees of freedom, df, are conservatively estimated as the smaller of $n_1 - 1$ and $n_2 - 1$

(Do not use the pooled version.)

SOURCES

Bureau of Labor Statistics. American Time Use Survey. http://www.bls.gov/tus/ (accessed January 2014).

Dansinger, M., J. Gleason, J. Griffith, H. Selker, and E. Schaefer. 2005. Comparison of the Atkins, Ornish, Weight Watchers, and Zone diets for weight loss and heart disease risk reduction: A randomized trial. *Journal of the American Medical Association* 293(1), 43–53.

Dunn, P. 2012. Assessing claims made by a pizza chain. *Journal of Statistics Education* 20(1). http://www.amstat.org/publications/jse/v20n1/dunn.pdf

Integrated Postsecondary Data System. http://nces.ed.gov/ipeds/ (accessed December 2013).

Kaplan, D. 2009, 2014. www.cherryblossum.org/results (accessed February 2014). Updated data scraped by authors in 2014.

Lundström, J., J. Boyle, and M. Jones-Gotman. 2006. Sit up and smell the roses better: Olfactory sensitivity to phenyl ethyl alcohol is dependent on body position. *Chemical Senses* 31(3), 249–252. doi: 10.1093/chemse/bjj025.

Mangen, A., B. Walgermo, and K. Bronnick. 2013. Reading linear texts on paper versus computer screens: Effects on reading comprehension. *International Journal of Educational Research* 58, 61–68. http://dx.doi.org/10.1016/j.ijer.2012.12.002

Meador, K. J., et al. 2009. Cognitive function at 3 years of age after fetal exposure to antileptic drugs. *New England Journal of Medicine* 360(16), 1597–1605.

National Health and Nutrition Examination Survey (NHANES). Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2003–2004.

SECTION EXERCISES

SECTION 9.1

9.1 Subjects A survey of 500 random full-time students at a college showed the mean number of subjects that the students had opted for was 7.9 with a standard deviation of 0.9 subjects.

a. Are these numbers statistics or parameters? Explain.

b. Label both numbers with their appropriate symbol (such as $\bar{x}$, μ, s, or σ).

9.2 Weight A study of all the employees at an office showed a mean weight of 60.4 kilograms and a standard deviation of 1.5 kilograms.

a. Are these numbers statistics or parameters? Explain.

b. Label both numbers with their appropriate symbol (such as $\bar{x}$, μ, s, or σ).

9.3 Exam Scores The distribution of the scores on a certain exam is $N(70, 10)$, which means that the exam scores are Normally distributed with a mean of 70 and standard deviation of 10.

a. Sketch the curve and label, on the x-axis, the position of the mean, the mean plus or minus one standard deviation, the mean plus or minus two standard deviations, and the mean plus or minus three standard deviations.

b. Find the probability that a randomly selected score will be bigger than 80. Shade the region under the Normal curve whose area corresponds to this probability.

9.4 Exam Scores The distribution of the scores on a certain exam is $N(70, 10)$, which means that the exam scores are Normally distributed with a mean of 70 and standard deviation of 10.

a. Sketch the curve and label, on the x-axis, the position of the mean, the mean plus or minus one standard deviation, the mean plus or minus two standard deviations, and the mean plus or minus three standard deviations.

b. Find the probability that a randomly selected score will be between 50 and 90. Shade the region under the Normal curve whose area corresponds to this probability.

9.5 Rodents A scientist is interested in studying the effects that applying pesticide to crops has on the local rodent population. The scientist collects 31 rodents from the field where the pesticide is used and finds the mean weight of this sample to be 84.6 grams. Assuming the selected rodents are a random sample, he concludes that because the sample mean is an unbiased estimator of the population mean, the mean weight of rodents in the population is also 84.6 grams. Explain why this is an incorrect interpretation of what it means to have an unbiased estimator.

9.6 Cellphone Calls Answers.com claims that the mean length of all cell phone conversations in the United States is 3.25 minutes (3 minutes and 15 seconds). Assume that this is correct, and also assume that the standard deviation is 4.2 minutes. (Source: wiki .answers.com, accessed January 16, 2011)

* a. Describe the shape of the distribution of the length of cell phone conversations in this population. Do you expect it to be approximately Normally distributed, right-skewed, or left-skewed? Explain your reasoning.
  b. Suppose that, using a phone company's records, we randomly sample 100 phone calls. We calculate the mean length from this sample and record the value. We repeat this thousands of times. What will be the (approximate) mean value of the distribution of these thousands of sample means?
  c. Refer to part b. What will be the standard deviation of this distribution of thousands of sample means?

9.7 Production Time A supervisor of a large factory takes a random sample of 100 laborers from the factory database. He calculates the mean time taken by them to produce one unit of the product. He records this value and repeats the process: He takes another random sample of 100 laborers and calculates the mean time taken. After he has done this 500 times, he makes a histogram of the mean time taken. Is this histogram a display of the population distribution, the distribution of a sample, or the sampling distribution of means?

9.8 Retirement Age From time to time, *The Telegraph* takes random samples from the UK population. One such survey is the Old Age Pension Survey. The most recent such survey, based on a large (several thousand) sample of randomly selected citizens, estimates the mean retirement age in the United Kingdom to be 64.7 years. Suppose we were to make a histogram of all of the retirement ages from this sample. Would the histogram be a display of the population distribution, the distribution of a sample, or the sampling distribution of means?

TRY **9.9 Sale of Air Conditioners (Example 1)** The average number of air conditioners sold in 2015 was 3600 per day in a city, and that was larger than the average for any other appliance. Suppose the standard deviation is 1404 and the distribution is right-skewed. Suppose we take a random sample of 81 days in the year.

a. What value should we expect for the sample mean? Why?
b. What is the standard error for the sample mean?

9.10 Sale of Microwaves The average number of microwaves sold in 2015 was 2700 per day in the same city, and that was larger than the average for any other appliance but less than that of the air conditioners. Suppose the standard deviation is 1551 and the distribution is right-skewed. Suppose we take a random sample of 121 days in the year.

a. What value should we expect for the sample mean? Why?
b. What is the standard error for the sample mean?

SECTION 9.2

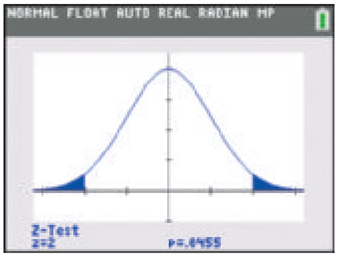
TRY **9.11 Babies Weights (Example 2)** Some sources report that the weights of full-term newborn babies have a mean of 7 pounds and a standard deviation of 0.6 pound and are Normally distributed.

a. What is the probability that one newborn baby will have a weight within 0.6 pound of the mean—that is, between 6.4 and 7.6 pounds, or within one standard deviation of the mean?
b. What is the probability the average of four babies' weights will be within 0.6 pound of the mean; will be between 6.4 and 7.6 pounds?
c. Explain the difference between a and b.

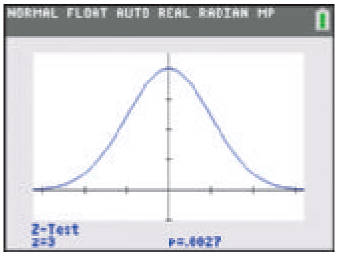
9.12 Babies' Weights, Again Some sources report that the weights of full-term newborn babies have a mean of 7 pounds and a standard deviation of 0.6 pound and are Normally distributed. In the given outputs, the shaded areas (reported as p=) represent the probability that the mean will be larger than 7.6 or smaller than 6.4. One of the outputs uses a sample size of 4, and one uses a sample size of 9.

a. Which is which, and how do you know?
b. These graphs are made so that they spread out to occupy the room on the face of the calculator. If they had the same horizontal axis, one would be taller and narrower than the other. Which one would that be, and why?

(A)



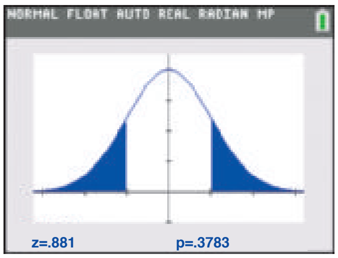
(B)



TRY **9.13 Sale of Air Conditioners (Example 3)** The average number of air conditioners sold in 2015 in a city was 3600 per day. Suppose the standard deviation is 1404 and the distribution is right-skewed. Suppose we take a random sample of 81 days in the year.

a. Is the sample size large enough to use the Central Limit Theorem for means? Explain.
b. What are the mean and standard error of the sampling distribution? Refer to Exercise 9.9.
c. What is the probability that the sample mean will be more than 2808 units away from the population mean?

* **9.14 Sale of Microwaves** The average number of microwaves sold in the city in 2015 was 2700 per day. Suppose the standard deviation is 1551 and the distribution is right-skewed. Suppose we take a random sample of 121 days in the year. We want to find the probability that the sample mean will be more than 141 units away from the population mean. The TI-84 output is shown.



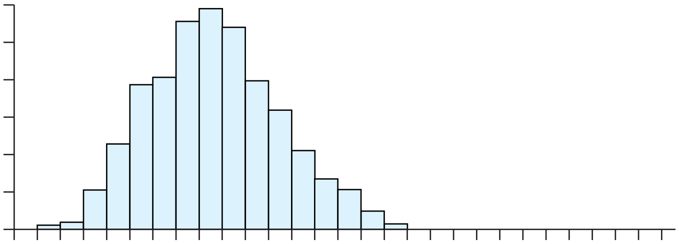
a. Why is the distribution Normal and not right-skewed like the population?

b. Why is the z-score 1?

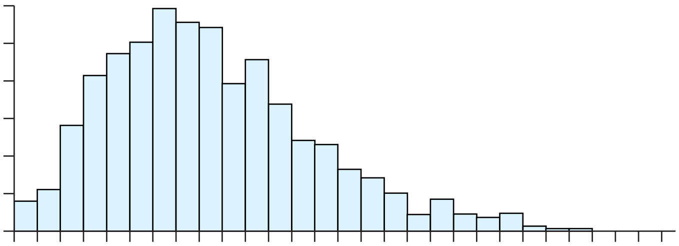
c. What is the probability that the sample mean will be more than 141 units away from the population mean?

TRY **9.15 CLT Shapes (Example 4)** One of the histograms is a histogram of a sample (from a population with a skewed distribution) one is the distribution of many means of repeated random samples of size 5, and one is the distribution of repeated means of random samples of size 25; all the samples are from the same population. The scale is the same across all histograms. State which is which and how you know.

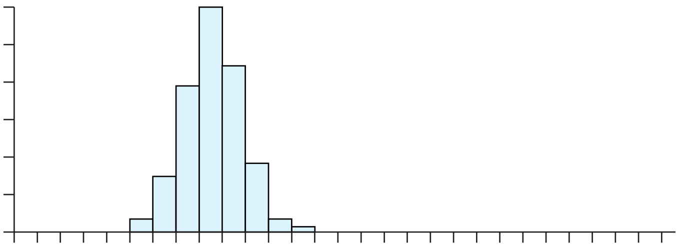
(A)



(B)

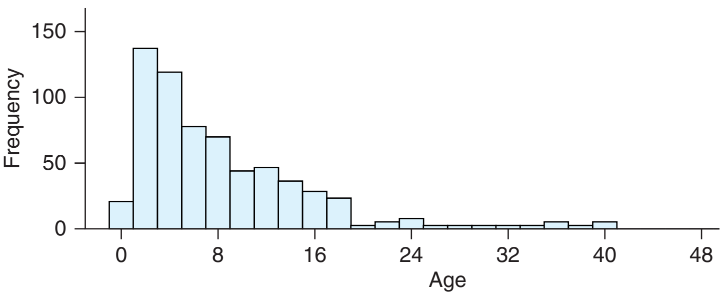


(C)

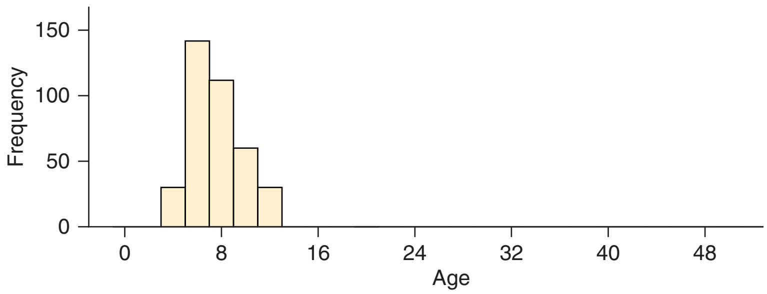


9.16 Used Car Ages One histogram shows the distribution of ages of all 638 used cars for sale in the *Ventura County Star* Sunday newspaper in 2013. The other three graphs show distributions of means from random samples taken from the same population of used cars. One histogram shows means based on samples of 2 cars, another shows means based on samples of 5 cars, and another shows means based on samples of 10 cars. Each graph based on means was done with many repetitions. Which distribution is which, and why?

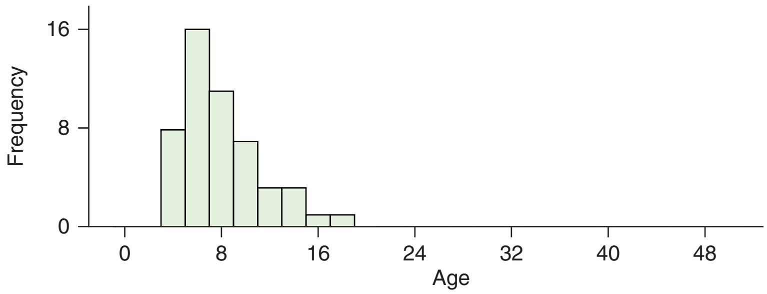
(A)



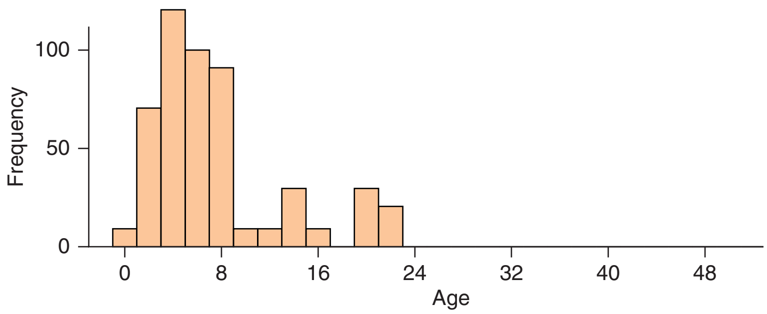
(B)



(C)



(D)



TRY **9.17 Used Car Ages (Example 5)** The mean age of all 638 used cars for sale in the *Ventura Country Star* one Saturday in 2013 was 7.9 years, with a standard deviation of 7.7 years. The distribution of ages is right-skewed. For a study to determine the reliability of classified ads, a reporter randomly selects 40 of these used cars and plans to visit each owner to inspect the cars. He finds that the mean age of the 40 cars he samples is 8.2 years and the standard deviation of those 40 cars is 6.0 years.

a. Which of these four numerical values are parameters and which are statistics?

b. $\mu = ?$ $\sigma = ?$ $s = ?$ $\bar{x} = ?$

c. Are the conditions for using the CLT fulfilled? What would be the shape of the approximate sampling distribution of a large number of means, each from a sample of 40 cars?

9.18 Student Heights The mean height of all 1800 fifth-grade students in a small school is 128 cm with a standard deviation of 16 cm, and the distribution is right-skewed. A random sample of 5 students' heights is obtained, and the mean is 124 with a standard deviation of 12 cm.

a. $\mu = ?$ $\sigma = ?$ $\bar{x} = ?$ $s = ?$

b. Is μ a parameter or a static?

c. Are the conditions for using the CLT fulfilled? What would be the shape of the approximate sampling distribution of many means, each from a sample of 5 students? Would the shape be right-skewed, Normal, or left-skewed?

SECTION 9.3

TRY **9.19 Four-year Graduation Rate (Example 6)** A random sample of 10 colleges from Kiplinger's 100 Best Values in Public Education was taken. The mean rate of graduation within four years was 43.5% with a margin of error of 6.0%. The distribution of graduation rates is Normal. (Source: http://portal.kiplinger.com/tool/college/T014-S001-kiplinger-s-best-values-in-public-colleges/index.php#colleges. Accessed via StatCrunch. Owner: Webster West.)

a. Decide whether each of the following statements is worded correctly for the confidence interval, and fill in the blanks for the correctly worded one(s).

i. We are 95% confident that the sample mean is between ___% and ___%.

ii. We are 95% confident that the population mean is between __% and ___%.

iii. There is a 95% probability that the population mean is between __ % and ___%.

b. Can we reject a population mean percentage of 50% on the basis of these numbers? Explain.

9.20 Spread of Diabetes A random sample of 100 people from different age groups was taken. The mean age of diabetic patients was 31.32 years with a margin of error of 17.4 years. The distribution of age is normal. (Source: https://www.StatCrunch.com/5.0/viewreport.php?reportid=60000. Accessed via StatCrunch. Owner: pachecodl79.)

a. Decide whether each of the following statements is worded correctly for the confidence interval, and fill in the blanks for the correctly worded one(s).

i. We are 95% confident that the boundaries of the interval are __ and ___.

ii. We are 95% confident that the population mean is between ___ and ___.

iii. We are 95% confident that the sample mean is between __ and ___.

b. Can we reject a population mean of 27 years on the basis of these numbers? Explain.

9.21 Oranges A statistics instructor randomly selected four bags of oranges, each bag labeled 10 pounds, and weighed the bags. They weighed 10.2, 10.5, 10.3, and 10.3 pounds. Assume that the distribution of weights is Normal. Find a 95% confidence interval for the mean weight of all bags of oranges. Use technology for your calculations.

a. Decide whether each of the following three statements is a correctly worded interpretation of the confidence interval, and fill in the blanks for the correct option(s).

i. I am 95% confident that the population mean is between ____ and ____.

ii. There is a 95% chance that all intervals will be between ___ and ___.

iii. I am 95% confident that the sample mean is between ___ and ___.

b. Does the interval capture 10 pounds? Is there enough evidence to reject the null hypothesis that the population mean weight is 10 pounds? Explain your answer.

9.22 Carrots The weights of four randomly chosen bags of horse carrots, each bag labeled 20 pounds, were 20.5, 19.8, 20.8, and 20.0 pounds. Assume that the distribution of weights is Normal. Find a 95% confidence interval for the mean weight of all bags of horse carrots. Use technology for your calculations.

a. Decide whether each of the following three statements is a correctly worded interpretation of the confidence interval, and fill in the blanks for the correct option(s).

i. 95% of all sample means based on samples of the same size will be between ___ and ____.

ii. I am 95% confident that the population mean is between ____ and ____.

iii. We are 95% confident that the boundaries are ____ and ____ .

b. Can you reject a population mean of 20 pounds? Explain.

TRY **9.23 College Admission Rates (Example 7)** A random sample of 10 colleges from Kiplinger's 100 Best Values in Public Education was taken. A 95% confidence interval for the mean admission rate was (52.8%, 75.0%). The rates of admission were Normally distributed. Which of the following statements is a correct interpretation of the confidence *level,* and which is the correct interpretation of the confidence *interval*? (Source: http://portal.kiplinger.com/tool/college/T014-S001-kiplinger-s-best-values-in-public-colleges/index.php#colleges. Accessed via StatCrunch. Owner: Webster West.)

a. We are confident that the mean admission rate is between 52.8% and 75.0%.

b. In about 95% of all samples of 10 colleges, the confidence interval will contain the population mean admission rate.

9.24 Even-numbered Digits If you take samples of 25 lines from a C-value table and find that the confidence interval for the proportion of even-numbered digits captures 95% 21 times out of the 25 lines, is it the confidence *interval* or confidence *level* you are estimating with the 21 out of 25?

TRY **9.25 t^* (Example 8)** A researcher collects one sample of 27 measurements from a population and wants to find a 95% confidence interval. What value should he use for t^*? (Recall that df $= n - 1$ for a one-sample t-interval.)

	C-level		
df	**90%**	**95%**	**99%**
24	1.711	2.064	2.797
25	1.708	2.060	2.787
26	1.706	2.056	2.779
27	1.703	2.052	2.771
28	1.701	2.048	2.763

	C-level		
df	**90%**	**95%**	**99%**
28	1.701	2.048	2.763
29	1.699	2.045	2.756
30	1.697	2.042	2.750
34	1.691	2.032	2.728

9.26 t^* A researcher collects a sample of 25 measurements from a population and wants to find a 99% confidence interval.

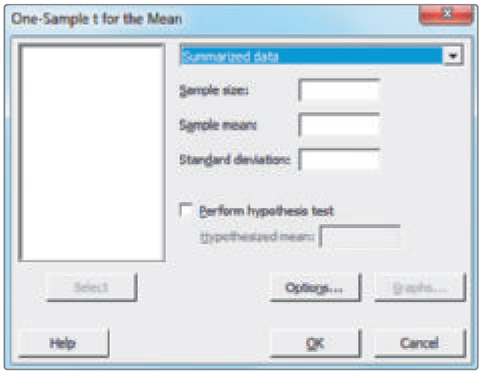
a. What value should he use for t^*? (Recall that $df = n - 1$ for a one-sample t-interval.) Use the table given for Exercise 9.25.

* b. Why is the answer to this question larger than the answer to Exercise 9.25?

TRY **9.27 Hamburgers (Example 9)** A hamburger chain sells large hamburgers. When we take a sample of 30 hamburgers and weigh them, we find that the mean is 0.51 pounds and the standard deviation is 0.2 pound.

a. State how you would fill in the numbers below to do the calculation with Minitab.

b. Report the confidence interval in a carefully worded sentence.



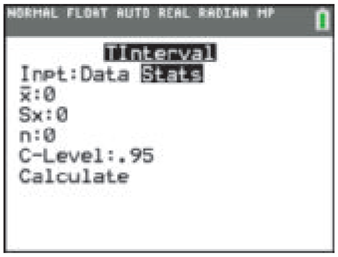
One-Sample T

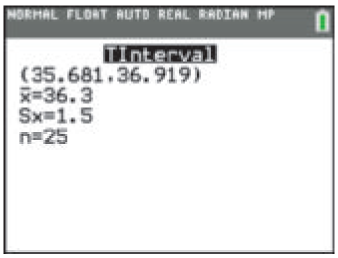
N	Mean	StDev	SE Mean	95% CI
30	0.5100	0.2000	0.0365	(0.4353, 0.5849)

9.28 Drinks A fast-food chain sells drinks that they call HUGE. When we take a sample of 25 drinks and weigh them, we find that the mean is 36.3 ounces with a standard deviation of 1.5 ounces.

a. State how you would fill in the numbers below to do the calculation with a TI-84.

b. Report the confidence interval in a carefully worded sentence.





TRY **9.29 Men's Pulse Rates (Example 10)** A random sample of 25 men's resting pulse rates shows a mean of 72 beats per minute and a standard deviation of 13.

a. Find a 95% confidence interval for the population mean pulse rate for men, and report it in a sentence. You may use the table given for Exercise 9.25.

b. Find a 99% confidence interval.

c. Which interval is wider and why?

9.30 Number of Children A random sample of 100 women from the General Social Survey showed that the mean number of children reported was 1.85 with a standard deviation of 1.5. (Interestingly, a sample of 100 men showed a mean of 1.49 children.)

a. Find a 95% confidence interval for the population mean number of children for women. Because the sample size is so large, you can use 1.96 for the critical value of t (which is the same as the critical value of z) if you do the calculations manually.

b. Find a 90% confidence interval. Use 1.645 for the critical value of t, which is the critical value of z.

c. Which interval is wider, and why?

TRY **9.31 GPAs (Example 11)** In finding a confidence interval for a random sample of 30 students GPAs, one interval was (2.60, 3.20) and the other was (2.65, 3.15).

a. One of them is a 95% interval and one is a 90% interval. Which is which, and how do you know?

b. If we used a larger sample size ($n = 120$ instead of $n = 30$), would the 95% interval be wider or narrower than the one reported here?

9.32 Confidence A student of statistics was given two intervals for the same data, one for 75% confidence and one for 90% confidence.

a. How would a 70% interval compare? Would it be narrower than both, wider than both, or between the two in width? Explain.

b. If we wanted to use a 70% confidence level and get a broader width, how could we change our data collection?

9.33 Confidence Interval Changes State whether each of the following changes would make a confidence interval wider or narrower. (Assume that nothing else changes.)

a. Changing from a 90% confidence level to a 99% confidence level.

b. Changing from a sample size of 30 to a sample size of 200.

c. Changing from a standard deviation of 20 pounds to a standard deviation of 25 pounds.

9.34 Changes in Confidence Interval State whether each of the following changes would make a confidence interval wider or narrower. (Assume that nothing else changes.)

a. Changing from an 80% level of confidence to an 85% level of confidence.

b. Changing from a sample size of 25 to a sample size of 40.

c. Changing from a standard deviation of 500 grams to a standard deviation of 750 grams.

9.35 Players The heights of four randomly and independently selected baseball players were found to be 196 cm, 198 cm, 193 cm, and 175 cm. Assume Normality.

a. Find a 95% confidence interval for the mean height of all players of the baseball team.

b. Does the interval capture 170 cm? Is there enough evidence to reject a mean height of 170 cm?

9.36 Tomatoes The weights of four randomly and independently selected bags of tomatoes labeled 5 pounds were found to be 5.1, 5.0, 5.3, and 5.1 pounds. Assume Normality.

a. Find a 95% confidence interval for the mean weight of all bags of tomatoes.

b. Does the interval capture 5.0 pounds? Is there enough evidence to reject a mean weight of 5.0 pounds?

SECTION 9.4

TRY g **9.37 Human Body Temperatures (Example 12)** A random sample of 10 independent healthy people showed the following body temperatures (in degrees Fahrenheit):

98.5, 98.2 99.0, 96.3, 98.3, 98.7, 97.2, 99.1, 98.7, 97.2

Test the hypothesis that the population mean is not 98.6°F, using a significance level of 0.05. *See page 492 for guidance.*

9.38 Reaction Distance Data on the text's website show reaction distances in centimeters for the dominant hand for a random sample of 40 independently chosen college students. Smaller distances indicate quicker reactions.

a. Make a graph of the distribution of the sample, and describe its shape.

b. Find, report, and interpret a 95% confidence interval for the population mean.

c. Suppose a professor said that the population mean should be 10 centimeters. Test the hypothesis that the population mean is not 10 cm, using the four-step procedure, with a significance level of 0.05.

9.39 Potatoes Use the data from Exercise 9.35.

a. If you use the four-step procedure with a two-sided alternative hypothesis, should you be able to reject the hypothesis that the population mean is 20 pounds using a significance level of 0.05? Why or why not? The confidence interval is reported here: I am 95% confident that the population mean is between 20.4 and 21.7 pounds.

b. Now test the hypothesis that the population mean is not 20 pounds using the four-step procedure. Use a significance level of 0.05.

c. Choose one of the following conclusions:

i. We cannot reject a population mean of 20 pounds.

ii. We can reject a population mean of 20 pounds.

iii. The population mean is 21.05 pounds.

9.40 Tomatoes Use the data from Exercise 9.36.

a. Using the four-step procedure with a two-sided alternative hypothesis, should you be able to reject the hypothesis that the population mean is 5 pounds using a significance level of 0.05? Why or why not? The confidence interval is reported here: I am 95% confident the population mean is between 4.9 and 5.3 pounds.

b. Now test the hypothesis that the population mean is not 5 pounds using the four step procedure. Use a significance level of 0.05 and number your steps.

9.41 Frequency In a music class, it was suggested that the frequency of an ideal guitar chord should be 2.81 Hz or less. Inspection of four randomly and independently selected guitars from the class showed the results in the Minitab output given.

Minitab: One-Sample T

Test of μ = 64.5 vs < 64.5

N	Mean	StDev	SE Mean	T	P
30	3.2497	0.4903	0.0895	4.91	0.000

Test the hypothesis that the mean guitar frequency is more than 2.81 using a significance level of 0.05. Assume that conditions are met.

9.42 BMI A body mass index of more than 25 is considered unhealthful. The Minitab output given is from 50 randomly and independently selected people from the NHANES study.

One-Sample T

Test of μ = 25 vs > 25

N	Mean	StDev	SE Mean	T	P
50	27.874	6.783	0.959	3.00	0.002

Test the hypothesis that the mean BMI is more than 25 using a significance level of 0.05. Assume that conditions are met.

9.43 Male Height In the United States, the population mean height for 3-year-old boys is 38 inches (http://www.kidsgrowth.com). Suppose a random sample of 15 non-U.S. 3-year-old boys showed a sample mean of 37.2 inches with a standard deviation of 3 inches. The boys were independently sampled. Assume that heights are Normally distributed in the population.

a. Determine whether the population mean for non-U.S. boys is significantly different from the U.S. population mean. Use a significance level of 0.05.

b. Now suppose the sample consists of 30 boys instead of 15, and repeat the test.

c. Explain why the *t*-values and p-values for parts a and b are different.

9.44 Vegetarians' Weights The mean weight of all 20-year-old women is 128 pounds (http://www.kidsgrowth.com). A random sample of 40 vegetarian women who are 20 years old showed a sample mean of 122 pounds with a standard deviation of 15 pounds. The women's measurements were independent of each other.

a. Determine whether the mean weight for 20-year old vegetarian women is significantly less than 128, using a significance level of 0.05.

b. Now suppose the sample consists of 100 vegetarian women who are 20 years old, and repeat the test.

c. Explain what causes the difference between the p-values for parts a and b.

9.45 GPAs Thirty GPAs from a randomly selected sample of statistics students at Oxnard College are available at this text's website. Assume that the population distribution is approximately Normal. The technician in charge of records claimed that the population mean GPA for the whole college is 2.81.

GPAs	GPAs	GPAs	GPAs
3.46	2.61	3.2	3.8
2.91	3	3.87	3.75
2.5	4	2.99	3.6
3.19	2.72	2.35	2.6
3.12	4	3.32	2.89
3.75	2.74	2.83	3.25
3	3.94	3.1	
3.9	3.5	3.6	

a. What is the sample mean? Is it higher or lower than the population mean of 2.81?

b. The chair of the mathematics department claims that statistics students typically have higher GPAs than the typical college student. Use the four-step procedure and the data provided to test this claim. Use a significance level of 0.05.

9.46 Dancers' Heights A random sample of 20 independent female college-aged dancers was obtained, and their heights (in inches) were measured. Assume the population distribution is Normal.

Dancers' Heights	Dancers' Heights	Dancers' Heights	Dancers' Heights
60	62	63	62
61	60	62	65
64	60.5	64.5	66
62	67	63.5	63
64	66	63	67

a. What is the sample mean? Is it above or below 64.5 inches?

b. Some people claim that the physical demands on dancers are such that dancers tend to be shorter than the typical person in the population. Use the four-step procedure to test the hypothesis that dancers have a smaller population mean height than 64.5 inches. Use a significance level of 0.05.

9.47 GPAs Using the data from Exercise 9.45 on GPAs, find a 95% confidence interval for the mean GPA. Also, if you had used a two-sided alternative (instead of the one-sided alternative in Exercise 9.45) and had done a test with a significance level of 0.05, would you have rejected a hypothesized mean GPA of 2.81?

9.48 Dancers' Heights Using the data from Exercise 9.46 on dancers' heights, find a 95% confidence interval for the mean height. Also, if you had used a two-sided alternative (Instead of the one-sided alternative used in Exercise 9.46) and had done a test with a significance level of 0.05, would you have rejected a hypothesized mean of 64.5 inches?

9.49 Atkins Diet Difference Ten randomly selected people went on a Atkins diet for a month. The weight losses experienced (in pounds) were

3, 8, 10, 0, 4, 6, 6, 4, 2, and -2

The negative weight loss is a weight gain. Test the hypothesis that the mean weight loss was more than 0, using a significance level of 0.05. Assume the population distribution is Normal.

9.50 Pulse Difference The following numbers are the differences in pulse rate (beats per minute) before and after running for 12 randomly selected people.

24, 12, 14, 12, 16, 10, 0, 4, 13, 42, 4, and 16

Positive numbers mean the pulse rate went up. Test the hypothesis that the mean difference in pulse rate was more than 0, using a significance level of 0.05. Assume the population distribution is Normal.

9.51 Employee Salaries Suppose that 500 employees each took a random sample (with replacement) of 100 employees at their office and recorded the salaries of the employees in their sample. Then each employee used his or her data to calculate an 80% confidence interval for the mean salary of all employees at the office. How many of the 500 intervals would you expect not to capture the true population mean? Explain by showing your calculation.

9.52 Concentrations A 95% confidence interval for the global concentration of carbon monoxide is (0–0.06 mg/m^3). Either interpret the interval or explain why it should not be interpreted.

SECTION 9.5

TRY **9.53 Independent or Paired (Example 13)** State whether each situation has independent or paired (dependent) samples.

a. A researcher wants to know whether pulse rates of people go down after brief meditation. She collects the pulse rates of a random sample of people before meditation and then collects their pulse rates after meditation.

b. A researcher wants to know whether women who text send more text messages than men who text. She gathers two random samples, one from men and one from women, and asks them how many text messages they sent yesterday.

9.54 Independent or Paired State whether each situation has independent or paired (dependent) samples.

a. A researcher wants to understand whether the brush strokes of an art student are more consistent than those of an amateur artist. She collects the data from a random sample of art students and amateur artists.

b. A researcher wants to know whether the wealth gathered by a person is directly related to the income level of the person. She surveys and collects data on the net worth of a random sample of people in the high tax bracket and the low tax bracket.

TRY **9.55 Televisions: CI (Example 14)** Minitab output is shown for a two-sample t-interval for the number of televisions owned in households of random samples of students at two different community colleges. Each individual was randomly chosen independently of the others. One of the schools is in a wealthy community (MC), and the other (OC) is in a less wealthy community.

Two-Sample T: CI

```
Sample       N      Mean      StDev      SE Mean
OCTVs       30      3.70       1.49         0.27
MCTVs       30      3.33       1.49         0.27

Difference = μ(1) - μ(2)
Estimate for difference:  0.370
95% CI for difference:  (-0.400, 1.140)
```

a. Are the conditions for using a confidence interval for the difference between two means met?

b. State the interval in a clear and correct sentence.

c. Does the interval capture 0? Explain what that shows.

9.56 Pulse and Gender: CI Using data from NHANES, we looked at the pulse rate for nearly 800 people to see whether it is plausible that men and women have the same population mean. NHANES data are random and independent. Minitab output follows.

Two-Sample T: CI

```
Sample       N      Mean      StDev      SE Mean
Women      384      76.3       12.8         0.65
Men        372      72.1       13.0         0.67

Difference = μ(1) - μ(2)
Estimate for difference: 4.200
95% CI for difference:  (2.357, 6.043)
```

a. Are the conditions for using a confidence interval for the difference between two means met?

b. State the interval in a clear and correct sentence.

c. Does the interval capture 0? Explain what that shows.

TRY **9.57 Televisions (Example 15)** The table shows the Minitab output for a two-sample *t*-test for the number of televisions owned in households of random samples of students at two different community colleges. Each individual was randomly chosen independently of the others; the students were not chosen as pairs or in groups. One of the schools is in a wealthy community (MC), and the other (OC) is in a less wealthy community. Test the hypothesis that the population means are not the same, using a significance level of 0.05. *See page 492 for guidance.*

Two-Sample T-Test and CI: OCTV, MCTV

```
Two-sample T for OCTV cs MCTV

           N      Mean      StDev     SE Mean
OCTV      30      3.70       1.49        0.27
MCTV      30      3.33       1.49        0.27

Difference = μ(OCTV) − μ(MCTV)
Estimate for difference: 0.367
95% CI for difference: (−0.404, 1.138)
T-Test of difference = 0 (vs ≠): T-Value = 0.95
                                 P-Value = 0.345
```

9.58 Pulse Rates Using data from NHANES, we looked at pulse rates of nearly 800 people to see whether men or women tended to have higher pulse rates. Refer to the Minitab output provided.

a. Report the sample means, and state which group had the higher sample mean pulse rate.

b. Use the Minitab output to test the hypothesis that pulse rates for men and women are not equal, using a significance level of 0.05. The samples are large enough so that Normality is not an issue.

Two-Sample T-Test and CI: Pulse, Sex

```
Two-sample T for Pulse

Sex          N      Mean      StDev     SE Mean
Female     384      76.3       12.8        0.65
Male       372      72.1       13.0        0.67

Difference = μ(Female) − μ(Male)
Estimate for difference: 4.248
95% CI for difference:  (2.406, 6.090)
T-Test of difference = 0 (vs ≠): T-Value = 4.53
                                 P-Value = 0.000
                                 DF = 752
```

9.59 Triglycerides Triglycerides are a form of fat found in the body. Using data from NHANES, we looked at whether men have higher triglyceride levels than women.

a. Report the sample means, and state which group had the higher sample mean triglyceride level. Refer to the Minitab output in figure (A).

b. Carry out a hypothesis test to determine whether men have a higher mean triglyceride level than women. Refer to the Minitab output provided in figure (A). Output for three different alternative hypotheses is provided—see figures (B), (C), and (D)—and you must choose and state the most appropriate output. Use a significance level of 0.05.

(A)

```
Two-Sample T-Test and CI:
Triglycerides, Gender

Two-sample T for Triglycerides

Gender    N    Mean   StDev   SE Mean
Female   44    84.4    40.2       6.1
Male     48   139.5    85.3        12

Difference = μ(Female) - μ(Male)
Estimate for difference: -55.1
95% CI for difference: (-82.5, -27.7)
```

(B)

```
9.49 B: T-Test of difference = 0 (vs <):
                               T-Value = -4.02
                               P-Value = 0.000
```

(C)

```
9.49 C: T-Test of difference = 0 (vs >):
                               T-Value = -4.02
                               P-Value = 1.000
```

(D)

```
9.49 D: T-Test of difference = 0 (vs ≠):
                               T-Value = -4.02
                               P-Value = 0.000
```

9.60 Systolic Blood Pressures When you have your blood pressure taken, the larger number is the systolic blood pressure. Using data from NHANES, we looked at whether men and women have different systolic blood pressure levels.

a. Report the two sample means, and state which group had the higher sample mean systolic blood pressure. Refer to the Minitab output in figure (A).

b. Refer to the Minitab output given in figure (A) to test the hypothesis that the mean systolic blood pressures for men and women are not equal, using a significance level of 0.05. Although the distribution blood pressures in the population are right-skewed, the sample size is large enough for us to use *t*-tests. Choose from figures (B), (C), and (D) for your t and p-values, and explain.

(A)

```
Two-sample T for BPSys

Gender     N    Mean   StDev   SE Mean
Female   404   116.8    22.7       1.1
Male     410   118.7    18.0      0.89

Difference = μ(Female) - μ(Male)
Estimate for difference: -1.93
95% CI for difference: (-4.75, 0.89)
```

(B)

```
9.50 B: T-Test of difference = 0 (vs ≠):
                               T-Value = -1.34
                               P-Value = 0.180
```

(C)

```
9.50 C: T-Test of difference = 0 (vs >): T-Value = -1.34  P-Value = 0.910
```

(D)

```
9.50 D: T-Test of difference = 0 (vs <): T-Value = -1.34   P-Value = 0.090
```

9.61 Triglycerides, Again Report and interpret the 95% confidence interval for the difference in mean triglyceride level for men and women (refer to the Minitab output in Exercise 9.59). Does this support the hypothesis that men and women differ in mean triglyceride level? Explain.

9.62 Blood Pressures, Again Report and interpret the 95% confidence interval for the difference in mean systolic blood pressure for men and women (refer to the Minitab output in Exercise 9.60). Does this support the hypothesis that men and women differ in mean systolic blood pressure? Explain.

9.63 Clothes Spending A random sample of 14 college women and a random sample of 19 college men were separately asked to estimate how much they spent on clothing in the last month. The table shows the data.

Test the hypothesis that the population mean amounts spent on clothes are different for men and women. Use a significance level of 0.05. Assume that the distributions are Normal enough for us to use the *t*-test.

Sex	moneyforclothes	Sex	moneyforclothes
m	175	f	80
f	200	m	200
m	150	m	80
f	200	m	100
f	100	m	120
f	100	m	80
f	200	m	25
m	100	f	80
m	100	m	50
f	200	m	100
m	200	m	30
m	200	f	20
m	200	f	50
f	250	m	60
f	150	f	100
m	100	f	350
m	0		

9.64 College Athletes' Weights A random sample of male college baseball players and a random sample of male college soccer players were obtained independently and weighed. The table shows the unstacked weights (in pounds). The distributions of both data sets suggest that the population distributions are roughly Normal. Determine whether the difference in means is significant, using a significance level of 0.05.

Baseball	Soccer	Baseball	Soccer
190	165	186	156
200	190	210	168
187	185	198	173
182	187	180	158
192	183	182	150
205	189	193	172
185	170	200	180
177	182	195	184
207	172	182	174
225	180	193	190
230	167	190	156
195	190	186	163
169	185		

* **9.65 Clothes Spending** In Exercise 9.63 you could not reject the null hypothesis that the mean amount spent by men and the mean amount spent by women for clothing are the same, using a two-tailed test with a significance level of 0.05.

a. If you found a 95% confidence interval for the difference between means, would it capture 0? Explain.

b. If you found a 99% confidence interval, would it capture 0? Explain.

c. Now go back to Exercise 9.63. Find a 95% confidence interval for the difference between means, and explain what it shows.

* **9.66 College Athletes' Weights** In Exercise 9.64, you could reject the null hypothesis that the mean weights of soccer and baseball players were equal using a two-tailed test with a significance level of 0.05.

a. If you found a 95% confidence interval for the difference between means, would it capture 0? Explain.

b. If you found a 90% interval, would it capture 0? Explain.

c. Now go back to Exercise 9.64. Find a 95% confidence interval for the difference between means, and explain what it shows.

TRY **9.67 Textbook Prices, UCSB vs. CSUN (Example 16)** The prices of a sample of books at University of California at Santa Barbara (UCSB) were obtained by statistics students Ricky Hernandez and Elizabeth Alamillo. Then the cost of books for the same subjects (at the same level) were obtained for California State University at Northridge (CSUN). Assume that the distribution of differences is Normal enough to proceed, and assume that the sampling was random. The data are at this text's website.

a. First find both sample means and compare them.

b. Test the hypothesis that the population means are different, using a significance level of 0.05.

9.68 Textbook Prices. OC vs. CSUN The prices of a random sample of comparable (matched) textbooks from two schools were recorded. We are comparing the prices at OC (Oxnard Community College) and CSUN (California State University at Northridge). Assume that the population distribution of differences is approximately Normal. Each book was priced separately; there were no books "bundled" together. The data are at this text's website.

a. Compare the sample means.

b. Determine whether the mean prices of all books are significantly different. Use a significance level of 0.05.

9.69 Females–Pulse Rates before and after a Fright (Example 17) In a statistics class taught by one of the authors, students took their pulses before and after being frightened. The frightening event was having the teacher scream and run from one side of the room to the other. The pulse rates (beats per minute) of the women before and after the scream were obtained separately and are shown in the table. Treat this as though it were a random sample of female community college students. Test the hypothesis that the mean of college women's pulse rates is higher after a fright, using a significance level of 0.05. *See page 493 for guidance*.

Women	
Pulse Before	**Pulse After**
64	68
100	112
80	84
60	68
92	104
80	92
68	72

Women	
Pulse Before	**Pulse After**
84	88
80	80
68	92
60	76
68	72
68	80

9.70 Males–Pulse Rates before and after a Fright Follow the instructions for Exercise 9.69, but use the data for the men in the class. Test the hypothesis that the mean of college men's pulse rates is higher after a fright, using a significance level of 0.05.

Men	
Pulse Before	**Pulse After**
50	64
84	72
96	88
80	72
80	88

Men	
Pulse Before	**Pulse After**
64	68
88	100
84	80
76	80

9.71 Organic Food A student compared organic food prices at Target and Whole Foods. The same items were priced at each store. The first three items are shown in Figure A. (Source: StatCrunch Organic food price comparison fall 2011. Owner: kerrypaulson) Choose the correct output (B or C) for the appropriate test, explaining why you chose that output. Then test the hypothesis that the population means are not equal using a significance level of 0.05.

Food	Target	Whole
Bananas/1 lb	0.79	0.99
Grape tomatoes/1 lb	4.49	3.99
Russet potato/5 lb	4.49	4.99

Figure A

Paired T-Test and CI: Target, Whole

```
Paired T for Target - Whole

              N      Mean     StDev    SE Mean
Target       30     2.879     1.197      0.219
Whole        30     3.144     1.367      0.250
Difference   30    -0.265     1.152      0.210

95% CI for mean difference: (-0.695, 0.165)
T-Test of mean difference = 0 (vs ≠ 0): T-Value = -1.26
                                         P-Value = 0.217
```

Figure B

Two-Sample T-Test and CI: Target, Whole

```
Two-sample T for Target vs Whole

              N      Mean     StDev    SE Mean
Target       30      2.88      1.20       0.22
Whole        30      3.14      1.37       0.25

Difference = (Target) - μ(Whole)
Estimate for difference: -0.265
95% CI for difference: (-0.930, 0.399)
T-Test of difference = 0 (vs ≠ 0): T-Value = -0.80
                                    P-Value = 0.427
```

Figure C

9.72 Smoking Mothers The birth weights of 35 babies whose mothers did not smoke and 22 babies whose mothers smoked were compared; weights were in grams. (Source: Smoking Mothers, Holcomb 2006, accessed via StatCrunch. Owner: kupresanin99)

Test the hypothesis that the population mean birth weight is larger for mothers who do not smoke. Assume that the sample is random and the distributions are Normal, and use a significance level of 0.05.

a. Explain why this data set is not paired.

b. Which sample mean is larger, and how do you know?

c. Test the hypothesis that the population mean birth weight is larger for mothers who do not smoke.

Options

Hypothesis test results:
μ_1 : Mean of No Smoking
μ_2 : Mean of Yes Smoking
$\mu_1 - \mu_2$: Difference between two means
$H_0 : \mu_1 - \mu_2 = 0$
$H_A : \mu_1 - \mu_2 > 0$
(without pooled variances)

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	725	227.58343	31.393082	3.185645	0.0016

9.73 Ages of Brides and Grooms Data for the ages of grooms and their brides for a random sample of 31 couples in Ventura County, California, were obtained and can be found at this text's website.

a. Compare the sample means.

b. Test the hypothesis that there is a significant difference in mean ages of brides and grooms, using a significance level of 0.05.

c. If the test had been done to determine whether the mean for the grooms was significantly larger than the mean for the brides, how would that change the alternative hypothesis and the p-value?

9.74 Surfers Surfers and statistics students Rex Robinson and Sandy Hudson collected data on the number of days on which surfers surfed in the last month for 30 longboard (L) users and 30 shortboard (S) users. Treat these data as though they were from two independent random samples. Test the hypothesis that the mean days surfed for all longboarders is larger than the mean days surfed for all shortboarders (because longboards can go out in many different surfing conditions). Use a level of significance of 0.05.

Longboard: 4, 9, 8, 4, 8, 8, 7, 9, 6, 7, 10, 12, 12, 10, 14, 12, 15, 13, 10, 11, 19, 19, 14, 11, 16, 19, 20, 22, 20, 22

Shortboard: 6, 4, 4, 6, 8, 8, 7, 9, 4, 7, 8, 5, 9, 8, 4, 15, 12, 10, 11, 12, 12, 11, 14, 10, 11, 13, 15, 10, 20, 20

TRY **9.75 Self-Reported Heights of Men (Example 18)** A random sample of students at Oxnard College reported what they believed to be their heights in inches. Then the students measured each others' heights in centimeters, without shoes. The data shown are for the men. Assume that the conditions for *t*-tests hold.

a. Convert heights in inches to centimeters by multiplying inches by 2.54. Find a 95% confidence interval for the mean difference as measured in centimeters. Does it capture 0? What does that show?

b. Perform a *t*-test to test the hypothesis that the means are not the same. Use a significance level of 0.05, and show all four steps.

Men	
Centimeters	Inches
166	66
172	68
184	73
166	67
191	76
173	68
174	69
191	76

Men	
Centimeters	Inches
178	70
177	69
181	71
175	69
171	67
170	67
184	72

9.76 Eating Out Jacqueline Loya, a statistics student, asked students with jobs how many times they went out to eat in the last week. There were 25 students who had part-time jobs and 25 students who had full-time jobs. Carry out a hypothesis test to determine whether the mean number of meals out per week for students with full-time jobs is greater than that for those with part-time jobs. Use a significance level of 0.05. Assume that the conditions for a two-sample *t*-test hold.

Full-time jobs: 5, 3, 4, 4, 4, 2, 1, 5, 6, 5, 6, 3, 3, 2, 4, 5, 2, 3, 7, 5, 5, 1, 4, 6, 7

Part-time jobs: 1, 1, 5, 1, 4, 2, 2, 3, 3, 2, 3, 2, 4, 2, 1,2, 3, 2, 1, 3, 3, 2, 4, 2, 1

9.77, 9.79, and 9.80 For these questions, the data set is given at this text's website. Assume that the data sets are from random samples and the distributions are Normal.

a. Find a 95% confidence interval for the difference between means, state whether it captures 0, and explain what that shows about the means.

b. Perform a hypothesis test to see whether the means are significantly different using a significance level of 0.05. Explain your conclusion clearly.

9.77 Backpack Weights Compare the weights of backpacks of men and women. (Source: StatCrunch: Backpack. Owner: Wikipeterson)

9.78 Navy Commissary Prices Amber Sanchez, a statistics student, collected data on the prices of the same items at the Navy commissary on the naval base in Ventura County, California, and a nearby Kmart. The items were matched for content, manufacturer, and size and were priced separately.

a. Report and compare the sample means.

b. Assume that they are a random sample of items, and use a significance level of 0.05 to test the hypothesis that the Navy commissary has a lower mean price. Assume that the population distribution of differences is approximately Normal.

9.79 Sleep Hours Compare the weekday and weekend/holiday hours of sleep. Each pair of numbers is from one randomly selected person. This is a different set of data from the one in the chapter. (Source: StatCrunch Survey: Responses to Sleep survey. Owner: scsurvey)

9.80 Shoes Compare the numbers of pairs of shoes for men and women. (Source: StatCrunch Survey: Responses to Shoe survey. Owner: scsurvey)

CHAPTER REVIEW EXERCISES

9.81 Choose a test for each situation: one-sample *t*-test, two-sample *t*-test, paired *t*-test, and no *t*-test.

a. A researcher goes to a clothing store and observes whether each person is male or female and whether they return the clothes to the correct racks (yes or no) after trying them on.

b. A random sample of restaurants is obtained. Then the researcher walks into each restaurant wearing ordinary clothes and finds out how long it takes (in minutes) for a waiter to approach the researcher. Later, the researcher goes into the same restaurant dressed in expensive clothes and finds out how long it takes for a waiter to approach.

c. The supervisor observes the number of working hours of a random sample of part-time workers and a random sample of regular workers.

9.82 Choose a *t*-test for each situation: one-sample *t*-test, two-sample *t*-test, paired *t*-test, and no *t*-test.

a. A random sample of students of a college is asked their statistics score. Our goal is to determine whether the mean score for students of that college is significantly different from the population mean score for all the students at the university.

b. A researcher goes to a department store and observes whether each person coming in is male or female and whether they finish their shopping within 15 minutes (yes or no).

c. A researcher visits a hospital at night and records the number of cases handled by the doctors on duty and the gender of the doctors.

9.83 Cones: 3 Tests A McDonald's fact sheet says their cones should weigh 3.18 ounces (converted from grams). Suppose you take a random sample of four cones, and the weights are 4.2, 3.4, 3.9, and 4.4 ounces. Assume that the population distribution is Normal, and, for all three parts, report the alternative hypothesis, the *t*-value, the p-value, and your conclusion. The null hypothesis in all three cases is that the population mean is 3.18 ounces.

a. Test the hypothesis that the cones do not have a population mean of 3.18 ounces.

b. Test the hypothesis that the cones have a population mean less than 3.18 ounces.

c. Test the hypothesis that the cones have a population mean greater than 3.18 ounces.

* **9.84 Colas: 3 Tests** A random sample of 10,000 UTA colas was taken to see whether the mean weight was 16 fluid ounces, as marked on the container. The null hypothesis is that the population mean is 16 ounces. For all three parts, report the alternative hypothesis, the *t*-value, the p-value, and your conclusion. Refer to the Minitab output. (Source: Statcrunch: UTA Cola. Owner: craig_slinkman)

a. Test the hypothesis that the colas do not have a population mean of 16 fluid ounces.

b. Test the hypothesis that the colas have a population mean less than 16 fluid ounces.

c. Test the hypothesis that the colas have a population mean greater than 16 fluid ounces.

One-Sample T: ounces

```
Test of μ = 16 vs ≠ 16

Variable     N     Mean    StDev  SE Mean         95% CI           T      P
ounces   10000  15.9988  0.0985   0.0010  (15.9969, 16.0007)   -1.21  0.225
```

9.85 Brain Size Brain size for 20 random women and 20 random men was obtained and is reported in the table (in hundreds of thousands of pixels shown in an MRI). Test the hypothesis that men tend to have larger brains than women at the 0.05 level. (Source: Willerman, L., Schultz, R., Rutledge, J. N., and Bigler, E. (1991), "In Vivo Brain Size and Intelligence," *Intelligence*, 15, 223–228.)

Brain Size (100,000s of pixels)	
Female	**Male**
8.2	10.0
9.5	10.4
9.3	9.7
9.9	9.0
8.5	9.6
8.3	10.8
8.6	9.2
8.8	9.5
8.7	8.9
8.5	8.9

Brain Size (100,000s of pixels)	
Female	**Male**
8.1	9.1
7.9	9.6
8.3	9.4
8.0	10.6
7.9	9.5
8.7	10.0
8.6	8.8
8.3	9.5
9.5	9.3
8.9	9.4

9.86 Happiness and Gender Users of StatCrunch were polled and asked to indicate their level of happiness from 1 to 100 (most happy). The output shows the results of a *t*-test. (Source: StatCrunch Survey: Responses to Happiness survey. Owner: Webster West) There were 297 females and 380 males polled. Assume the sample is random.

a. Which had a higher sample mean, and how do you know?

b. Test the hypothesis that the reported levels of happiness are not equal for men and women, assuming that alpha is 0.05.

Options

Hypothesis test results:
μ_1 : Mean of Happy where Gender=Male
μ_2 : Mean of Happy where Gender=Female
$\mu_1 - \mu_2$: Difference between two means
$H_0 : \mu_1 - \mu_2 = 0$
$H_A : \mu_1 - \mu_2 \neq 0$
(without pooled variances)

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	-4.0718324	2.033609	651.88988	-2.002269	0.0457

9.87 Heart Rate before and after Coffee Elena Lucin, a statistics student, collected the data in the table showing heart rate (beats per minute) for a random sample of coffee drinkers before and 15 minutes after they drank coffee. Carry out a complete analysis, using the techniques you learned in this chapter. Use a 5% significance level to test whether coffee increases heart rate. The same amount of caffeinated coffee was served to each person, and you may assume that conditions for a *t*-test hold.

Before	After
90	92
84	88
102	102
84	96
74	96
88	100
80	84
68	68

Before	After
74	78
72	82
72	76
92	96
86	88
90	92
80	74

* **9.88 Exam Grades** The final exam grades for a sample of daytime statistics students and evening statistics students at one college are reported. The classes had the same instructor, covered the same material, and had similar exams. Using graphical and numerical summaries, write a brief description about how grades differ for these two groups. Then carry out a hypothesis test to determine whether the mean grades are significantly different for evening and daytime students. Assume that conditions for a *t*-test hold. Select your significance level.

Daytime grades: 100, 100, 93, 76, 86, 72.5, 82, 63, 59.5, 53, 79.5, 67, 48, 42.5, 39

Evening grades: 100, 98, 95, 91.5, 104.5, 94, 86, 84.5, 73, 92.5, 86.5, 73.5, 87, 72.5, 82, 68.5, 64.5, 90.75, 66.5

9.89 Hours of Television Viewing Data on the text's website show the number of hours per week of television viewing for random samples of fifth grade boys and fifth grade girls. Each student logged his or her hours for one Monday-through-Friday period. Assume that the students were independent; for example, there were no pairs of siblings who watched the same shows.

Using graphical and numerical summaries, write a brief description of how the hours differed for the boys and girls. Then carry out a hypothesis test to determine whether the mean hours of television viewing are different for boys and girls. Evaluate whether the conditions for a *t*-test are met, and state any assumptions you must make in order to carry out a *t*-test.

9.90 Reaction Distances Data on the text's website show reaction distances in centimeters for a random sample of 40 college students. Shorter distances indicate quicker reactions. Each student tried the experiment both with his or her dominant hand, and with his or her nondominant hand, catching the meter stick. The subjects all started with their dominant hand.

Examine summary statistics, and explain what we can learn from them. Then do an appropriate test to see whether the mean reaction distance is shorter for the dominant hand. Use a significance level of 0.05.

9.91 Shift Sleep Hours A survey was done comparing the number of hours of sleep for workers on the day shift and for workers on the night shift. Assume the sample is random. Descriptive statistics are shown. The output of a *t*-test is also shown. (Source: StatCrunch Survey: Group Data. Owner: jlb4wolf)

Options

Summary statistics for Hours Sleep:
Group by: Shift

Shift	n	Mean	Std. dev.	Median	Range	Min	Max	Q1	Q3
day	75	6.67	1.3968836	7	6	4	10	6	8
night	52	5.9519231	1.5154745	6	6	3	9	5	7

Options

Hypothesis test results:
μ_1 : Mean of Hours Sleep where Shift=night
μ_2 : Mean of Hours Sleep where Shift=day
$\mu_1 - \mu_2$: Difference between two means
$H_0 : \mu_1 - \mu_2 = 0$
$H_A : \mu_1 - \mu_2 \neq 0$
(without pooled variances)

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	-0.71807692	0.26492209	103.92772	-2.7105212	0.0079

a. Compare the sample means descriptively in at least one sentence.
b. Test the hypothesis that those on the night shift and those on the day shift do not have the same population mean number of hours of sleep. Use a significance level of 0.05.
c. If you could find a 95% confidence interval for the mean difference in means, would it capture 0? Explain.

9.92 Grocery Prices Grocery prices of the same items were compared at Target and Whole Foods. Assume the sampling was random. The descriptive statistics are shown, and the results of a *t*-test are shown. (Source: StatCrunch. Organic food price comparison, Fall 2011 (ECO252). Owner: Keith Cox.)

Options

Summary statistics:

Column	n	Mean	Std. dev.	Median	Min	Max	Q1	Q3
Target	30	2.8516667	1.2290116	3.015	0.69	4.99	1.89	3.99
Whole Foods	30	3.144	1.3665804	3.19	0.79	6.99	2	3.99

Options

Hypothesis test results:
$\mu_D = \mu_1 - \mu_2$: Mean of the difference between Target and Whole Foods
$H_0 : \mu_D = 0$
$H_A : \mu_D \neq 0$

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
Target - Whole Foods	-0.29233333	0.20678918	29	-1.4136781	0.1681

a. Compare the sample means descriptively.
b. Perform a hypothesis test to determine whether the means are significantly different using an alpha of 0.05.
c. If you were to find a 95% confidence interval for the mean difference, would it capture 0? Explain.

9.93 Maximum Tax Rate A random sample of 10 Democrats and 10 Republicans were asked what is the largest value (in percentage) that one should be expected to pay as taxes. (Source: StatCrunch Survey: Responses to Taxes in the U.S. survey. Owner: scsurvey)

Dem	80	65	10	9	50	20	50	35	10	10
Rep	35	25	30	28	25	10	10	5	12	20

a. Test the hypothesis that Democrats have a higher mean than Republicans, using a 0.05 significance level. Assume the data sets are Normal.
b. Create a new data set that repeats each observation in the provided data set. (This means each observation should appear twice.) Test the same hypothesis.
c. Compare the results.

9.94 Boys Weight Perception Do boys who feel that they are "under weight" actually weigh less than boys who feel that they are "about right" in weight? If not, this suggests that boys do not have a good sense of their weight relative to others. The table has a random sample of observations from the Youth At Risk survey, and includes the weights (in kg) of 17 year-old boys. Assume that the distribution of weights in both populations is close to Normal.

a. Test the hypothesis that the mean weight of boys who feel they are 'under weight' is less than the mean weight of boys who are 'about right'. Use a significance level of 0.05.

b. Using only the first five observations in each group, repeat the hypothesis test.

c. Compare the results. Why do the conclusions differ?

Under Weight	About Right
69	68
61	95
74	68
77	86
59	61
63	68
79	77
59	77
67	65
61	73
51	59
58	57
57	79
61	64
59	67

9.95 Groceries The table shows the prices of identical groceries at 7-Eleven and at Vons.

a. Test the hypothesis that the mean price at 7-Eleven is more than the mean price at Vons, at the 0.05 level. Assume that the sampling is random and the samples are Normal.

b. Although a two-sample t-test is not appropriate for this data set, do it anyway to see what happens.

* c. Compare the results and explain.

	Prices	
	7-Eleven	Vons
1 gal Milk	3.29	2.99
12 Eggs	1.89	2.19
Sliced ham (6 oz)	2.89	2.28
Coke (12-pack)	3.69	3.33
Campbell's soup	1.19	1.19
Half gal Milk	1.99	1.83
Cheerios (15 oz)	3.49	3.35
Cheezit (10 oz)	2.89	2.29

9.96 Parents The table shows the heights (in inches) of a random sample of students and their parent of the same gender. Test the hypothesis that the mean for the students is more than the mean for the parents, at the 0.05 level. Assume the data are Normal.

a. Use the paired t-test that is appropriate.

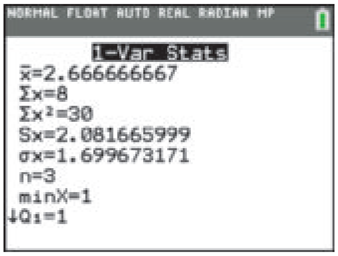
b. Use the two-sample t-test, even though it is not appropriate.

c. Compare the results. The numerator of both t-values is the difference in sample means, which is 1.12 inches. What must be causing the different t-values if the numerators are the same?

Height	
Student	Parent
70	65
71	71
61	60
63	65
65	67
68	66
70	72
63	63
65	64
63	62
65	65
73	70
68	64
72	70
71	69
68	65
69	68

* **9.97 Why Is $n - 1$ in the Sample Standard Deviation?** Why do we calculate s by dividing by $n - 1$, rather than just n?

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$



The reason is that if we divide by $n - 1$, then s^2 is an unbiased estimator of σ^2, the population variance.

We want to show that s^2 is an unbiased estimator of σ^2, sigma squared. The mathematical proof that this is true is beyond the scope of an introductory statistics course, but we can use an example to demonstrate that it is.

First we will use a very small population that consists only of these three numbers: 1, 2, and 5.

You can determine that the population standard deviation, σ, for this population is 1.699673 (or about 1.70), as show0n in the TI-84 output. So the population variance, sigma squared, σ^2, is 2.888889 (or about 2.89).

Now take all possible samples, with replacement, of size 2 from the population, and find the sample variance, s^2, for each sample. This process is started for you in the table. Average these sample variances (s^2), and you should get approximately 2.88889. If you do, then you have demonstrated that s^2 is an unbiased estimator of σ^2, sigma squared.

Sample	s	s^2
1, 1	0	0
1, 2	0.7071	0.5
1, 5	2.8284	8.0
2, 1		
2, 2		
2, 5		
5, 1		
5, 2		
5, 5		

Show your work by filling in the accompanying table and show the average of s^2.

9.98 Presidents of Sri Lanka The table shows the terms (in years) of office of the political parties as represented by the former prime ministers of Sri Lanka. Explain why it would be inappropriate to do a *t*-test with these data. UNP stands for the United National Party, and SLFP stands for the Sri Lanka Freedom Party. (Source: The Official Website of the Government of Sri Lanka)

UNP	5	6	3	4	11	4	1
SLFP	3	1	16	1	1	5	5

* **9.99** Construct two sets of body temperatures (in degrees Fahrenheit, such as 96.2°F), one for men and one for women, such that the sample means are different but the hypothesis test shows the population means are not different. Each set should have three numbers in it.

* **9.100** Construct GPAs for three or more sets of student pairs (six or more students). Make the students' GPAs similar but not exactly the same. Put all of the low-scoring students in Set A and all of the high-scoring students in Set B. Create the numbers such that a two-sample *t*-test will *not* show a significant difference in the mean GPAs of the low scorers of each pair and the mean GPAs of the high scorers of each pair, but the paired *t*-test *does* show a significant difference. (*Hint:* Make one of the pairs score the highest, one of the pairs score the lowest, and one of the pairs in-between.) Report all the numbers and the *t*- and p-values for the tests. Explain why the paired *t*-test shows a difference and the two-sample *t*-test does not show a difference.

GUIDED EXERCISES

g **9.37 Human Body Temperatures** A random sample of 10 independent healthy people showed body temperatures (in degrees Fahrenheit) as follows:

98.5, 98.2, 99.0, 96.3, 98.3, 98.7, 97.2, 99.1, 98.7, 97.2

The Minitab output of the results of a one-sample *t*-test is shown.

```
One-Sample T: Sample of 10
Test of mu = 98.6 vs not = 98.6

Variable   N     Mean   StDev  SE Mean           95% CI           T      P
sam10     10  98.1200  0.9187   0.2905  (97.4628, 98.7772)    -1.65  0.133
```

QUESTION Test the hypothesis that the population mean is not 98.6°F, using a significance level of 0.05. Write out the steps given, filling in the blanks.

Step 1 ▶ Hypothesize

H_0: $\mu = 98.6$

H_a: __________

Step 2 ▶ Prepare

A stemplot is shown that is not strongly skewed, suggesting that the distribution of the population is also approximately Normal. (A histogram would also be appropriate.) Comment on the data collection, and state the test to be used. State the significance level.

```
96  3
97  22
98  23577
99  01
```

Step 3 ▶ Compute to compare

t = __________

p-value = __________

Step 4 ▶ Interpret

Reject or do not reject H_0, and choose interpretation i, ii, or iii:

i. We cannot reject 98.6 as the population mean from these data at the 0.05 level.

ii. The population mean is definitely 98.6 on the basis of these data at the 0.05 level.

iii. We can reject the null hypothesis on the basis of these data, at the 0.05 level. The population mean is not 98.6.

g **9.57 Televisions** Minitab output is shown for a two-sample *t*-test for the number of televisions owned in households of random samples of students at two different community colleges. Each individual was randomly chosen independently of the others; the students were not chosen as pairs or in groups. One of the schools is in a wealthy community (MC), and the other (OC) is in a less wealthy community.

QUESTION Complete the steps to test the hypothesis that the mean number of televisions per household is different in the two communities, using a significance level of 0.05.

```
Two-Sample T-Test and CI: OCTV, MCTV

       N     Mean     StDev    SE Mean
OCTV  30     3.70      1.49       0.27
MCTV  30     3.33      1.49       0.27

Difference = mu OCTV - mu MCTV
Estimate for difference: 0.367
95% CI for difference: (-0.404, 1.138)
T-Test of difference = 0 (vs not =): T-Value = 0.95  P-Value = 0.345
```

Step 1 ▶ Hypothesize

Let μ_{oc} be the population mean number of televisions owned by families of students in the less wealthy community (OC), and let μ_{mc} be the population mean number of televisions owned by families of students in the wealthier community (MC).

H_0: $\mu_{oc} = \mu_{mc}$

H_a: ________ __

Step 2 ▶ Prepare

Choose an appropriate *t*-test. Because the sample sizes are 30, the Normality condition of the *t*-test is satisfied. State the other conditions, indicate whether they hold, and state the significance level that will be used.

Step 3 ▶ Compute to compare

$t =$ __________

p-value $=$ __________

Step 4 ▶ Interpret

Reject or do not reject the null hypothesis. Then choose the correct interpretation:

i. At the 5% significance level, we cannot reject the hypothesis that the mean number of televisions of all students in the wealthier community is the same as the mean number of televisions of all students in the less wealthy community.

ii. At the 5% significance level, we conclude that the mean number of televisions of all students in the wealthier community is different from the mean number of televisions of all students in the less wealthy community.

Confidence Interval

Report the confidence interval for the difference in means given by Minitab, and state whether it captures 0 and what that shows.

g **9.69 Female Pulse Rates before and after a Fright** In a statistics class taught by one of the authors, students took their pulses before and after being frightened. The frightening event was having the teacher scream and run from one side of the room to the other. The pulse rates (beats per minute) of the women before and after the scream were obtained separately and are shown in the table. Treat this as though it were a random sample of female community college students.

QUESTION Test the hypothesis that the mean of college women's pulse rates is higher after a fright, using a significance level of 0.05, by following the steps below.

Women	
Pulse Before	**Pulse After**
64	68
100	112
80	84
60	68
92	104
80	92
68	72

Women	
Pulse Before	**Pulse After**
84	88
80	80
68	92
60	76
68	72
68	80

Step 1 ▶ Hypothesize

μ is the mean number of beats per minute.

H_0: $\mu_{\text{before}} = \mu_{\text{after}}$

H_a: μ_{before} __________ μ_{after}

Step 2 ▶ Prepare

Choose a test: Should it be a paired *t*-test or a two-sample *t*-test? Why? Assume that the sample was random and that the distribution of differences is sufficiently Normal. Mention the level of significance.

Step 3 ▶ Compute to compare

$t =$ __________

p-value $=$ __________

Step 4 ▶ Interpret

Reject or do not reject H_0. Then write a sentence that includes "significant" or "significantly" in it. Report the sample mean pulse rate before the scream and the sample mean pulse rate after the scream.

CHECK YOUR TECH

Pulse Rates after Exercise: Understanding *t*

Pulse rates were observed for 35 people before and after running in place for 2 minutes. The Minitab output for a paired *t*-test is shown. We will check that the test statistic value reported by Minitab is correct by using a formula. We will not focus on the four steps.

Paired T-Test and CI: Pulse After Run, Pulse Before Run

```
Paired T for PulAftRun - PulBefRun

              N      Mean     StDev    SE Mean
PulAftRun    35     92.51     18.94       3.20
PulBefRun    35     73.60     11.44       1.93
Difference   35     18.91     15.05       2.54

95% CI for mean difference: (13.74, 24.08)
T-Test of mean difference = 0 (vs ≠ 0): T-Value = 7.44
                                        P-Value = 0.000
```

Paired t-test formula: $$t = \frac{\bar{x}_{\text{diff}} - 0}{SE_{\text{diff}}}$$

where

$$SE_{\text{diff}} = \frac{s_{\text{diff}}}{\sqrt{n}}$$

$\bar{x}$ is the average difference.

s_{diff} is the standard deviation of the differences.

n is the number of data pairs.

QUESTION Verify that t is 7.44 by following the steps below, using the formula given.

SOLUTION

Step 1 ▶ Find the numerator of t, x_{diff}

Find the difference: the average pulse after the run minus the average pulse before the run from the two means given. This is the numerator. Retain two decimal digits, which is what the output shows.

Step 2 ▶ Find the denominator of t, SE_{diff}

To verify the standard error (*SE*) of the difference, take the standard deviation of the difference (StDev) and divide by the square root of 35, which is the sample size. (You may round to two decimal digits, which is what Minitab does.)

Step 3 ▶ Find t

Divide the numerator (from step 1) by the denominator from step 2 and see whether you get 7.44 (or close to it), the t reported in the output.

Step 4 ▶ Understanding what influences t

In answering these questions, consider whether the variable is in the numerator or in the denominator.

a. If the difference between means were larger, and all else were the same, would that cause t to be larger or smaller? Why?

b. If the standard deviation (s_{diff}) were larger, and all else were the same, would that cause t to be larger or smaller? Why?

c. If the sample size were larger, and all else were the same, would that cause t to be larger or smaller? Why?

TechTips

General Instructions for All Technology

Because of the limitations of the algorithms, precision, and rounding involved in the various technologies, there can be slight differences in the outputs. These differences can be noticeable, especially for the calculated p-values involving *t*-distributions. It is suggested that the technology that was used be reported along with the p-value, especially for two-sample *t*-tests.

EXAMPLE A (ONE-SAMPLE *t*-TEST AND CONFIDENCE INTERVAL): ▶ McDonald's sells ice cream cones, and the company's fact sheet says that these cones weigh 3.2 ounces. A random sample of 5 cones was obtained, and the weights were 4.2, 3.6, 3.9, 3.4, and 3.3 ounces. Test the hypothesis that the population mean is 3.2 ounces. Report the *t*- and p-values. Also find a 95% confidence interval for the population mean.

EXAMPLE B (TWO-SAMPLE *t*-TEST AND CONFIDENCE INTERVAL): ▶ Below are the GPAs for random samples of men and women.

Men: 3.0, 2.8, 3.5

Women: 2.2, 3.9, 3.0

Perform a two sample *t*-test to determine whether you can reject the hypothesis that the population means are equal. Find the *t*- and p-values. Also find a 95% confidence interval for the difference in means.

EXAMPLE C (PAIRED *t*-TEST): ▶ Here are the pulse rates (in beats per minute) before and after exercise for three randomly selected people.

Person	Before	After
A	60	75
B	72	80
C	80	92

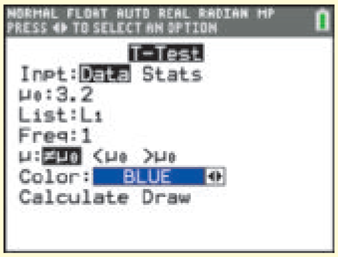
Determine whether you can reject the hypothesis that the population mean change is 0 (in other words, that the two population means are equal). Find the *t*- and p-values.

TI-84

One-Sample *t*-Test

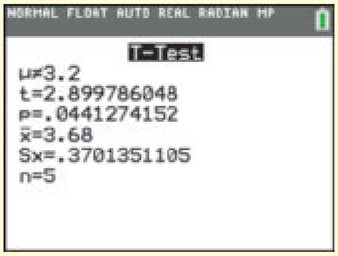
1. Press **STAT** and choose **EDIT**, and type the data into **L1** (list one).
2. Press **STAT**, choose **TESTS**, and choose **2: T-Test**.
3. Note this but don't do it: If you did not have the data in the list and wanted to enter summary statistics such as $\bar{x}$, s, and n, you would put the cursor over **Stats** and press **ENTER**, and put in the required numbers.
4. See Figure 9A. Because you have raw data, put the cursor over **Data** and press **ENTER**.

Enter: μ_0, **3.2**; **List**, **L1**; **Freq: 1**; put the cursor over $\neq$ and press **ENTER**; scroll down to **Calculate** and press **ENTER**.



▲ **FIGURE 9A** TI-84 Input for One-Sample *t*-Test

Your output should look like Figure 9B.



▲ **FIGURE 9B** TI-84 Output for One-sample *t*-Test

One-Sample *t*-Interval

1. Press **STAT**, choose **TESTS**, and choose **8: TInterval**.
2. Choose **Data** because you have raw data. (If you had summary statistics, you would choose **Stats**.) Choose the correct **List** (to select **L1**, press **2nd** and **1**) and **C-Level**, here 0.95. Leave **Freq:1**, which is the default. Scroll down to **Calculate** and press **ENTER**.

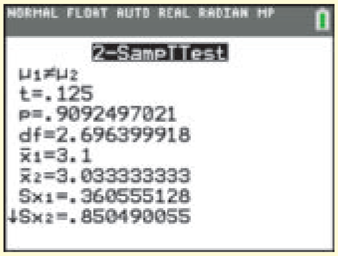
The 95% confidence interval reported for the mean weight of the cones (ounces) is (3.2204, 4.1396).

Two-Sample *t*-Test

1. Press **STAT**, choose **EDIT**, and put your data (GPAs) in two separate lists (unstacked). We put the men's GPAs into **L1** and the women's GPAs into **L2**.
2. Press **STAT**, choose **TESTS**, and choose **4:2-Samp-TTest**.
3. For **Inpt**, choose **Data** because we put the data into the lists. (If you had summary statistics, you would choose **Stats** and put in the required numbers.)

4. In choosing your options, be sure the lists chosen are the ones containing the data; leave the **Freq**s as 1, choose ≠ as the alternative, and choose **Pooled No** (which is the default). Scroll down to **Calculate** and press **ENTER**.

You should get the output shown in Figure 9C. The arrow down on the left-hand side means that you can scroll down to see more of the output.



▲ **FIGURE 9C** TI-84 Output for Two-Sample *t*-Test

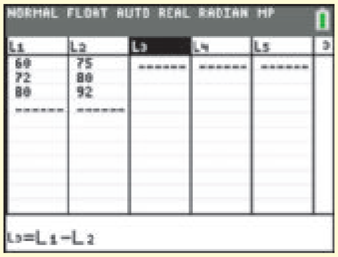
Two-Sample *t*-Interval

1. After entering your data into two lists, press **STAT**, choose **TESTS**, and choose **0:2-SampTInt**.
2. Choose **Data** because you have raw data. (If you had summary statistics, you would choose **Stats**.) Make sure the lists chosen are the ones with your data. Leave the default for **Freq1:1**, **Freq2:1** and **Pooled No**. Be sure the **C-Level** is **.95**. Scroll down to **Calculate** and press **ENTER**.

The interval for the GPA example will be (−1.744, 1.8774) if the men's data correspond to **L1** and the women's to **L2**.

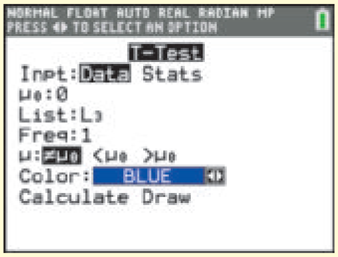
Paired *t*-Test

1. Enter the data given in Example 3 into **L1** and **L2** as shown in Figure 9D.



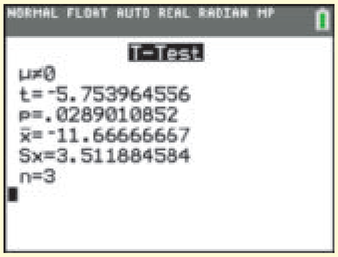
▲ **FIGURE 9D** Obtaining the List of Differences for the TI-84

2. See Figure 9D. Use your arrows to move the cursor to the top of **L3** so that you are in the label region. Then press **2ND L1 - 2ND L2**. For the minus sign, be sure to use the button above the plus button. Then press **ENTER**, and you should see all the differences in **L3**.
3. Press **STAT**, choose **TESTS**, and choose **2: T-Test**.
4. See Figure 9E. For **Inpt** choose **Data**. Be sure μ_0 is **0** because we are testing to see whether the mean difference is 0. Also be sure to choose **L3**, if that is where the differences are. Scroll down to **Calculate** and press **ENTER**.



▲ **FIGURE 9E** TI-84 Input for Paired *t*-Test

Your output should look like Figure 9F.

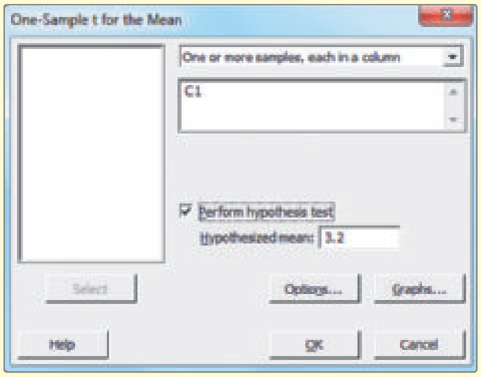


▲ **FIGURE 9F** TI-84 Output for Paired *t*-Test

MINITAB

One-Sample *t*-Test and Confidence Interval

1. Type the weight of the cones in **C1** (column 1).
2. **Stat > Basic Statistics > 1-Sample t**
3. See Figure 9G. Click in the empty white box below **One or more samples…**. A list of columns containing data will appear to the left. Double click **C1** to choose it. Check **Perform hypothesis test**, and put in **3.2** as the **Hypothesized mean**. (If you wanted a one-sided test or a confidence level other than 95%, you would use **Options**.)
4. Click **OK**.



▲ **FIGURE 9G** Minitab Input Screen for One-Sample *t*-Test

The output is shown in Figure 9H.

```
One-Sample T: C1

Test of μ = 3.2 vs ≠ 3.2

Variable  N   Mean  StDev  SE Mean       95% CI          T      P
C1        5  3.680  0.370    0.166  (3.220, 4.140)  2.90  0.044
```

▲ **FIGURE 9H** Minitab Output for One-Sample *t*-Test and Confidence Interval

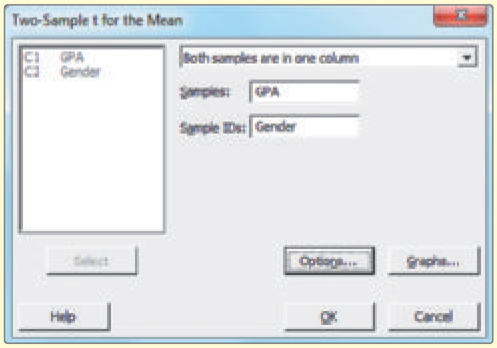
Note that the Minitab output (Figure 9H) includes the 95% confidence interval for the mean weight (3.22, 4.14) as well as the *t*-test.

Two-Sample *t*-Test and Confidence Interval

Use stacked data with all the GPAs in one column. The second column will contain the categorical variable that designates groups: male or female.

1. Upload the data from the disk or use the following procedure: Enter the GPAs in the first column. In the second column put the corresponding **m** or **f**. (Complete words or coding is also allowed for the second column, but you must decide on a system for one data set and stick to it. For example, using F one time and f the other times within one data set will create problems.) Use headers for the columns: **GPA** and **Gender.**
2. **Stat** > **Basic Statistics** > **2-Sample t**
3. Refer to Figure 9I. Choose **Both samples are in one column,** because we have stacked data. Click in the small box to the right of **Samples** at the top to activate the box. Then double click **GPA** and then double click **Gender** to get it into the **Sample IDs** box. If you wanted to do a one-sided test or to use a confidence level other than 95%, you would click **Options.**

   (If you had unstacked data, you would choose **Each sample is in its own column**, and choose both columns of data.)
4. Click **OK**.



▲ **FIGURE 9I** Minitab Input Screen for Two-Sample *t*-Test and Confidence Interval

The output is shown in Figure 9J. Note that the confidence interval is included.

```
Two-sample T for GPA

Gender  N   Mean  StDev  SE Mean
F       3  3.033  0.850     0.49
M       3  3.100  0.361     0.21


Difference = μ (F) - μ (M)
Estimate for difference:  -0.067
95% CI for difference:  (-2.361, 2.228)
T-Test of difference = 0 (vs ≠): T-Value = -0.13  P-Value = 0.912  DF = 2
```

▲ **FIGURE 9J** Minitab Output for Two-Sample *t*-Test and Confidence Interval

Paired *t*-Test and Confidence Interval

1. Type the numbers (pulse rates) in two columns. Label the first column "**Before**" and the second column "**After**."
2. **Stat** > **Basic Statistics** > **Paired t**
3. Click in the small box to the right of **First sample** to activate the box. Then double click **Before** and double click **After** to get it in the **Second sample** box. (If you wanted a one-sided test or a confidence level other than 95%, you would click **Options**.)
4. Click **OK**.

Figure 9K shows the output.

```
Paired T for Before - After

            N    Mean  StDev  SE Mean
Before      3   70.67  10.07     5.81
After       3   82.33   8.74     5.04
Difference  3  -11.67   3.51     2.03


95% CI for mean difference: (-20.39, -2.94)
T-Test of mean difference = 0 (vs ≠ 0): T-Value = -5.75  P-Value = 0.029
```

▲ **FIGURE 9K** Minitab Output for Paired *t*-Test and Confidence Interval

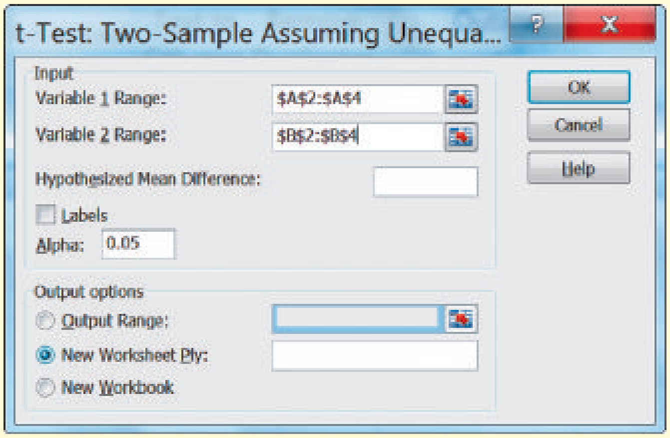
EXCEL

We are saving the one-sample *t*-test for last, because we have to treat it strangely.

Two-Sample *t*-Test

1. Type the GPAs in two columns side by side.
2. Click on **Data** and **Data Analysis,** and then scroll down to **t-Test: Two-Sample Assuming Unequal Variances** and double click it.
3. See Figure 9L. For the **Variable 1 Range** select one column of numbers (don't include any labels). Then click inside the box for **Variable 2 Range**, and select the other column of numbers.

   You may leave the hypothesized mean difference empty, because the default value is 0, and that is what you want.
4. Click **OK**.



▲ **FIGURE 9L** Excel Input for Two-Sample *t*-Test

To see all of the output, you may have to click **Home**, **Format** (in the **Cells** group), and **AutoFit Column Width**.

Figure 9M shows the relevant part of the output.

t Stat	0.125
P(T<=t)one-tail	0.454214709
P(T<-t)two-tail	0.908429419

▲ **FIGURE 9M** Part of the Excel Output for Two-Sample *t*-Test

For a one-sided alternative hypothesis, Excel always reports one-half the p-value for the two-sided hypothesis. This is the correct p-value only when the observed value of the test statistic is consistent with the direction of the alternative hypothesis. (In other words, if the alternative hypothesis is ">", then the observed test statistic must be positive; if "<", then the observed value must be negative.) If this is not the case, to find the correct p-value, calculate 1 minus the reported one-tailed p-value.

Paired *t*-Test

1. Type the data into two columns.
2. Click on **Data**, **Data Analysis**, and **t-Test: Paired Two Sample for Means**, and follow the same procedure as for the two-sample *t*-Test.

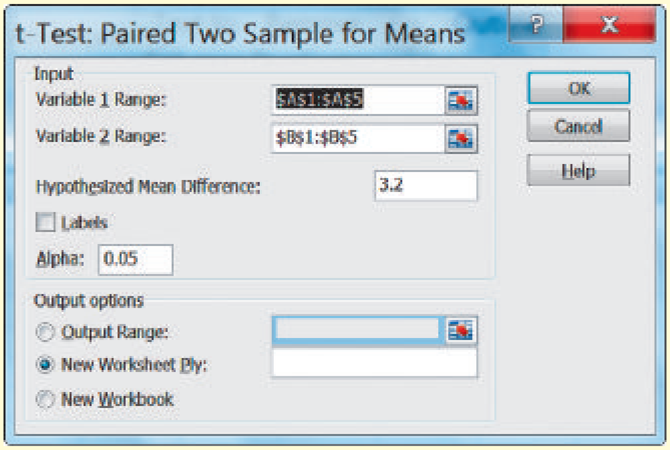
One-Sample *t*-Test

1. You need to use a trick to force Excel to do this test. Enter the weights of the cones in column A, and put zeros in column B so that the columns are equal in length, as shown in Figure 9N.

A	B
4.2	0
3.6	0
3.9	0
3.4	0
3.3	0

▲ **FIGURE 9N** Excel Data Entry for One-Sample *t*-Test

2. Click **Data**, **Data Analysis**, and **t-Test: Paired Two Sample for Means**.
3. See Figure 9O. After selecting the two groups of data, you need to put the hypothesized mean in the box labeled **Hypothesized Mean Difference**. For the way we set up this example, it is **3.2**. (If you had entered 3.2's in column B, you would put in 0 for the Hypothesized Mean Difference.)
4. Click **OK**.



▲ **FIGURE 9O** Excel Input for One-Sample *t*-Test

Figure 9P shows the relevant part of the Excel output.

Hypothesized Mean Difference	3.2
Df	4
P(T<=t) one-tail	0.022063708
P(T<=t) two-tail	0.044127415

▲ **FIGURE 9P** Relevant Part of the Excel Output for One-Sample *t*-Test

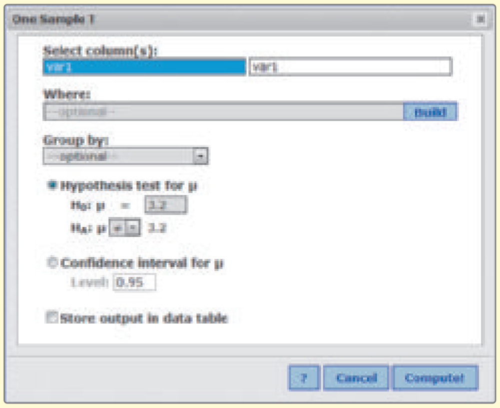
Again, the p-value for the one-sided alternative hypothesis is consistent with the alternative hypothesis that the mean weight was *more* than 3.2 ounces.

STATCRUNCH

One-Sample *t*-Test

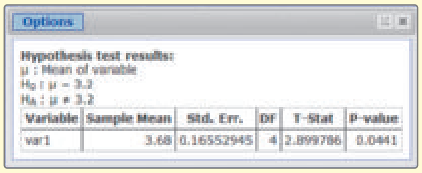
1. Type the weights of the cones into the first column.
2. **Stat > T Stats > One Sample > With Data**

   (If you had summary statistics, then after **One Sample**, you would choose **With Summary**.)
3. See Figure 9Q. Click on the column containing the data, **var1**. Put in the **H_0: μ**, which is **3.2**. Leave the default not equal for the **H_A: μ**, and click **Compute!**



▲ **FIGURE 9Q** StatCrunch Input for One-Sample *t*-Test

You will get the output shown in Figure 9R.



▲ **FIGURE 9R** StatCrunch Output for One-Sample *t*-Test

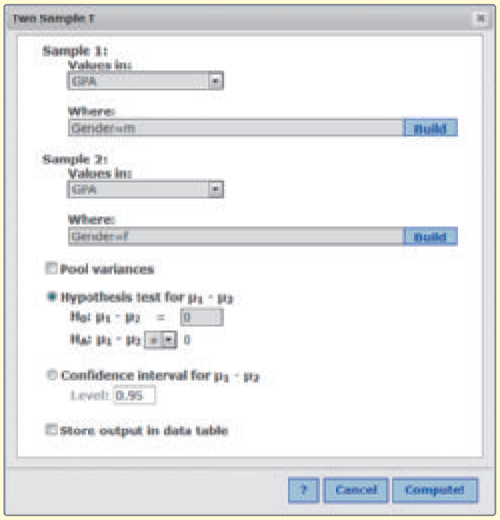
One-Sample Confidence Interval

Go back and perform the same steps as for the one-sample *t*-test, but when you get to step 3, check **Confidence interval for μ**. You may change the confidence level from the default 0.95 if you want. See Figure 9Q.

Two-Sample *t*-Test

Use stacked data with all the GPAs in one column. The second column will contain the categorical variable that designates groups: male or female.

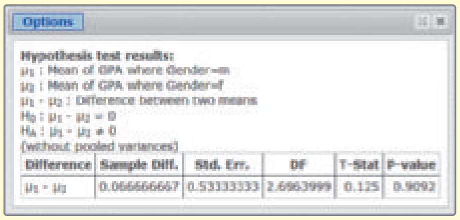
1. Upload the data from the disk or follow these steps: Enter the GPAs in column 1 (**var1**). Put **m** or **f** in column 2 (**var2**) as appropriate. (Complete words or coding for column 2 is also allowed, but whatever system you use must be maintained within the data set.) Put labels at the top of the columns; change **var1** to **GPA** and **var2** to **Gender**.
2. **Stat > T Stats > Two Sample > With Data**
3. See Figure 9S. For **Sample 1 Values in:** choose **GPA**, and for **Where** put **Gender=m**. For **Sample 2** choose **GPA** and put **Gender=f**. Click off **Pool variances**. (If you had unstacked data, you would choose the two lists for the two samples and not use the **Where** boxes.)



▲ **FIGURE 9S** StatCrunch Input for Two-Sample *t*-Test (Stacked Data)

4. Click **Compute!**

Figure 9T shows the output.



▲ **FIGURE 9T** StatCrunch Output for Two-Sample *t*-Test

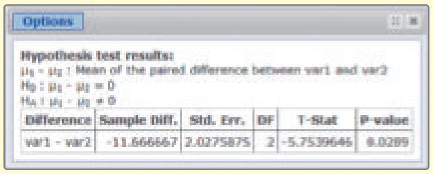
Two-Sample Confidence Interval

1. Go back and do the preceding first three numbered steps.
2. Check **Confidence interval for $\mu_1 - \mu_2$**.
3. Click **Compute!**

Paired *t*-Test

1. Type the pulse rates Before in column 1 and the pulse rates After in column 2. The headings for the columns are not necessary.
2. **Stat > T Stats > Paired**
3. Select the two columns. Ignore the **Where** and **Group by** boxes.
4. Click **Compute!**

Figure 9U shows the output.



▲ **FIGURE 9U** StatCrunch Output for Paired *t*-Test

10 Analyzing Categorical Variables and Interpreting Research

THEME

Distributions of categorical variables are often summarized in two-way tables. We can make inferences about these distributions by calculating how many observations we would expect in each cell if the null hypothesis were true, and then comparing this to the actual counts.

As we have shown in the last few chapters, statistical tests are comparisons between two different views. In one view, if two samples have different means, then it might be because the populations are truly different. In the other view, randomness rules, and this difference is due merely to chance. A sample of people who had watched violent TV as children may behave more aggressively toward their partners when adults. But this might just be due to chance. Changing the placement of an advertisement on an Internet search results page might lead to a greater number of "follow-through" visits to the company that placed the ad. Or the difference could be due to chance—that is, if we were to repeat the study, we might see a very different outcome.

In this chapter we ask the same question as before: Are the results we see due to chance, or might something else be going on? In Chapter 8, we asked this question about one-sample and two-sample proportions. We now ask the same question with respect to categorical variables with multiple categories. For example, we might compare two categorical variables: the results of popping a batch of popcorn ("good" and "bad") for different amounts of oil ("no oil," "medium amount of oil," and "maximum oil").

This chapter goes beyond analyzing categorical variables by including suggestions for reading and interpreting published scientific studies. At this point in your studies, you know enough statistics to be able to critically evaluate scientific findings published in academic journals or reported on the nightly news, on blogs, and in newspapers. Although few of us will carry out formal experiments, we all need to know how to evaluate the strength of the conclusions of these studies, since many studies, particularly medical studies, can directly impact our lives.

CASE STUDY

Popping Better Popcorn

You're planning a movie night and decide to cook up some popcorn. Many factors might affect how good the popcorn tastes: the brand, for example, or how much oil you use, or how long you let the popcorn pop before stopping. Some researchers (Kukuyeva et al. 2008) investigated factors that might determine how to pop the perfect batch of popcorn. They decided that if more than half of the kernels in a bag were popped in the first 75 seconds, then the batch was a "success." If fewer than half, then it was a "failure." They popped 36 bags under three different treatments: no oil, "medium" amount of oil (1/2 tsp), and "maximum" oil (1 tsp). The bags were randomly assigned to an oil group. Each bag had

	Oil Amount		
Result	**No Oil**	**Medium Oil**	**Maximum Oil**
Failure	23	22	33
Success	13	14	3

50 kernels. The outcome is shown in the accompanying table. From the table it looks as though it is bad to use too much oil. But might this just be due to chance? In other words, if other investigators were to do this experiment the same way, would they also find so few successful bags with the maximum amount of oil?

This question—Is the outcome due to chance?—is one we've asked before. In this study the two variables are the amount of oil (with three values: none, medium, and maximum), and the result (success or failure). Both *Oil* and *Result* are categorical variables. In this chapter, we will see how to test the hypothesis that the amount of oil had an effect on the outcome.

SECTION 10.1

The Basic Ingredients for Testing with Categorical Variables

Hypothesis tests that involve categorical variables follow the same four-step recipe you studied in Chapters 8 and 9. However, the basic ingredients are slightly different. Here we introduce these ingredients: data, expected counts, the chi-square statistic (our test statistic), and the chi-square distribution.

Details

Pronunciation
Chi is pronounced "kie" (rhymes with "pie"), with a silent h.

1. The Data

Recall that categorical variables are those whose values are categories, as opposed to numbers. A respondent's income level can be given as a category (such as high, moderate, or low.)

When two categorical variables are analyzed, as we are doing in this chapter, they are often displayed in a **two-way table**, a summary table that displays frequencies for the outcomes. Such tables can give the impression that the data are numerical, because you are seeing numbers in the table. But it's important to keep in mind that the numbers are *summaries* of variables whose values are *categories*.

For instance, the General Social Survey (GSS) asked respondents, "There are always some people whose ideas are considered bad or dangerous by other people. For instance, somebody who is against churches and religion. If such a person wanted to make a speech in your (city/town/community) against churches and religion, should he be allowed to speak, or not?" These respondents were also asked about their income and, depending on their response, were assigned to one of four annual income levels: \$0–\$20K, \$20–\$40K, \$40–\$70K, and \$70K and more. The first four lines of the actual data would look something like Table 10.1. A value on the boundary would be assigned to the group with the larger incomes, so an income of 40 thousand would be in the 40–60 group.

▶ **TABLE 10.1** The first four lines of the raw data (based on an actual data set) showing responses to two questions on the General Social Survey.

Observation ID	Response	Income Level
1	allowed	20–40
2	allowed	20–40
3	not allowed	70 and over
4	allowed	0–20
. . .	. . .	. . .

We have chosen to display another variable, *Observation ID*, which is useful simply for keeping track of the order in which the observations are stored.

A two-way table summarizes the association between the two variables *Response* and *Income Level* as in Table 10.3.

Response	Income			
	0–20	20–40	40–70	70+
Allowed	248	311	223	229
Not Allowed	100	105	55	21

TABLE 10.2 Two-way table summarizing the association between responses to the question about whether speeches opposed to church and religion should be allowed, and income level. (Source: GSS 2012–2014, http://www.teachingwithdata.org)

2. Expected Counts

The **expected counts** are the numbers of observations we would see in each cell of the summary table if the null hypotheses were true.

Consider a study of the link between TV viewing and violent behavior. Researchers compared TV viewing habits for children, and then interviewed them about violent behavior many years later (Husemann et al. 2003). Subjects were asked if they had ever pushed, grabbed, or shoved their partner. Table 10.3 summarizes the results.

There are two categorical variables: *TV Violence* ("High" or "Low") and *Physical Abuse* ("Yes" or "No"). We wish to know whether these variables are associated. The null hypothesis says that they are not—that these variables are independent. In other words, any patterns you might see are due purely to chance.

Details

Expectations

Expected counts are actually long-run averages. When we say that we "expect" 10 observations in a cell of a table, we mean that if the null hypothesis were true and we were to repeat this data collection many, many times, then, on average, we would see 10 observations in that cell.

	High TV Violence	Low TV Violence	Total
Yes, Physical Abuse	25	57	82
No Physical Abuse	41	206	247
Total	66	263	329

TABLE 10.3 A two-way summary of TV violence and later abusive behavior.

What counts should we expect if these variables are truly not related to each other? That is, if the null hypothesis is true, what would we expect the table to look like? There are two ways of answering this question, and they both lead to the same answer. Let's look at them both.

Starting with the* Physical Abuse *Variable We notice that out of the entire sample, 82/329 (a proportion of 0.249240, or about 24.92%) said that they had physically abused their partners. If abuse is independent of TV watching—that is, if there is no relationship between these two variables—then we should expect to find the same percentages of violent abuse in those who watched high TV violence and in those with low TV violence.

So when we consider the 66 people who watched high TV violence as children, we expect 24.92% of them to have abused their partners. This translates to $(0.249240 \times 66) = 16.4498$ people. When we consider the 263 people who watched low TV violence, we expect again that 24.92%, or $0.249240 \times 263 = 65.5501$ people to have abused their partners.

We can use similar reasoning to find the other expected counts. We know that if a proportion of 0.249240 abused their partners, then $1 - 0.249240 = 0.750760$ did not.

This proportion should be the same no matter which level of TV violence was watched. Thus, we expect the following:

In the "no physical abuse and high TV violence" group,

$$0.750760 \times 66 = 49.5502$$

In the "no physical abuse and low TV violence" group,

$$0.750760 \times 263 = 197.4499$$

(We are avoiding rounding in these intermediate steps so that our answers for the expected values will be as accurate as possible.)

We summarize these calculations in Table 10.4, which shows the expected counts in parentheses. We include the actual counts in the same table so that we can compare. Note that we rounded to two decimal digits for ease of presentation.

▶ **TABLE 10.4** TV Violence: A summary including expected values (in parentheses).

	High TV Violence	Low TV Violence	Total
Yes, Physical Abuse	25 (16.45)	57 (65.55)	82
No Physical Abuse	41 (49.55)	206 (197.45)	247
Total	66	263	329

Details

Rows and Columns
Some statisticians refer to the row and column totals as "marginal totals."

If we call the values 82 and 247 row totals (for obvious reasons, we hope!), 66 and 263 column totals, and 329 the grand total, we can generate a formula for automatically finding expected counts for each cell. This formula is rarely needed. First, you can and should always think through the calculations as we did here. Second, most software will do this automatically for you.

Formula 10.1: $\text{Expected count for a cell} = \dfrac{(\text{row total}) \times (\text{column total})}{\text{grand total}}$

Details

Fractions of People
Does it bother you that we have fractions of people in each category? It *is* a little strange, until you think about this in terms of an ideal model. These expected counts are like averages. We say the average family has 2.4 children, and we know very well that there is no single family with a 0.4 child. This number 2.4 is a description of the collection of all families. Our claim that we expect 16.45 people (in this group) who have seen high TV violence to be abusive is a similar idealization.

Starting with the TV Violence Variable The other way of finding the expected counts is to begin by considering the *TV Violence* variable, rather than beginning with the *Physical Abuse* variable. We see that $66/329 = 0.200608$, or 20.06%, watched high TV violence. The rest, $263/329 = 0.799392$, or 79.94%, watched low TV violence.

If these variables are not related, then when we look at the 82 people who committed physical abuse, we should expect about 20.06% of them to fall in the High TV Violence category and the rest to fall in the Low TV Violence category.

Also, when we look at the 247 who did not commit physical abuse, we should expect 20.06% of them to have viewed high TV violence.

Among the abusive, the expected number with high TV violence is $0.200608 \times 82 = 16.45$. Among the nonabusive, the expected number with high TV violence is $0.200608 \times 247 = 49.55$.

You see that you will get the same result no matter which variable you consider first.

EXAMPLE 1 Gender and Opinion on Same-Sex Marriage

Do men and women feel differently about the issue of same-sex marriage? In 2012, the General Social Survey took a random sample of 1287 Americans, recording their gender and their level of agreement with the statement "Homosexuals should have the right to marry." The results are summarized in Table 10.5.

Option	Male	Female	Total
Strongly Agree	120	203	323
Agree	142	173	315
Neutral	64	88	152
Disagree	88	95	183
Strongly Disagree	157	157	314
Total	571	716	1287

◀ **TABLE 10.5** Summary of gender, and opinion that homosexuals should have the right to marry.

QUESTIONS Assuming that the two variables *Opinion* and *Gender* are *not* associated, find the number of males who would be expected to agree strongly and the number of females who would be expected to agree strongly.

SOLUTION We consider Gender first. The percentage of men in the sample is $(571/1287) \times 100\% = 44.3667\%$. The percentage of women is therefore $100\% - 44.3667\% = 55.6333\%$. If Gender is not associated with Opinion, then if we look at the 323 people who strongly agree, we should see that about 44.4% of those who strongly agree are male, and about 55.6% are female. In other words:

CONCLUSION Expected count of males who strongly agree $= 323 \times 0.443667 = 143.3044$, or about 143.30.

Expected count of females who strongly agree $= 323 \times 0.556333 = 179.6956$, or about 179.70.

TRY THIS! Exercise 10.7

3. The Chi-Square Statistic

The **chi-square statistic** is a statistic that measures the amount that our expected counts differ from our observed counts. This statistic is shown in Formula 10.2.

$$\textbf{Formula 10.2:} \quad X^2 = \sum_{\text{cells}} \frac{(O - E)^2}{E}$$

where

O is the observed count in each cell
E is the expected count in each cell
$\sum$ means add the results from each cell

Why does this statistic work? The term $(O - E)$ is the difference between what we observe and what we expect under the null hypothesis. To measure the total amount of deviation between Observed and Expected, it is tempting to just add together the individual differences. But this doesn't work, because the expected counts and the observed counts always add to the same value; if we sum up the differences, they will always add to 0.

You can see that the differences between Observed and Expected add to 0 in Table 10.6 on the next page, where we've listed the data from Example 1, but we now show the expected counts as well as the differences (Observed minus Expected).

One reason why the chi-square statistic uses squared differences is that by squaring the differences, we always get a positive value, because both negative and positive numbers multiplied by themselves result in positive numbers:

$$\frac{(-23.3)^2}{143.3} + \frac{23.3^2}{179.7} + \frac{2.24^2}{139.76} + \frac{(-2.24)^2}{175.24} + \cdots$$

▶ **TABLE 10.6** Gender and opinion on same-sex marriage, emphasizing the Observed minus Expected values.

Outcome	Observed Counts	Expected Counts	Observed minus Expected
Strongly Agree Male	120	143.3	−23.3
Strongly Agree Female	203	179.7	23.3
Agree Male	142	139.76	2.24
Agree Female	173	175.24	−2.24
Neutral Male	64	67.44	−3.44
Neutral Female	88	84.56	3.44
Disagree Male	88	81.19	6.81
Disagree Female	95	101.81	−6.81
Strongly Disagree Male	157	139.31	17.69
Strongly Disagree Female	157	174.69	−17.69

Why divide by the expected count? The reason is that a difference between the expected and actual counts of, say, 2 is a small difference if we were expecting 1000 counts. But if we were expecting only 5 counts, then this difference of 2 is substantial. By dividing by the expected count, we're controlling for the size of the expected count. Basically, for each cell, we are finding what proportion of the expected count the squared difference is.

If we apply this formula to the data in Example 1, we get $X^2 = 12.26$. We must still decide whether this value discredits the null hypothesis that gender and the answer to the question are independent. Keep reading.

EXAMPLE 2 Viewing Violent TV as a Child and Abusiveness as an Adult

Table 10.7 shows summary statistics from a study that asked whether there was an association between watching violent TV as a child and aggressive behavior toward one's spouse later in life. The table shows both actual counts and expected counts (in parentheses).

▶ **TABLE 10.7** A two-way summary of the effect of viewing TV violence on later abusiveness (expected values are shown in parentheses).

	High TV Violence	Low TV Violence	Total
Yes, Physical Abuse	25 (16.45)	57 (65.55)	82
No Physical Abuse	41 (49.55)	206 (197.45)	247
Total	66	263	329

QUESTION Find the chi-square statistic to measure the difference between the observed counts and expected counts for the study of the effect of violent TV on future behavior.

SOLUTION We use Formula 10.2 with the values for O and E taken from Table 10.7.

$$X^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

$$= \frac{(25 - 16.45)^2}{16.45} + \frac{(57 - 65.55)^2}{65.55} + \frac{(41 - 49.55)^2}{49.55} + \frac{(206 - 197.45)^2}{197.45} = 7.4047$$

CONCLUSION

$$X^2 = 7.40$$

Later we will see whether this is an unusually large value for two independent variables.

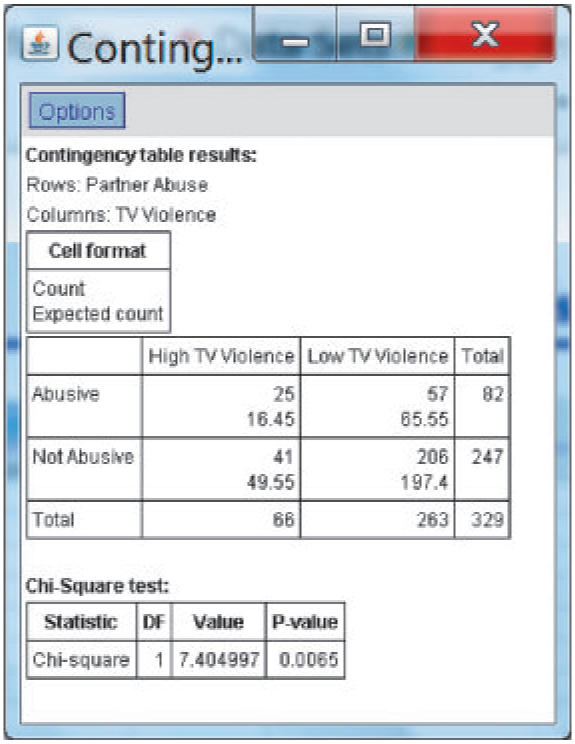
TRY THIS! Exercise 10.9

As you might expect, for tables with many cells, these calculations can quickly become tiresome. Fortunately, technology comes to our rescue. Most statistical software will calculate the chi-square statistic for you, given data summarized in a two-way table (as in Table 10.7) or presented as raw data (as in Table 10.1), and some software will even display the expected counts alongside the observed counts. Figure 10.1 shows the output from StatCrunch for these data.

When the null hypothesis is true, our real-life observations will usually differ slightly from the expected counts just by chance. When this happens, the chi-square statistic will be a small value.

If reality is very different from what our null hypothesis claims, then our observed counts should differ substantially from the expected counts. When that happens, the chi-square statistic is a big value.

The trick, then, is to decide what values of the chi-square statistic are "big." Big values discredit the null hypothesis. To determine whether an observed value is big, we need to know its probability distribution when the null hypothesis is true.



Conting...

Options

Contingency table results:

Rows: Partner Abuse

Columns: TV Violence

Cell format
Count Expected count

	High TV Violence	Low TV Violence	Total
Abusive	25 16.45	57 65.55	82
Not Abusive	41 49.55	206 197.4	247
Total	66	263	329

Chi-Square test:

Statistic	DF	Value	P-value
Chi-square	1	7.404997	0.0065

◀ **FIGURE 10.1** StatCrunch output for TV violence and abusiveness. The expected values are below the observed values.

Tech

KEY POINT If the data conform to the null hypothesis, then the value of the chi-square statistic will be small. For this reason, large values of the chi-square statistic make us suspicious of the null hypothesis.

Caution

Symbols
The symbol X^2 is used to represent the chi-square *statistic*. The symbol χ^2 is used to represent the chi-square *distribution*. Do not confuse the two. One is a statistic whose value is based on data; the other provides (approximate) probabilities for the values of that statistic.

4. Finding the p-Value for the Chi-Square Statistic

We are wondering whether television-viewing habits as a child are associated with violent behavior as an adult. If there is *no* association, then the observed counts in Table 10.7 should be close to the expected counts, and our chi-square statistic should be small. We found that $X^2 = 7.40$. Is this small?

To help us determine whether a particular value for a chi-square statistic is big or small, we need to know the probability distribution of X^2. Recall that the probability distribution of a statistic, such as X^2, is called a sampling distribution. Using the sampling distribution for X^2, we can find the probability that, just through chance, a chi-square statistic would have the value of 7.40 or bigger. If this probability is large, it means that 7.40 is not an unusual value. If the probability is small, it suggests that the value might be unusually large.

If the sample size is large enough, there is a probability distribution that gives a fairly good approximation to the sampling distribution. Not surprisingly, this approximate distribution is called the **chi-square distribution**. The chi-square distribution is often represented with the Greek lowercase letter chi (χ) raised to the power of 2—that is, χ^2.

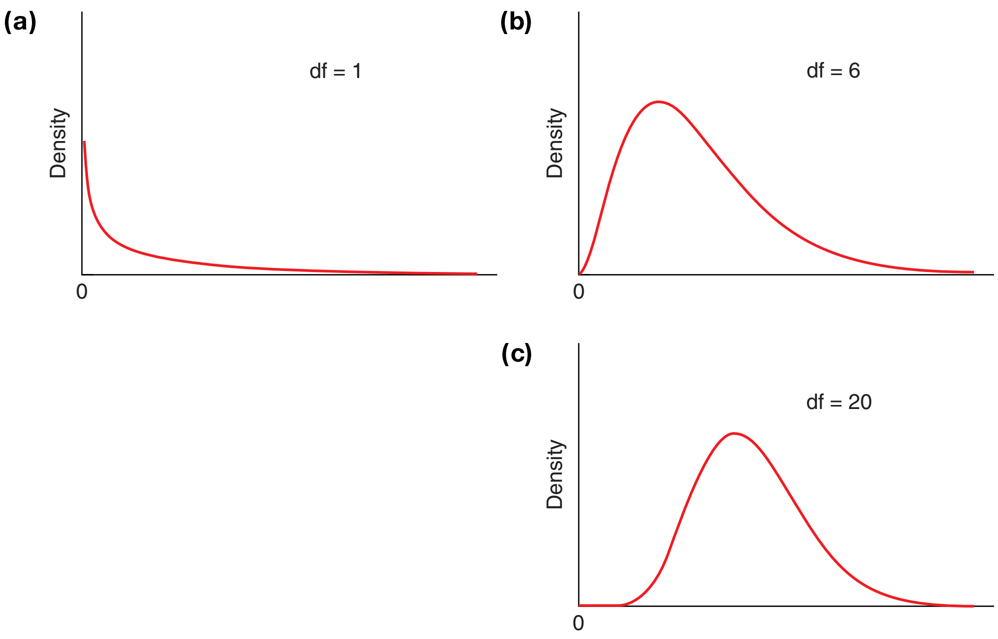
Unlike the normal distribution and the t-distribution, the χ^2 distribution allows for only positive values. It also differs from the other sampling distributions you've seen in that it is (usually) not symmetric and is instead right-skewed.

Details

Degrees of Freedom (df)
The degrees of freedom for the chi-square distribution are determined by the number of categories, not by the number of observations.

Like the shape of the t-distribution, the shape of the chi-square distribution depends on a parameter called the **degrees of freedom**. The lower the degrees of freedom, the more skewed the shape of the chi-square distribution. Figure 10.2 shows the chi-square distribution for several different values of the degrees of freedom. Sometimes the degrees of freedom are indicated using this notation: χ^2_{df}. For example, χ^2_6 represents a chi-square distribution with 6 degrees of freedom, as in Figure 10.2b.

The degrees-of-freedom parameter is different for different tests, but in general it depends on the number of categories in the summary table.



▶ **FIGURE 10.2** Three chi-square distributions for **(a)** df = 1, **(b)** df = 6, and **(c)** df = 20. Note that the shape becomes more symmetric as the degrees of freedom (df) increase. No negative values are possible with the chi-square distribution; the smallest possible value is 0.

The chi-square distribution χ^2 is only an approximation of the true sampling distribution of the statistic X^2. The approximation is usually quite good if all of the expected counts are 5 or higher.

KEY POINT The chi-square distribution provides a good approximation to the sampling distribution of the chi-square statistic only if the sample size is large. For many applications, the sample size is large enough if each expected count is 5 or higher.

We asked whether the value $X^2 = 7.40$ was small for the problem of determining whether viewing TV violence was associated with abusiveness later in life. We'll answer this question in Section 10.2.

SECTION 10.2

Chi-Square Tests for Associations between Categorical Variables

There are two tests to determine whether two categorical variables are associated. Which test you use depends on how the data were collected.

We often summarize two categorical variables in a two-way table, in part because it helps us see whether associations exist between these variables. When we examine a two-way table (or any data summary, for that matter), one aspect that gets hidden is the method used to collect the data. We can collect data that might appear in a two-way table in either of two ways.

The first method is to collect two or more distinct, independent samples, one from each population. Each object sampled has a categorical value that we record. For example, we could collect a random sample of men and a distinct random sample of women. We could then ask them to what extent they agree with the statement that same-sex marriage should be allowed: Strongly Agree, Agree, Neutral, Disagree, or Strongly Disagree. We now have one categorical response variable, *Opinion*. We also have another categorical variable, *Gender*, that keeps track of which population the response belongs to. Hence we have two samples, one categorical response variable, and one categorical grouping variable.

The second method is to collect just one sample. For the objects in this sample, we record two categorical response variables. For example, we might collect a large sample of people and record their marital status (single, married, divorced, or widowed) and their educational level (high school, college, graduate school). From this one sample, we get two categorical variables: *Marital Status* and *Educational Level*.

In both data collection methods, we are interested in knowing whether the two categorical variables are related or unrelated. However, because the data collection methods are different, the ways in which we test the relationship between variables differ also. That's the bad news. The good news is that this difference is all behind the scenes. The result of careful calculation shows us that no matter which method we use to collect data, we can use the same chi-square statistic and the same chi-square distribution to test the relation between variables.

These two methods have different names. If we test the association, based on two independent samples, between the grouping variable and the categorical response variable (the first method), the test is called a test of **homogeneity**. If we base the test on one sample (the second method), the test is called a test of **independence**. Two different data collection methods, two different names—but, fortunately, the same test!

Details

Homogeneity
The word *homogeneity* is based on the word *homogeneous*, which means "of the same, or similar, kind or nature."

KEY POINT There are two tests to determine whether two categorical variables are associated. For two or more samples and one categorical response variable, we use a test of homogeneity. For one sample and two categorical response variables, we use a test of independence.

EXAMPLE 3 Independence or Homogeneity?

Perhaps you've taken an online class or two. What are your feelings about the value of online classes? The Pew Foundation surveyed two distinct groups: the general American public and presidents of colleges and universities. About 29% of the sample from the general public said that online courses "offer an equal value compared with courses taken in the classroom." By comparison, over half of the sample of college presidents felt that online courses and classroom courses had the same value. The Pew Researchers carried out a hypothesis test to determine whether the variables *Online Course Value* (which recorded whether a respondent agreed or disagreed that online courses were equivalent to classroom courses) and *President* (which recorded whether or not a respondent was the president of a college or university).

QUESTION Will this be a test of independence or of homogeneity?

SOLUTION The Pew Foundation took two distinct samples of people: presidents and the general public. The variable *President* simply tells us which group a person belongs to. There is only one response variable: *Online Course Value*.

CONCLUSION This is a test of a homogeneity.

TRY THIS! Exercise 10.11

EXAMPLE 4 Independence or Homogeneity?

Do movie critics' opinions align with the public's? Even though the answer may be an obvious "no" to you, it is nice to know that we can examine data to answer this question. We took a random sample of about 200 movies from the Rotten Tomatoes website. Rotten Tomatoes keeps track of movie reviews and summarizes the general opinion about each movie. The critics' summary opinion is given as (from Best to Worst) "Certified Fresh," "Fresh," or "Rotten." The audience general opinions are classified as "Upright" (which is good) or "Spilled" (which is bad.) Thus, for each movie in our sample, we have two variables: *Critics' Opinion* and *Audience Opinion*.

QUESTION If we test whether there is an association between Critics' Opinion and Audience Opinion, is this a test of homogeneity or of independence?

CONCLUSION There is one sample, consisting of about 200 movies. Each movie provides two responses: a critics' opinion and an audience opinion. This is a test of independence.

TRY THIS! Exercise 10.13

Tests of Independence and Homogeneity

Again, the tests follow the four-step procedure of all hypothesis tests. We'll give you an overview and then fill in the details with an example.

Step 1: Hypothesize

The hypotheses are always the same.

H_0: There is *no* association between the two variables (the variables are independent).

H_a: There is an association between the two variables (the variables are not independent).

Although the hypotheses are always the same, you should phrase these hypotheses in the context of the problem.

Step 2: Prepare
Whether you are testing independence or homogeneity, the test statistic you should use to compare counts is the chi-square statistic, shown in Formula 10.2 and repeated here.

$$X^2 = \sum_{\text{cells}} \frac{(O - E)^2}{E}$$

If the conditions are right, then this statistic follows, approximately, a chi-square distribution with

$$\text{df} = (\text{number of rows} - 1)(\text{number of columns} - 1)$$

Conditions:

1. *Random Samples.* All samples were collected randomly.
2. *Independent Samples and Observations.* All samples are independent of each other. Always, the observations within a sample must be independent of each other.
3. *Large Samples.* The expected count must be 5 or more in each cell.

Note that in a test of independence, there is always only one sample. But a test of homogeneity might have several independent samples.

Step 3: Compute to compare
This step is best done with technology. The p-value is the probability, assuming the null hypothesis is true, of getting a value as large as or larger than the observed chi-square statistic. In other words, the p-value is the probability, if the variables really are not associated, that we would see an outcome as large or larger than the one observed. A small p-value therefore means a large test statistic, which casts doubt on the hypothesis that the variables are not associated.

Step 4: Interpret
If the p-value is less than or equal to the stated significance level, we reject the null hypothesis and conclude that the variables are associated.

EXAMPLE 5 Education and Marital Status

Does a person's educational level affect his or her decision about marrying? With observational data, we can't know for certain, but we can see whether the data are consistent with or contradict the idea that people's educational level affects their decisions about marrying. From the U.S. Census data, we took a random sample of 665 people and measured their marital status (single, married, divorced, or widow/widower) and their educational level (less than high school, high school degree, college degree or higher). Figure 10.3 shows output from StatCrunch.

QUESTION Use the provided output to test whether marital status and educational level are associated. Is this a test of homogeneity or of independence?

SOLUTION Because there is one sample with two response variables, this is a test of independence.

Step 1: Hypothesize

H_0: Among all U.S. residents, marital status and educational level are independent.

H_a: Among all U.S. residents, marital status and educational level are associated.

Step 2: Prepare
We can see from the output that the expected counts are 5 or more. (The smallest expected count is 11.47.) This means that the chi-square distribution will provide a good

approximation for the p-value. The data are summarized in Figure 10.3. This table displays the actual counts for a random sample of 665 people, the expected counts, and the calculated values.

▶ **FIGURE 10.3** Summary table with expected counts, chi-square statistic, and p-value to test whether marital status and educational level are associated.

Tech

Contingency Table with data

Options

Contingency table results:

Rows: marital

Columns: education

Cell format
Count Expected count

	College or higher	HS	Less HS	Total
Divorced	15 18.06	59 50.15	10 15.79	84
Married	98 87.74	240 243.6	70 76.69	408
Single	27 24.08	68 66.86	17 21.05	112
Widow/Widower	3 13.12	30 36.42	28 11.47	61
Total	143	397	125	665

Statistic	**DF**	**Value**	**P-value**
Chi-square	6	39.96996	<0.0001

Step 3: Compute to compare
The chi-square statistic is $X^2 = 39.97$ (rounding to two decimal digits). The chi-square distribution has 6 degrees of freedom, and the p-value is reported as smaller than 0.0001.

Step 4: Interpret
Because the p-value is less than 0.05, we reject the null hypothesis.

CONCLUSION Marital status and educational level are associated.

TRY THIS! Exercise 10.17

EXAMPLE 6 Hungry Monkeys

Research in the past has suggested that mice and rats that are fed less food live longer and healthier lives. Recently, a study of Rhesus monkeys was done that involved caloric restriction (less food). It is believed that monkeys have many similarities to humans, which is what makes this study so interesting.

Seventy-six Rhesus monkeys, all young adults, were randomly divided into two groups. Half of the monkeys (38) were assigned to caloric restriction. Their food was decreased about 10% per month for three months, and as a result they were fed about 30% less food than the other 38 monkeys for the duration of the experiment.

For those on the normal diet, 14 out of 38 had died of age-related causes by the time the article was written. For those on caloric restriction, only 5 out of 38 had died of age-related causes (Colman et al. 2009).

QUESTION Because this is a randomized study, the hypothesis that diet is associated with aging can be stated as a cause-and-effect hypothesis. Therefore, test the hypothesis that diet causes differences in aging. Will this be a test of homogeneity or of independence? Minitab output is shown in Figure 10.4.

```
Chi-Square Test: normal diet, caloric Restriction
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

        Normal      Caloric
          Diet  Restriction  Total
Died        14            5     19
          9.50         9.50
         2.132        2.132

Not         24           33     57
         28.50        28.50
         0.711        0.711

Total       38           38     76

Chi-Sq = 5.684, DF = 1, P-Value = 0.017
```

◀ **FIGURE 10.4** Minitab output showing *Diet* (normal or caloric restriction) and *Aging* (died from age-related causes or not) for 76 monkeys.

Details

Random Assignment
Experiments that randomly assign subjects to treatment groups result in distinct, independent samples, so if other conditions are satisfied, they can be analyzed with tests of homogeneity.

SOLUTION There are two samples (monkeys with caloric restriction and monkeys without) and one outcome variable: whether the monkey died of age-related causes. Therefore, this is a test of homogeneity.

Step 1: Hypothesize

H_0: For these monkeys, the amount of calories in the diet is independent of aging.

H_a: For these monkeys, the amount of calories in the diet causes differences in aging. Because we do not have a random sample, our results do not generalize beyond this group of monkeys. But because we have randomized assignment to treatment groups, we are able to conclude that any differences we see in the aging process are caused by the diet.

Step 2: Prepare
Because we wish to use the chi-square test, we must confirm that the expected counts are all 5 or more. This is the case here, as Figure 10.4 confirms.

Step 3: Compute to compare
From the Minitab output, Figure 10.4, you can see that the chi-square value is 5.68 and the p-value is 0.017.

Step 4: Interpret
Because the p-value is less than 0.05, we can reject the null hypothesis.

CONCLUSION The monkeys' deaths from age-related causes were caused by differences in the number of calories in the diet.

TRY THIS! Exercise 10.23

Looking Back

Comparing Two Proportions
In Section 8.4, you learned to test two population proportions based on data from independent samples using the *z*-test. The test of diet in monkeys could have been done using this test. The chi-square test provides an alternative, but equivalent, approach.

The article contains other information that suggests that monkeys on a restricted diet are generally healthier than monkeys on a normal diet. Figure 10.5 shows a photo of two monkeys. The one on the right had the restricted diet and shows fewer characteristics of old age.

▶ **FIGURE 10.5** The healthier monkey (on the right) was one of those on caloric restriction.

Random Samples and Randomized Assignment

You have now seen randomization used in two different ways. Random sampling is the practice of selecting objects in our sample by choosing them at random from the population, as is done in many surveys. We can make generalizations about the population only if the sample is selected randomly, because this is the only way of ensuring that the sample is representative of the population. The General Social Survey is an example of studies based on random sampling. When we conclude, as we did in Example 5, that marital status and educational level are associated, we are stating a conclusion about the entire population—in this case, all adults in the United States. We are confident that these variables are associated in the population, because our sample was selected at random.

In Example 6, on the other hand, there was no random sample. However, the monkeys were randomly assigned to a treatment group (low-calorie diet) or the control group (normal diet). Because the monkeys were not selected randomly, we have no means of generalizing about the population as a whole, statistically speaking. (There might be a biological argument, or an assumption, that a diet that works on one group of monkeys would work on any other group, but as statisticians, we have no data to support this assumption.) However, the researchers performed this study because they were interested in a cause-and-effect relationship: Does changing the calories in a monkey's diet change the monkey's health and longevity?

Because researchers controlled which monkeys got which diet, this is a controlled experiment. And because they used random assignment, and because we rejected the null hypothesis that diet and health were independent, we can conclude that in fact the caloric restriction *did* affect the monkeys' health.

Looking Back

Data Collection

Controlled experiments are those in which experimenters determine how subjects are assigned to treatment groups. In contrast, observational studies are those in which subjects place themselves into treatment groups, by behavior or innate characteristics such as gender. Causal conclusions cannot be based on a single observational study.

SNAPSHOT CHI-SQUARE TESTS OF INDEPENDENCE AND HOMOGENEITY

WHAT IS IT? ▶ A test of whether two categorical variables are associated.

WHAT DOES IT DO? ▶ Using the chi-square statistic, we compare the observed counts in each outcome category with the counts we would expect if the variables were *not* related. If observations are too different from expectations, then the assumption that there is no association between the categorical variables looks suspicious.

HOW DOES IT DO IT? ▶ If the sample size is large enough and basic conditions are met, then the chi-square statistic follows, approximately, a chi-square distribution with df = (number of rows − 1) × (number of columns − 1). The p-value is the probability of getting a value as large as or larger than the observed chi-square statistic, using the chi-square distribution.

HOW IS IT USED? ▶ To compare distributions of two categorical variables.

Relation to Tests of Proportions

In the special case in which both categorical variables have only two categories, the test of homogeneity is identical to a z-test of two proportions, using a two-sided alternative hypothesis. The following analysis illustrates this.

In a landmark study of a potential AIDS vaccine published in 2009, researchers from the U.S. Army and the Thai Ministry of Health randomly assigned about 8200 volunteers to receive a vaccine against AIDS and another 8200 to receive a placebo. (We rounded the numbers slightly to make this discussion easier.) Both groups received counseling on AIDS prevention measures and were promised life-time treatment should they contract AIDS. Of those who received the vaccine, 51 had AIDS at the end of the study (three years later). Of those that received the placebo, 74 had AIDS (http://www.hivresearch.org/, accessed September 29, 2009). We will show two ways of testing whether an association existed between receiving the vaccine and getting AIDS. The data are summarized in Table 10.8.

	Vaccine	No Vaccine	Total
AIDS	51	74	125
No AIDS	8149	8126	16275
Total	8200	8200	16400

◀ **TABLE 10.8** Effect of the vaccine on AIDS.

If we use the approach of this chapter, we recognize that this is a test of homogeneity, because there are two samples (*Vaccine* and *Placebo*) and one outcome variable (*AIDS*). Although we cannot generalize to a larger population (because the volunteers were not randomly selected), we can make a cause-and-effect conclusion about whether differences in AIDS rates are due to the vaccine, because this is a controlled, randomized study.

As a first step, we calculate the expected counts, under the assumption that the two variables are not associated.

Because the proportion of those who got AIDS was $125/16400 = 0.007622$, if the risk of getting AIDS had nothing to do with the vaccine, then we should see about the same proportion of those getting AIDS in both groups. If the proportion of people who got AIDS in the *Vaccine* group was 0.007622, then we would expect $8200 \times 0.007622 = 62.5$ people to get AIDS in the *Vaccine* group.

Both groups are the same size, so we would expect the same number of AIDS victims in the *Placebo* group. This means that in both groups, we would expect $8200 - 62.5 = 8137.5$ not to get AIDS.

The results, with expected counts in parentheses to the right of the observed counts, are shown in Table 10.9.

▶ **TABLE 10.9** Expected counts, assuming no association between variables, are shown in parentheses.

	Vaccine	No Vaccine	Total
AIDS	51 (62.5)	74 (62.5)	125
No AIDS	8149 (8137.5)	8126 (8137.5)	16275
Total	8200	8200	16400

Note that all expected counts are much greater than 5.

The chi-square statistic is not difficult to calculate:

$$X^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

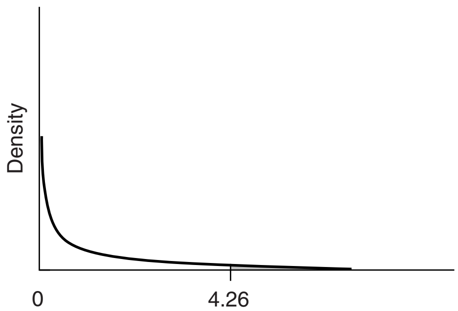
$$= \frac{(51 - 62.5)^2}{62.5} + \frac{(74 - 62.5)^2}{62.5} + \frac{(8149 - 8137.5)^2}{8137.5} + \frac{(8126 - 8137.5)^2}{8137.5}$$

$$= 4.26$$

The degree of freedom of the corresponding chi-square distribution is

$$(\text{Number of rows} - 1)(\text{number of columns} - 1) = (2 - 1)(2 - 1) = 1 \times 1 = 1$$

The p-value is illustrated in Figure 10.6. It is the area under a chi-square distribution with 1 degree of freedom and to the right of 4.26. The p-value turns out to be 0.039. We therefore reject the null hypothesis and conclude that there is an association between getting the vaccination and contracting AIDS. The difference in the numbers of AIDS victims was caused by the vaccine.



▶ **FIGURE 10.6** The area to the right of 4.26 represents the p-value to test whether there is an association between receiving the AIDS vaccine and contracting AIDS. The distribution is a chi-square distribution with 1 degree of freedom. The p-value is 0.039.

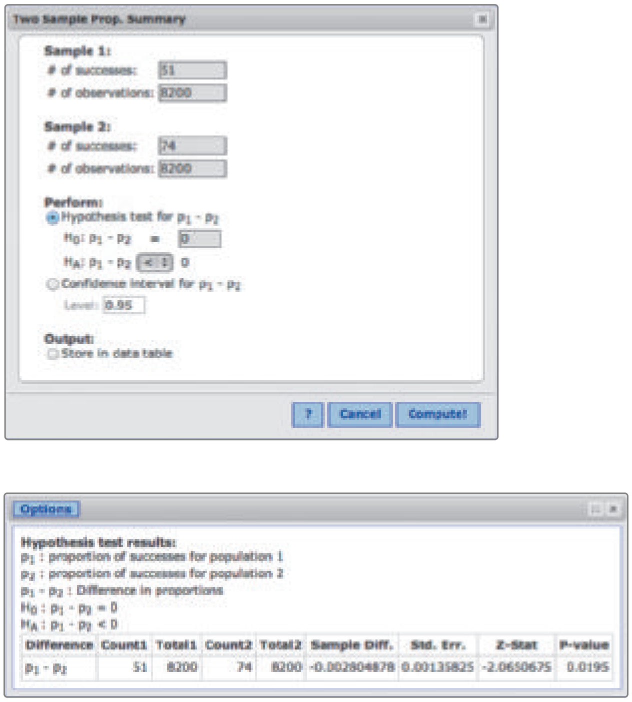
One thing that's disappointing about this conclusion is that the alternative hypothesis states only that the variables are associated. That's nice, but what we really want to know is *how* they are associated. Did the vaccine decrease the number of people who got AIDS? That's what the researchers wanted to know. They didn't want to know merely whether there was an association. They had a very specific direction in mind for this association.

One drawback with chi-square tests is that they reveal only whether two variables are associated, not how they are associated. Fortunately, when both categorical variables have only two categories, we can instead do a two-proportion z-test.

By doing a two-proportion z-test, we can test for the *direction* of the effect: whether the vaccine improved AIDS infection rates. We'll use Figure 10.7, which shows StatCrunch input and output, to test this hypothesis using a two-proportion z-test.

(a)

(b)



◀ **FIGURE 10.7** StatCrunch enables us to carry out a two-proportion *z*-test by entering the summary information as shown in part **(a)**. The resulting output is shown in part **(b)**.

Step 1: Hypothesize

H_0: $p_1 - p_2 = 0$ (*or* $p_1 = p_2$)

H_a: $p_1 - p_2 < 0$ (*or* $p_1 < p_2$)

In words, our null hypothesis is that the proportion of AIDS victims will be the same in the vaccine and the placebo groups. The alternative hypothesis states that the proportion of AIDS victims will be lower in the vaccine group.

Step 2: Prepare

The conditions are satisfied for a two-proportion *z*-test. (We'll leave it to you to confirm this, but it is important to note that because of the random assignment—the vaccine was randomly assigned to some subjects and a placebo to others—the two samples are independent of each other.)

Step 3: Compute to compare

The observed value of the *z*-statistic is −2.07, with a p-value of 0.0195.

Step 4: Interpret

Based on the p-value, we reject the null hypothesis and conclude that people who receive the vaccine are less likely to contract AIDS than those who do not. Because this was a randomized, controlled study, we can conclude that the vaccine caused a decrease in the proportion of people getting AIDS.

Note that if we had doubled the p-value—that is, if we had instead used a two-sided alternative hypothesis—we would have got $2 \times 0.019458 = 0.039$, exactly what we got for the chi-square test.

Although medical researchers were very excited about this study—not too many years before, it was thought to be nearly impossible to develop an AIDS vaccine—they caution that this vaccine lowers risk by only about 30%. Many medical professionals consider a vaccine to be useful only if it lowers risk by at least 60%. There was also some controversy over this study: Some argued that a few of the subjects who were dropped from the initial analysis (because of preexisting medical conditions) should have been included. If these subjects had been included, then the vaccine would no longer have been judged effective, statistically speaking.

KEY POINT For a 2-by-2 contingency table of counts, a two-proportion z-test with a two-sided alternative is equivalent to a test of homogeneity.

When should you use the two-proportion z-test, and when should you use the test for homogeneity? If you need to use a one-sided alternative hypothesis, then you should use the z-test. However, if you plan to use a two-sided alternative hypothesis, then it doesn't matter which test you use.

SECTION 10.3

Reading Research Papers

One goal of this text is to teach you enough of the basic concepts of statistics that you can critically evaluate published research. The medical literature, in particular, records many findings that can have major consequences in our lives. When we rely on the popular media to interpret these findings, we often get contradictory messages. However, you now know enough statistics so that, with a few guiding principles, you can often make sense of research results yourself.

Before discussing ways of evaluating individual research articles, we offer a few over-arching, guiding principles.

1. *Pay attention to how randomness is used.* Random sampling is used to obtain a representative sample, so that we can make inferences about a larger population. Random assignment is used to test causal associations so that we can conclude that the treatment was truly effective (or was not) on a particular sample of subjects. Many medical studies use random assignment but do not use random sampling. This means the results are not necessarily applicable to the entire population. Table 10.10 summarizes these possibilities.

TABLE 10.10 Four different study designs and the inferences possible.

	Sample Selected Randomly	No Random Sample
Random Assignment	You can make a causal conclusion and conclude that the entire population would be affected similarly.	You can make a causal conclusion, but we do not know whether everyone would respond similarly.
No Random Assignment	You can assume that an association between the variables exists in the population, but you cannot conclude that it is a causal relationship.	You can conclude that an association exists within the sample but not in the entire population, and you cannot conclude that there is a causal relationship.

2. *Don't rely solely on the conclusions of any single paper.* Research advances in small steps. A single research study, even when the conclusions are grand and ambitious, can tell us only a small part of the real story. For example, the *Los Angeles Times* reported, on the basis of a few published studies, that many people considered vitamin D to be useful for preventing cancer, cardiovascular disease, depression, and other maladies. But a panel of medical experts concluded that these beliefs were based on preliminary studies. The body of medical research, they concluded, was in fact quite mixed in its view of just how effective vitamin D really is (Healy 2011).

3. *Extraordinary claims require extraordinary evidence.* This advice was a favorite piece of wisdom of magician and professional skeptic (The Amazing) James Randi. If someone claims to have done something that had been believed impossible, don't believe it until you've seen some very compelling evidence.

4. *Be wary of conclusions based on very complex statistical or mathematical models.* Complex statistics often require complex assumptions, and one consequence of this is that the findings might be correct and true, but only in a limited set of circumstances. Also, some research studies are essentially "what if" studies: "We're studying *what* will happen *if* these assumptions hold." It is important to remember that those assumptions may not be practical—or even possible.

5. *Stick to peer-reviewed journals.* **Peer review** means that papers were read by two or three (and sometimes more) knowledgeable and experienced researchers in the same field. These reviewers can prevent papers from being published if they do not think the methods employed were sound or if they find too many mistakes. They can also demand that the authors make changes and submit the paper again. Many published papers have gone through several rounds of reviews. But be careful: Not all peer-reviewed journals are equal. The more prestigious journals have greater resources to check papers for mistakes, and they have much more careful and knowledgeable reviewers on hand to find sometimes subtle errors. Good journals have editorial boards that reflect the general practice of the community and are not restricted to people who reflect a minority point of view.

EXAMPLE 7 Improving Tips

A sociologist wonders whether a waitress's tip is affected by whether or not she touches her customers. Below are two study designs that the sociologist might use. Read these, and then answer the questions that follow.

Design A: The sociologist chooses four large restaurants in his city and gets all waitresses at these restaurants to agree to participate in his study. On several nights during the next few weeks, the sociologist visits the restaurants and records the number of times each waitress touches her customers on the customers' back or shoulder. He also records the total amount of tips earned by each waitress. He finds that waitresses who sometimes touched their customers earned larger tips, on average, than those who did not.

Design B: The sociologist chooses four large restaurants in his city and gets all waitresses at these restaurants to agree to participate in his study. At each restaurant, half of the waitresses are randomly assigned to either the "touch" group or the "no touch" group. The "touch" group is instructed to lightly touch each customer on the back or shoulder two or three times during the meal. The "no touch" group is instructed not to touch customers at all. At the end of the study, the sociologist finds that waitresses in the "touch" group earned larger tips, on average, than those in the "no touch" group.

QUESTION For each study design, state whether it is possible to generalize the results to a larger population and whether the researcher can make a cause-and-effect conclusion.

SOLUTION For Design A, the waitresses (or restaurants) were not randomly selected from a larger population, so we cannot generalize the results beyond the sample. This was an observational study: The waitresses themselves chose whether to be "touchers" or "no touchers." Because random assignment was not used, we cannot conclude that the difference in tips was caused by touching. A possible confounding effect is the restaurant itself, which might encourage touching and which might have a clientele that tends to tip more generously than the clientele at other restaurants.

For Design B also, we cannot generalize to a larger population. But here random assignment was used. Because random assignment was used, we can conclude that the touching caused the increased amount of tips.

TRY THIS! Exercise 10.33

Reading Abstracts

An **abstract** is a short paragraph at the beginning of a research article that describes the basic findings. If you look up science papers using Google Scholar, say, clicking on the link will usually take you to an abstract. Often, you will be able to read the abstract at no charge. (Reading the papers themselves often requires that you subscribe to the journal or that you read from the computers at a library that has purchased a subscription.) In addition, the websites of many journals display the abstracts to their published papers, even if they do not provide access to the papers themselves.

For example, an article published in the *New England Journal of Medicine* (Poordad et al. 2011) reported the findings of a study on the effectiveness of a new treatment for chronic hepatitis C virus (HCV). Currently, the standard treatment, peginterferon-ribavirin, has a relatively low success rate. A success, in treating HCV, is called a "sustained virologic response." In a sustained virologic response, the virus is not eliminated from the body, but is undetectable for a long period of time. Researchers hope that adding a new medicine, boceprevir, to the standard treatment would lead to a great proportion of patients achieving a sustained virologic response. An excerpt from the abstract follows:

> **Methods:** We conducted a double-blind study in which previously untreated adults with HCV genotype 1 infection were randomly assigned to one of three groups. In all three groups, peginterferon alfa-2b and ribavirin were administered for 4 weeks (the lead-in period). Subsequently, group 1 (the control group) received a placebo plus peginterferon-ribavirin for 44 weeks; group 2 received boceprevir plus peginterferon-ribavirin for 24 weeks, and those with a detectable HCV RNA level between weeks 8 and 24 received a placebo plus peginterferon-ribavirin for an additional 20 weeks; and group 3 received boceprevir plus peginterferon-ribavirin for 44 weeks. Nonblack patients and black patients were enrolled and analyzed separately.
>
> **Results:** A total of 938 nonblack and 159 black patients were treated. In the nonblack cohort, a sustained virologic response was achieved in 125 of the 311 patients (40%) in group 1, in 211 of the 316 patients (67%) in group 2 ($P < 0.001$), and in 213 of the 311 patients (68%) in group 3 ($P < 0.001$). In the black cohort, a sustained virologic response was achieved in 12 of the 52 patients (23%) in group 1, in 22 of the 52 patients (42%) in group 2 ($P = 0.04$), and in 29 of the 55 patients (53%) in group 3 ($P = 0.004$). In group 2, a total of 44% of patients received peginterferon-ribavirin for 28 weeks. Anemia led to dose reductions in 13% of controls and 21% of boceprevir recipients, with discontinuations in 1% and 2%, respectively.
>
> **Conclusions:** The addition of boceprevir to standard therapy with peginterferon-ribavirin, as compared with standard therapy alone, significantly increased the rates of sustained virologic response in previously untreated adults with chronic HCV genotype 1 infection. The rates were similar with 24 weeks and 44 weeks of boceprevir. (Funded by Schering-Plough [now Merck]; SPRINT-2 ClinicalTrials.gov, number NCT00705432.)

To help you evaluate this abstract, and others like it, answer these questions:

1. What is the research question that these investigators are trying to answer?
2. What is their answer to the research question?
3. What were the methods they used to collect data?
4. Is the conclusion appropriate for the methods used to collect data?
5. To what population do the conclusions apply?
6. Have the results been replicated (that is, reproduced) in other articles? Are the results consistent with what other researchers have suggested?

The sixth question is important, although it is often difficult for a layperson to answer. Research is a difficult activity, in part because of the great variability in nature. Studies are often done with samples that do not allow generalizing to populations or are subject to bias, and sometimes researchers just make mistakes. For that reason, you should not believe in the claims of a single study unless it is consistent with currently accepted theory and supported by other research. For all of these reasons, wait until there is some accumulation of knowledge before changing your lifestyle or making a major decision on the basis of scientific research.

Let's see how we would answer these questions for this abstract.

1. *What is the research question that these investigators are trying to answer?* Does the addition of boceprevir to the standard treatment lead to a higher proportion of hepatitis C patients achieving a sustained virologic response?
2. *What is their answer to the research question?* Yes, the additional drug improved patients' responses. Researchers report an increased rate of patients achieving sustained virologic response with boceprevir plus the standard therapy compared to those receiving only standard therapy.
3. *What were the methods they used to collect data?* Patients were randomly assigned to one of three treatment groups. Assignments were double-blind. Patients were examined after 24 and 44 weeks to determine if they had achieved sustained virologic response.
4. *Is the conclusion appropriate for the methods used to collect data?* The fact that subjects were randomly assigned to treatment groups means that we can make a causal conclusion and say that the difference in virologic response rates was due to the the addition of boceprevir to the standard treatment. We also know that the observed results were too large to be explained by chance. Note that p-values are given as, for example, "$P < 0.001$". For instance, we know that group 2 (nonblack cohort) had a greater rate of sustained virologic response than the treatment group, with p-value < 0.001.
5. *To what population do the conclusions apply?* The sample was not randomly selected, so although we *can* conclude that the treatment was effective, we do not know if we would see the same sized effect on other samples or in the population of all hepatitis C patients. Because it is known that African Americans and non-African Americans have different responses to the standard treatment, these two groups were treated separately in the analysis.
6. *Have the results been replicated in other articles? Are the results consistent with what other researchers have suggested?* We cannot tell from the abstract.

Note that even though we don't know which statistical procedures were carried out (we could learn this from reading the article itself), we still have a good sense of the reliability of the study. We know that the reported success rates for standard treatment for hepatitis C (40% in the nonblack cohort and 23% in the black cohort) were lower than for the treatment that included boceprevir (67% for nonblack cohort after 24 weeks, 68% after three weeks; 42% for the black cohort after 24 weeks, and 53% after 44 weeks.) We know this difference cannot be plausibly assigned to chance, and we know that, because of random assignment, that boceprevir was the cause of the difference in rates.

EXAMPLE 8 Brain Games

Read the following abstract, which discusses the effects of "brain games" on elderly Americans' cognitive functioning (Rebok et al. 2014).

Objectives: To determine the effects of cognitive training on cognitive abilities and everyday function over 10 years. **Design**: Ten-year follow-up of a randomized, controlled single-blind trial (Advanced Cognitive Training for Independent and Vital Elderly (ACTIVE)) with three intervention groups and a no-contact control group. **Participants**: A volunteer sample of 2,832 persons (mean baseline age 73.6; 26% African American) living independently. **Intervention**: Ten training sessions for memory, reasoning, or speed of processing; four sessions of booster training 11 and 35 months after initial training. **Measurements**: Objectively measured cognitive abilities and self-reported and performance-based measures of everyday function. **Results**: Participants in each intervention group reported less difficulty with instrumental activities of daily living (IADLs). At a mean age of 82, approximately 60% of trained participants, versus 50% of controls ($P < 0.05$), were at or above their baseline level of self-reported IADL function at 10 years. The reasoning and speed-of-processing interventions maintained their effects on their targeted cognitive abilities at 10 years. Memory training effects were no longer maintained for memory performance. Booster training produced additional and durable improvement for the reasoning intervention for reasoning performance and the speed-of-processing intervention for speed-of-processing performance. **Conclusion**: Each Advanced Cognitive Training for Independent and Vital Elderly cognitive intervention resulted in less decline in self-reported IADL compared with the control group. Reasoning and speed, but not memory, training resulted in improved targeted cognitive abilities for 10 years.

QUESTION Write a short paragraph describing this research. Use the six questions above as guidance.

SOLUTION The study attempts to determine whether cognitive training (which isn't defined in the abstract) can improve cognitive functions such as reasoning, speed of processing, and memory, over a 10-year period. The researchers randomly assigned subjects to one of three different training programs or to a control group that received no training. They concluded that, in fact, the training does improve reasoning and speed of processing, but not memory. They also found that those who received cognitive training had less difficulty with daily activities (instrumental activities of daily living, or IADLs). The p-value for comparing the IADL scores of the control subjects to those of the subjects who received the cognitive training was reported as less than 0.05, so we know that at a 5% significance level, we are confident that the differences are not explained by chance. Because the researchers used random assignment, their conclusion that the differences between the intervention groups and the control groups were caused by the cognitive training is justified. The results apply to only those 2,832 people who participated, because this is not a random sample. All of the participants were older than the general population (their mean age was 73.6 years at the start of the study). The abstract does not tell us whether these results have been replicated.

TRY THIS! Exercise 10.49

Buyer Beware

Leo Tolstoy's *Anna Karenina* begins, "Happy families are all alike; every unhappy family is unhappy in its own way." To (very loosely) paraphrase the great Russian novelist, there are few ways that a study can be good, but there are many ways that studies can go wrong. We've given you some tips that should help you recognize a good study, but you should also be aware of some warning signs and features that indicate that a study might not be good.

Data Dredging Hypothesis testing is designed to test claims that result from a theory. The theory makes predictions about what we should see in the data; for example, students who write about their anxieties will do better on an exam, so the mean exam score of students who wrote about their anxieties should be greater than the mean score of those who did not. We next collect the data and then do a hypothesis test to determine whether the theory was correct. **Data dredging** is the practice of stating our hypotheses after first looking at the data. Data dredging makes it more likely that we will mistakenly reject the null hypothesis. Even when the null hypothesis is true, our data sometimes show surprising outcomes, just because of chance variation. If we first look at the test statistic to decide what the hypotheses should be, we are rigging the system in favor of the alternative hypothesis.

The situation is analogous to betting on a horse race after the race has begun. You are supposed to place a bet before the race starts, so that everyone is on equal footing. The odds on which horse will win (which determine the payoffs) are meant to estimate probabilities of a horse winning. But if you wait until the horses are running, you have a better chance than everyone else of correctly predicting which horse will win. You have "snooped" at the data to make your decision. Your probability of winning is not the same as what everyone else believed it to be.

Theories should be based on data, but the correct procedure is to use data to formulate a theory and then to collect additional data in an independent study to test that theory. If it is too costly to collect additional data, one common approach is to randomly split the data set into two (or even more) pieces. One piece of the data is then set aside—"locked away"—and the researchers are forbidden to look. The researchers can then examine the first piece as much as they want and use these data to generate hypotheses. After they have generated hypotheses, they can test these hypotheses on the second, "locked away" data set. Another possibility is to use the data for generating hypotheses and then go out and collect *more* data.

Publication Bias Most scientific and medical journals prefer to publish "positive" findings. A positive finding is one in which the null hypothesis is rejected (with the result that the researcher concludes that the tested treatment is effective). Some journals prefer this sort of finding, because these are generally the results that advance science. However, suppose a pharmaceutical company produces a new drug that it claims can cure depression, and suppose this drug does not work. If many researchers are interested in studying this drug and they do statistical tests with a 5% significance level, then about 5% of them will conclude that the drug is effective, even though we know it is not. If a journal favors positive findings over negative findings, then we will read only about the studies that find the drug works, even though the vast majority of researchers came to the opposite conclusion.

Publication bias is one reason why it is important to consider several different studies of the same drug before making decisions. A new, and somewhat controversial, form of statistical analysis called meta-analysis has been developed in recent years, in part to help with problems such as these. A **meta-analysis** considers all studies done to test a particular treatment and tries to reconcile different conclusions, attempting to determine whether other factors, such as publication bias, played a role in the reported outcomes.

Psychologists conducted a meta-analysis to conclude that violent video games are not associated with violent behavior in children, despite the fact that several studies had concluded otherwise (Ferguson and Kilburn 2010). They concluded that papers that found an association between video game violence and real-life violence were more likely to be published, so researchers who found no such associations were less likely to be published. This result is not likely to settle the controversy, but it points to a potential danger that arises when journals publish only positive findings.

Profit Motive Much scientific research is now sponsored by corporations that hope to establish that their products make life better for people. Researchers are usually required to disclose to a journal whether they are themselves making money off the drug or product that they are researching, but this does not tell us who funded the project in the first place. There is no reason to reject the conclusions of a study simply because it was paid for by a corporation or business or some other organization with a vested interest. You should always evaluate the methods of the study used and decide whether the methods are sound.

However, you should be aware that sometimes the corporation funding the research can influence whether results get published and which results get published. For example, a researcher might be funded by a pharmaceutical company to test a new drug. If he finds the drug doesn't work, the drug company might decide that it doesn't want to publicize this fact. Or perhaps the drug works on only a small subset of people. The company might then publicize that the drug is effective but fail to mention that it is effective on only a small group.

For example, a 2007 study concluded that playing "active" video games (such as the Wii) was healthier than playing "passive" games and found that children burned more calories playing the active games (Neale 2007). The study also suggested that playing real sports was much more healthful than playing any video game, but still the researchers conclude, "Nevertheless, new generation computer games stimulated positive activity behaviors. Given the current prevalence of childhood overweight and obesity, such positive behaviors should be encouraged." A press release notes that the study was funded by Cake, the "marketing arm" of Nintendo UK, which manufactures the Wii. This does not mean that the results of the study are wrong (the children playing the active games most likely did in fact burn more calories than those who played the passive games), but it may account for the positive spin despite the finding that active video games were not nearly as effective as playing sports at promoting weight loss.

Media The media—newspapers, magazines, television shows, and radio broadcasts—are also profit-motivated. A good journalist strives to get to the heart of the matter. Nonetheless, scientific and medical research findings are complex, and when condensing complex ideas into easily digested sound bites (and doing so in a way that entertains those who pay for the papers and magazines), the truth sometimes gets distorted.

The media often use catchy headlines, and these headlines do not always capture the true spirit of a study. The most common problem is that headlines often suggest a cause-and-effect relationship, even though the wise statistics student will quickly recognize that such a conclusion is not supported by the data. Statistician Jonathan Mueller keeps a website of such headlines. A couple of examples: "Studies say lots of candy could lead to violence." "Texting lifts crash risk by large margin." These headlines all suggest causality: Eating candy will make you violent. Texting will increase the risk of a car accident. But you now know that such a conclusion can be made only for controlled experiments, and even then we must be cautious. Ideally, you can learn what you need to know by reading the news article. But not always. The information you need to judge whether the conclusions of a study are strong is often missing from news reports.

Clinical Significance vs. Statistical Significance An outcome of an experiment or study that is large enough to have a real effect on people's health or lifestyle is said to have **clinical significance**. Sometimes, researchers discover that a treatment is statistically significant (meaning that the outcome is too large to be due to chance) but too small to be meaningful (so it is not clinically significant). Studies with very large sample size have large power and so are capable of detecting even very small differences between treatment groups. This does not mean the differences matter. For example, a drug might truly lower cholesterol levels, but not enough to make a real difference in someone's health. Or playing Wii might burn more calories, but maybe not enough for it to serve as a form of exercise. Treatments that cause meaningful effects are called clinically significant. Sometimes, statistically significant results are not clinically significant.

Imagine a rare disease that only 1 person in 10 million people gets. A controlled experiment finds that a new drug "significantly" reduces your risk of getting this disease; specifically, it cuts the risk in half. Here, *significantly* means that the reduction in risk is statistically significant. But given that the disease is so rare already, is it worth taking medicine to cut your risk from 1 in 10 million to 1 in 20 million? Particularly if the drug is expensive or has side effects, most people would probably decide that this treatment is not clinically significant.

CASE STUDY REVISITED

To better understand the ideal conditions for popping corn, experimenters designed a randomized, controlled study to observe how well the popcorn popped under different conditions. Of interest here was how the amount of oil affected the outcome. A summary of results was shown in the Case Study at the beginning of the chapter, where we observed that it looked as though using the maximum amount of oil did not work well, because so few popcorn bags were successful by the criteria adopted for success.

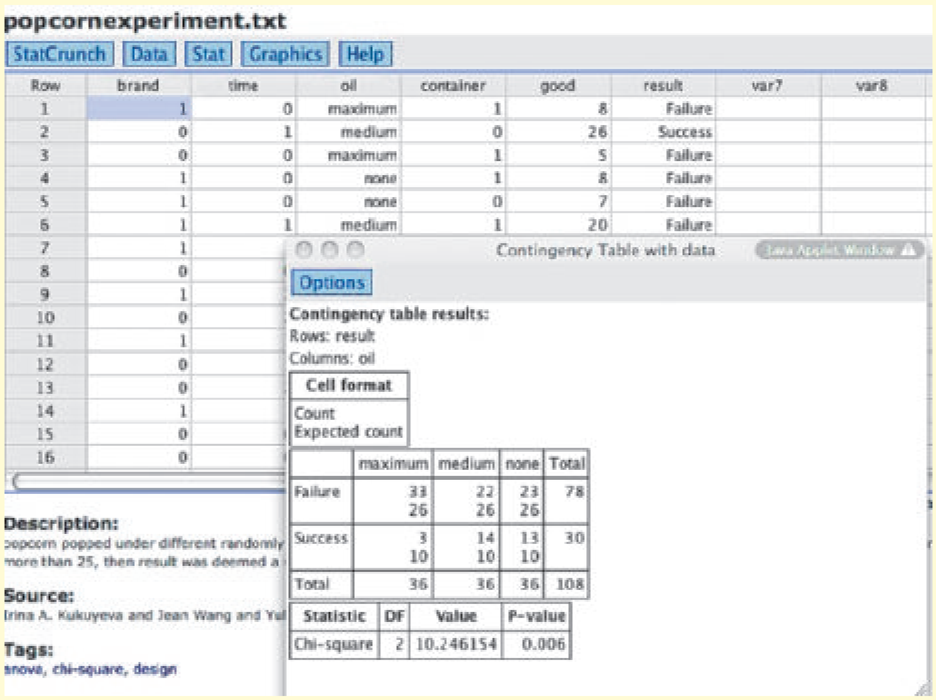
Did the amount of oil affect the resulting quality of the popcorn? To test this using the methods of this chapter, we carry out a test of homogeneity, because we have three independent samples (no oil, medium oil, and maximum oil) and one categorical response variable (*Result*: success or failure). We have three independent samples, because bags were randomly assigned to one of these three groups. Each group was assigned 36 bags of popcorn, and each bag had 50 kernels.

The hypotheses follow.

H_0: The quality of popcorn and the amount of oil are independent.
H_a: The amount of oil affects the quality of the popcorn.

The results (and some of the raw data) are shown in Figure 10.7 in output generated by StatCrunch. The output shows that the expected counts are all greater

▶ **FIGURE 10.8** StatCrunch output shows the results of the analysis in the foreground and the raw data in the background. The expected values in the table are given below the observed values.



than 5, so our sample sizes are large enough for the chi-square distribution (with 2 degrees of freedom) to produce a good approximation to the p-value.

From the output, we see that the test statistic has a value of 10.25, with a p-value of 0.006. This is quite small. Certainly it is less than 0.05, so at the 5% significance level we reject the null hypothesis. We conclude that quality is affected by the amount of oil used. (At least, it is if you believe that quality is measured by the number of kernels popped after 75 seconds.)

EXPLORING STATISTICS
CLASS ACTIVITY

Skittles

GOALS	MATERIALS
Apply a chi-square test to check whether two bags of Skittles candies contain the same proportions of colors.	• One small bag of Skittles for each student • Computer or TI-83/84

ACTIVITY

Open a bag of Skittles and count how many of each color are in your bag. Fill in the numbers in the table below. Then find a partner and fill in the colors from his or her bag. (If someone does not have a partner, form a group of three and use three rows.)

	Purple	Red	Orange	Yellow	Green	Total
Yours						
Partner's						

BEFORE THE ACTIVITY

1. Do you think that you and your partner(s) will get exactly the same number in each category?
2. Do you think you and your partner(s) will have significantly different distributions of colors? Why or why not?

AFTER THE ACTIVITY

1. Perform a hypothesis test to test whether the two bags have a significantly different distribution of colors, using a significance level of 0.05.
2. Throw away, save, or eat the Skittles.

CHAPTER REVIEW

KEY TERMS

two-way table, *502*
expected counts, *503*
chi-square statistic, *505*
chi-square distribution, *508*
degrees of freedom, *508*
homogeneity, *509*
independence, *509*
peer review, *519*
abstract, *520*
data dredging, *523*
publication bias, *523*
meta-analysis, *523*
clinical significance, *524*

LEARNING OBJECTIVES

After reading this chapter and doing the assigned homework problems, you should

- Distinguish between tests of homogeneity and tests of independence.
- Understand when it is appropriate to use a chi-square statistic to test whether two categorical variables are associated; know how to perform this test and interpret the results.
- Understand how random assignment is used to allow cause-and-effect inference, and understand how random sampling is used to allow generalization to a larger population.
- Be prepared to apply knowledge about collecting and analyzing data to critically evaluate abstracts in the science literature.

SUMMARY

We presented two types of tests for analyzing categorical variables. Although the test of homogeneity is conceptually different from the test of independence, they are exactly the same in terms of the calculations required. Both of these tests attempt to determine whether two categorical variables are associated. The only difference is in the way the data for the study were collected. When researchers collect two or more independent samples and measure one categorical response variable, they are performing a test of homogeneity. When instead they collect one sample and measure two categorical response variables, they are performing a test of independence.

Both tests rely on the chi-square statistic. For each cell of a two-way summary table, we compare the observed count with the count we would expect if the null hypothesis were true. If the chi-square statistic is big, it means that these two counts don't agree, and it discredits the null hypothesis.

An approximate p-value is calculated by finding the area to the right of the observed chi-square statistic using a chi-square distribution. To do this, you need to know the degrees of freedom for the chi-square distribution.

Many research questions can be divided into two categories: those that ask questions about causality and those that ask about associations between variables. Questions about causality can be answered only with controlled experiments, whereas observational studies can answer questions about associations.

Interpreting conclusions in scientific studies is complex, because many things can go wrong in a study. Remember that extraordinary results must be supported by extraordinary evidence, and that you should trust studies that have been replicated (repeated, resulting in the same conclusion) over studies that have not.

Formulas

Expected Counts

Formula 10.1: $$\text{Expected count for a cell} = \frac{(\text{row total}) \times (\text{column total})}{\text{grand total}}$$

Chi-Square Statistic

Formula 10.2: $$X^2 = \sum_{\text{cells}} \frac{(O - E)^2}{E}$$

Test of Homogeneity and Independence

Hypotheses

H_0: The variables are independent.
H_a: The variables are associated.

Conditions (Homogeneity)

1. *Random Samples*. Two or more samples, all sampled randomly.
2. *Independent Samples and Observations*. Samples are independent of each other. The observations within each sample are independent.
3. *Large Samples.* At least 5 expected counts in each cell of the summary table.

Conditions (Independence)

1. *Random Sample.* One sample, selected randomly.
2. *Independent Observations.* Observations are independent of each other.
3. *Large Sample.* There are at least 5 expected counts in each cell of the summary table.

Sampling Distribution
If conditions hold, the sampling distribution follows a chi-square distribution with degrees of freedom = (number of rows − 1) × (number of columns − 1).

SOURCES

Colman, R. J., et al. 2009. Caloric restriction delays disease onset and mortality in Rhesus monkeys. *Science* 325, 201.

Ferguson, C. J., and J. Kilburn. 2010. Much ado about nothing: The misestimation and overinterpretation of violent video game effects in Eastern and Western nations: Comment on Anderson et al. *Psychological Bulletin* 136(2), 174–178, discussion 182–187.

Healy, M. 2010. New vitamin D recommendations: What they mean. *Los Angeles Times,* December 6, www.latimes.com (accessed February 3, 2011).

Husemann, L. R., J. Moise-Titus, C. Podolski, and L. D. Eron. 2003. Longitudinal relations between children's exposure to TV violence and their aggressive and violent behavior in young adulthood: 1977–1992. *Developmental Psychology* 39(2), 201–221.

I. A. Kukuyeva, J. Wang, and Y. Yaglovskaya. 2008. Popcorn popping yield: An investigation, JSM

Neale, T. 2007. Even active video games not good enough for kids' fitness. MedPage Today, http://www.medpagetoday.com (accessed April 22, 2010).

Poordad, F., McCone, J., Bacon, B., Bruno, S., Manns, M., et al., 2011. Boceprevir for Untreated Chronis HCV Genotype 1 Infection. *New England Journal of Medicine,* v.364 (13) 1195–1206, http://www.nejm.org/doi/pdf/10.1056/NEJMoa1010494 (accessed April 26, 2011)

Rebok, G. W., K. Ball, L. Guey, R. Jones, H-Y. Kim, H. King, M. Marsiske, J. Morris, S. Tennstedt, F. Unverzagt, and S. Willis. 2014. Ten-year effects of the advanced cognitive training for independent and vital elderly cognitive training trial on cognition and everyday functioning in older adults. *Journal of the American Geriatrics Society* 62, 16–24.

SECTION EXERCISES

SECTION 10.1

10.1 Tests

a. In Chapter 8, you learned some tests of proportions. Are tests of proportions used for categorical or numerical data?

b. In this chapter, you are learning to use chi-square tests. Do these tests apply to categorical or numerical data?

10.2 In Chapter 9, you learned some tests of means. Are tests of means used for numerical or categorical data?

10.3 Gender and Toys A study was conducted by an organization that looked at children to understand whether there is an association between gender and types of toys chosen (mechanical toys or stuffed toys). Mechanical includes guns, cars, walkie-talkies, etc.; stuffed toys include teddy bears, plush puppets, sock monkeys, etc. The raw data are shown in the accompanying table; m for mechanical, s for stuffed toys, b for boy, and g for girl.

Gender	Toys
b	m
b	m
g	s
g	m
g	s
D	s
g	s
g	m
b	m
b	m
g	m
b	s
g	s
b	m
b	m

Gender	Toys
b	m
b	m
g	m
b	m
g	m
g	s
g	s
g	m
b	s
b	m
g	s
g	s
b	s
g	m
b	m

Gender	Toys
b	s
g	m
g	s
g	m
g	s
g	s
g	m
b	m
b	m
g	m
b	s
b	s
b	m
g	s
b	s

Create a two-way table to summarize these data. Notice that the two variables are categorical, as can be seen from the raw data. If you are doing this by hand, create a table with two rows and two columns. Label the columns Boys and Girls (across the top). Label the rows Mechanical Toys and Stuffed Toys. Begin with a big table, making a tally mark in one of the four cells for each observation, and then summarize the tally marks as counts.

10.4 Age and Traffic Rules The table shows the raw data for the results of a traffic survey of 24 people and whether they obeyed traffic rules or not. In the Age column, "Adult" means the person surveyed was an adult and "Not" means the person surveyed was not an adult. In the Rules column, "Yes" means the rules are followed and "No" means the rules are not followed.

Create a two-way table to summarize these data. Use Yes and No for the columns (across the top) and Adult or Not for the rows. (We gave you an orientation of the table so that your answers would be easy to compare.)

Are the two variables numerical or categorical?

Age	Rules
Adult	Yes
Not	Yes
Not	No
Not	Yes
Adult	No
Not	Yes
Adult	Yes
Not	No
Not	Yes
Adult	Yes
Not	Yes
Adult	Yes

Age	Rules
Adult	No
Not	No
Adult	Yes
Adult	No
Not	Yes
Not	No
Not	No
Adult	No
Adult	Yes
Adult	No
Not	No
Adult	Yes

10.5 The table summarizes the outcomes of a study that employees of an organization carried out to determine whether the employees in production department had a higher mean salary than the employees in sales department. Identify both of the variables, and state whether they are numerical or categorical. If numerical, state whether they are continuous or discrete.

	Mean Salary
Production	42,684
Sales	51,489

10.6 Playing Basketball There is a theory that playing basketball makes teenagers grow in height. The table shows a summary of the outcomes of a study that a researcher carried out to determine whether playing basketball makes one grow taller or not. Identify both of the variables, and state whether they are numerical or categorical. If numerical, state whether they are continuous or discrete.

	Tall	Short
Play Basketball	18	10
Do Not Play Basketball	4	12

TRY **10.7 Effects of Parental Education on Boys' Education (Example 1)** A study done by the *IZA Journal of Labour Economics* in the United Kingdom in 2013 compared 15-year-old boys who are either attending school or have dropped out, in order to understand the impact of parental education on them. The table shows the relationship between boys' education and parental education.

	Educated Parents	Uneducated Parents
Studying	42	23
Not Studying	14	8

a. Find the row, column, and grand totals, and prepare a table showing these values as well as the counts given.

b. Find the percentage of boys who are studying.

c. Find the expected number of boys having educated parents who would study, if the variables are independent. Multiply the proportion overall that were studying times the number of boys having educated parents. Do not round off to a whole number. Round to two decimal digits.

d. Find the other expected counts using your knowledge so that the expected counts must add to the row and column totals. Report them in a table with the same orientation as the one given for the data.

10.8 Effects of Parental Education on Girls' Education Refer to Exercise 10.7. This data table compares 15-year-old girls who are either attending school or have dropped out, in order to understand the impact of parental education on them. The table shows the relationship between girls' education and parental education.

	Educated Parents	Uneducated Parents
Studying	13	68
Not Studying	11	6

a. Find the row, column, and grand totals, and prepare a table showing these values as well as the counts given.

b. Find the percentage of girls who are studying.

c. Find the expected number of girls having educated parents who would study, if the variables are independent. Multiply the proportion overall that were studying times the number of girls having educated parents. Do not round off to a whole number. Round to two decimal digits.

d. Find the other expected counts. Report them in a table with the same orientation as the one for the data.

TRY **10.9 Effects of Parental Education on Boys' Education (Example 2)** Refer to Exercise 10.7. The data table compares 15-year-old boys who are either attending school or have dropped out, in order to understand the impact of parental education on them. Report the observed value of the chi-square statistic.

	Educated Parents	Uneducated Parents
Studying	42	23
Not Studying	14	8

10.10 Effects of Parental Education on Girls' Education Refer to Exercise 10.8. The data table compares 15-year-old girls who are either attending school or have dropped out, in order to understand the impact of parental education on them. Report the observed value of the chi-square statistic.

	Educated Parents	Uneducated Parents
Studying	13	68
Not Studying	11	6

SECTION 10.2

TRY **10.11 Party and Right Direction (Example 3)** Suppose a polling organization asks a random sample of people if they are Democrat, Republican, or Other and also asks them if they think the country is headed in the right direction or the wrong direction. If we wanted to test whether party affiliation and answer to the question were associated, would this be a test of homogeneity or a test of independence? Explain.

10.12 Antibiotic or Placebo A large number of surgery patients get infections after surgery, which can sometimes be quite serious. Researchers randomly assigned some surgery patients to receive a simple antibiotic ointment after surgery, others to receive a placebo, and others to receive just cleansing with soap. If we wanted to test the association between treatment and whether or not patients get an infection after surgery, would this be a test of homogeneity or of independence? Explain. (Source: Hospitals could stop infections by tackling bacteria patients bring in, studies find. *New York Times*, January 6, 2010.)

TRY **10.13 Jet Lag Drug (Example 4)** A recent study was conducted to determine whether the drug Nuvigil was effective at helping east-bound jet passengers adjust to jet lag. Subjects were randomly assigned either to one of three different doses of Nuvigil (low, medium, high) or to a placebo. Subjects were flown to France in a plane in which they could not drink alcohol or coffee or take sleeping pills, and then were examined in a lab where their state of wakefulness was measured and classified into categories (low, normal, alert). If we test whether treatments for jet lag are associated with wakefulness, are we doing a test of independence or of homogeneity? Explain. (Source: A drug's second act: Battling jet lag. *New York Times*, January 6, 2010.)

10.14 Most Important Problem A Gallup Poll in September 2013 asked people what they considered to be the most important problem in the United States today. The people were also classified by race. If we wanted to test whether there was an association between response to the question and race of the respondent, should we do a test of independence or of homogeneity?

10.15 Oil Leaders The table shows the world's five largest crude oil producing countries and their total oil output in percentage for the years 2011 and 2012 (www.whichcountry.co). Give two reasons why you should not do a chi-square test with these data.

	Russia	Saudi Arabia	United States	China	Canada
2011	14.05	13.09	12.23	5.15	4.54
2012	13.99	13.51	9.86	5.81	4.2

10.16 Literacy Rates World literacy rates for individuals of 15 years of age or older are given in the data table as a percentage. Give two reasons why a chi-square test is not appropriate for this set of data.

	Male	Female
2011	87.6	76.8
2012	88.4	75.5
2013	84.5	79.9

TRY g **10.17 Obesity and Marital Status (Example 5)** A study reported in the medical journal *Obesity* in 2009 analyzed data from the National Longitudinal Study of Adolescent Health. Obesity was defined as having a body mass index (BMI) of 30 or more. The research subjects were followed from adolescence to adulthood, and all the people in the sample were categorized in terms of whether they were obese and whether they were dating, cohabiting, or married. Test the hypothesis that relationship status and obesity are associated, using a significance level of 0.05. Can we conclude from these data that living with someone is making some people obese and that marrying is making people even more obese? Can we conclude that obesity affects relationship status? Explain. *See page 539 for guidance.*

	Dating	Cohabiting	Married
Obese	81	103	147
Not Obese	359	326	277

(Source: N. S. The and P. Gordon-Larsen. 2009. Entry into romantic partnership is associated with obesity. *Obesity* 17(7), 1441–1447.)

10.18 Weight Loss Overweight or obese adults from psychiatric programs were recruited and randomly assigned to a treatment group or control group. Patients in the treatment group received both weight-management sessions plus exercise sessions plus their usual care. Patients in the control group received their usual treatment for mental illness and no additional treatment. After 18 months, some of the patients had lost 5% or more of their weight, and some had not. The table summarizes the data.

	Treatment Group	Control Group
Lost 5% or more	53	32
Did not lose 5% or more	87	108

a. Find the percentage of each group that lost 5% or more, and compare them descriptively. That is, report both percentages, and indicate what these sample percentages suggest about the effectiveness of the treatment program.

b. Test the hypothesis that the treatment and result are independent using a significance level of 0.05.

(Source: G. L. Daumit et al. 2013. A behavioral weight-loss intervention in persons with serious mental illness. *New England Journal of Medicine* 368, 1594–1602, April 25.)

10.19 Effects of Television Violence on Men The data table compares men who viewed television violence with those who did not, in order to study the differences in physical abuse of the spouse. For the men in the table, test whether television violence and abusiveness are associated, using a significance level of 0.05. Refer to the Minitab output.

	High TV Violence	Low TV Violence
Yes, Physical Abuse	13	27
No Physical Abuse	18	95

(Source: L. R. Husemann et al. 2003. Longitudinal relations between children's exposure to TV violence and their aggressive and violent behavior in young adulthood: 1977–1992. *Developmental Psychology* 39(2), 201–221.)

```
Chi-Square Test: High TV Violence, Low TV Violence
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

         High TV    Low TV
        Violence  Violence   Total
Yes, Ab       13        27      40
            8.10     31.90
           2.957     0.751

No            18        95     113
           22.90     90.10
           1.047     0.266

Total         31       122     153

Chi-Sq = 5.021, DF = 1, P-Value = 0.025
```

10.20 Effects of Television Violence on Women The data table compares women who viewed television violence with those who did not, in order to study the differences in physical abuse of the spouse (Husemann et al. 2003). Test whether television violence and abusiveness are associated, using a significance level of 0.05. Refer to the Minitab output.

	High TV Violence	Low TV Violence
Yes, Physical Abuse	12	30
No Physical Abuse	23	111

```
Chi-Square Test: High TV Violence, Low TV
Violence
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

         High TV     Low TV
        Violence   Violence   Total
Yes, ab       12         30      42
            8.35      33.65
           1.593      0.395

No            23        111     134
           26.65     107.35
           0.499      0.124
Total         35        141     176

Chi-Sq = 2.612, DF = 1, P-Value = 0.106
```

10.21 Gender and Education The table shows the results of a two-way table of gender and whether a person is pursuing higher studies, according to data obtained from an educational survey.

Higher Studies × Respondent's Gender Cross Tabulation Count

		Respondent's Gender		
		Male	**Female**	**Total**
Higher Studies	Graduate	634	481	1115
	Postgraduate	287	237	524
	Doctorate	79	77	156
Total		1000	795	1795

a. If we carry out a test to determine whether these variables are associated, is this a test of independence, homogeneity, or goodness of fit?

b. Do a chi-square test with a significance level of 0.05 to determine whether gender and higher studies are associated.

c. Does this suggest that males and females tend to pursue different levels of higher studies or that their levels of studies are about equal?

10.22 Is Smiling Independent of Age? Randomly chosen people were observed for about 10 seconds in several public places, such as malls and restaurants, to see whether they smiled during that time. The table shows the results for different age groups.

	Age Group				
	0–10	**11–20**	**21–40**	**41–60**	**61 +**
Smile	1131	1748	1608	937	522
No Smile	1187	2020	3038	2124	1509

(Source: M. S. Chapell. 1997. Frequency of public smiling over the life span. *Perceptual and Motor Skills* 85, 1326.)

a. Find the percentage of each age group that were observed smiling, and compare these percentages.

b. Treat this as a single random sample of people, and test whether smiling and age group are associated, using a significance level of 0.05. Comment on the results.

TRY **10.23 Preschool Attendance and High School Graduation Rates (Example 6)** The Perry Preschool Project was created in the early 1960s by David Weikart in Ypsilanti, Michigan. One hundred and twenty three African American children were randomly assigned to one of two groups: One group enrolled in the Perry Preschool, and one did not enroll. Follow-up studies were done for decades to answer the research question of whether attendance at preschool had an effect on high school graduation. The table shows whether the students graduated from regular high school or not. Students who received GEDs were counted as not graduating from high school. This table includes 121 of the original 123. This is a test of homogeneity, because the students were randomized into two distinct samples.

	Preschool	No Preschool
HS Grad	37	29
No HS Grad	20	35

a. For those who attended preschool, the high school graduation rate was 37/57, or 64.9%. Find the high school graduation rate for those not attending preschool, and compare the two. Comment on what the rates show for *these* subjects.

b. Are attendance at preschool and high school graduation associated? Use a 0.05 level of significance.

(Source: L. J. Schweinhart et al. 2005. Lifetime effects: The High/Scope Perry Preschool Study through age 40. *Monographs of the High/Scope Educational Research Foundation,* 14. Ypsilanti, Michigan: High/Scope Press.)

10.24 Preschool Attendance and High School Graduation Rates for Females The Perry Preschool Project data presented in Exercise 10.23 (Schweinhart et al. 2005) can be divided to see whether the preschool attendance effect is different for males and females. The table shows a summary of the data for females, and the figure shows Minitab output that you may use.

```
Chi-Square Test: Preschool, No Preschool: Girls
Expected counts are printed below observed counts

                           No
          Preschool  Preschool  Total

    Grad         21          8     29
              14.50      14.50

    No Grad       4         17     21
              10.50      10.50

Total            25         25     50

Chi-Sq = 13.875, DF = 1, P-Value = 0.000
```

	Preschool	No Preschool
HS Grad	21	8
HS Grad No	4	17

a. Find the graduation rate for those females who went to preschool, and compare it with the graduation rate for females who did not go to preschool.

b. Test the hypothesis that preschool and graduation rate are associated, using a significance level of 0.05.

10.25 Preschool Attendance and High School Graduation Rates for Males The Perry Preschool Project data presented in Exercise 10.23 can be divided to see whether there are different effects for males and females. The table shows a summary of the data for males (Schweinhart et al. 2005).

	Preschool	No Preschool
HS Grad	16	21
HS Grad No	16	18

a. Find the graduation rate for males who went to preschool, and compare it with the graduation rate for males who did not go to preschool.

b. Test the hypothesis that preschool and graduation are associated, using a significance level of 0.05.

c. Exercise 10.24 showed an association between preschool and graduation for just the females in this study. Write a sentence or two giving your advice to parents with preschool-eligible children about whether attending preschool is good for their children's future academic success, based on this data set.

10.26 Odd-Even Formula A survey was taken of a random sample of people noting their gender and asking whether they agreed with the Odd-Even Formula (OEF) to control the alarming levels of air pollution. Minitab results are shown.

```
Chi-Square Test for Association:
Opinion, Gender
Rows: Ooinion     Columns: Gender

             Male     Female       All

Disagree       42         44        86
            42.17      43.83
Agree
              110        114       224
           109.83     114.17
All
              152        158       310
Cell Contents:         Count
                       Expected count

Pearson Chi-Square = 0.002, DF = 1, P-Value = 0.966
```

a. Find the percentage of men and women in the sample who agreed with the OEF method, and compare these percentages.

b. Test the hypothesis that opinions about OEF and gender are independent using a significance level of 0.05.

c. Does this suggest that men and women have significantly different views about the OEF method?

10.27 Bariatric Surgery for Diabetes Mingrone et al. reported the results of an experiment on severely obese patients who had diabetes for at least 5 years. Sixty patients were randomly divided into three groups. One group received medical therapy only (control group), a second group received gastric bypass surgery, and a third group received another kind of surgery called biliopancreatic diversion. It was reported that none of the patients assigned to the control group were free from diabetes after 2 years but that 75% of the gastric-bypass group were free of diabetes and 95% of those receiving biliopancreatic diversion were free from diabetes. Assume that 20 patients were assigned to each group.

a. Find the number of people in each group who were free from diabetes after 2 years.

b. Create a two-way table of the data with Control, Gastric, and Bilio across the top.

c. Test the hypothesis that the treatment and freedom from diabetes are independent using a significance level of 0.05.

(Source: G. Mingrone et al. 2012. Bariatric surgery versus conventional medical therapy for Type 2 diabetes. *New England Journal of Medicine* 366. 577–1585, April 26.)

10.28 Antiretrovirals to Prevent HIV A study conducted in Uganda and Kenya looked at heterosexual couples in which one of the partners was HIV-positive and the other was not. The person in each couple who was not HIV-positive was randomly assigned to one of three study regimens: tenofovir (TDF), combination tenofovir–emtricitabine (TDF–FTC), or placebo and was followed for up to 36 months. Seventeen of the 1584 people assigned to TDF became positive for HIV, as did 13 of the 1579 assigned to TDF–FTC and 52 of the 1584 assigned to the placebo.

a. Find the percentage in each group in the sample that became HIV positive, and compare these percentages.

b. Create a two-way table with the treatment labels across the top.

c. Test the hypothesis that treatment and HIV status are associated using a significance level of 0.05.

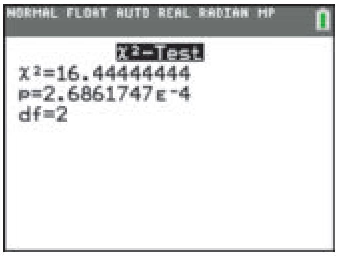
(Source: J. Baelen et al. 2012. Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. *New England Journal of Medicine* 367, 399–410, August.)

10.29 Confederates and Compliance A study was done to see whether participants would ignore a sign that said, "Elevator may stick between floors. Use the stairs." The people who used the stairs were said to be compliant, and those who used the elevator were noncompliant. The study was done in a university dormitory on the ground floor of a three-story building. There were three different situations, two of which involved confederates. A confederate (Conf) is a person who is secretly working with the experimenter. In the first situation, there was no other person using the stairs or elevator—that is, no confederate. In the second, there was a compliant confederate (one who used the stairs). In the third, there was a noncompliant confederate (one who used the elevator). A summary of the data is given in the table, and TI-84 output is given.

	No Conf	Compliant Conf	Noncompliant Conf
Participant Used Stairs	6	16	5
Participant Used Elevator	12	2	13

a. Find the percentage of participants who used the stairs in all three situations. What do these sample percentages say about the association between compliance and the existence of confederates?

b. From the figure, is the p-value 2.69? Explain. Report the actual p-value.



c. Determine whether there is an association between the three situations and whether the participant used the stairs (was compliant) or not. Use a significance level of 0.05.

(Source: L. Shaffer and M. R. Merrens. 2001. *Research Stories for Introductory Psychology*. Allyn and Bacon, Boston. Original Source: M. S. Wogalter, et al. 1987. Effectiveness of warnings. *Human Factors* 29, 599–612.)

10.30 Endocarditis Kang et al. reported on a randomized trial of early surgery for patients with infective endocarditis (a heart infection). Of the 37 patients assigned to early surgery, 1 had a bad result (died, had an embolism, or had a recurrence of the problem within 6 months). Of the 39 patients with conventional treatment (of whom more than half had surgery later on), 11 had a bad result.

a. Find and compare the sample percentages of those who had a bad result for each group.

b. Create a two-way table with the labels Early Surgery and Conventional across the top.

c. Test the hypothesis that early treatment and a bad result are independent at the 0.05 level.

d. Does the treatment cause the effect? Why or why not?

e. Can you generalize to other people? Why or why not?

(Source: D. Kang et al. 2012. Early surgery versus conventional treatment for infective endocarditis. *New England Journal of Medicine* 366, 2466–2473, June 28.)

10.31 Temperature and Jumping Performance A study was done on spotted grass frogs to see whether temperature affects their jumping performance. Twenty frogs were randomly selected to be kept under a temperature-regulated structure (TRS). Twenty similar frogs were randomly selected and kept separately in normal atmosphere (NA). The selected frogs were of the same age range. The frogs, were observed for 30 days. Jumping performance improved in two NA frogs whereas the performance improved in 15 TRS frogs.

a. What percentage of the NA frogs and the TRS frogs in the sample showed improved jumping performance? Compare these percentages and comment.

b. Create a two-way table showing the observed values. Label the columns (across the top) with TRS and NA.

c. Test the hypothesis that the temperature is associated with jumping performance (at the 0.05 level of significance).

10.32 Removal of Healthy Appendixes Computed tomography (CT) scans are used to diagnose the need for the removal of the appendix. CT scans give the patient a large level of radiation, which has risks, especially for young people. There is a new form of CT scanning called low-dose CT, which was tested to see whether it was inferior when diagnosing appendicitis. Negative appendectomies are appendectomies that were done even though the appendix was healthy. The negative appendectomy rate was 6 of 172 patients randomly assigned to the low-dose CT and 6 out of 186 patients randomly assigned to the standard-dose group.

a. Find the negative appendectomy rates for both samples and compare them.

b. Test the hypothesis that the negative appendectomy rate and dosage are independent at the 0.05 level.

(Source: K. Kim et al. 2012. Low-dose abdominal CT for evaluating suspected appendicitis. *New England Journal of Medicine* 366, 1596–1605, April 26.)

SECTION 10.3

TRY **10.33 Effect of Confederates on Compliance (Example 7)** A study was done to see whether participants would ignore a sign that read, "Elevator may stick between floors. Use the stairs." The people who used the stairs were classified as compliant, those who used the elevator as noncompliant. The study was done in a university dorm on the ground floor of a building that had three floors. There were three different situations, two of which involved a person who was secretly working with the experimenter. (This person is called a confederate.) In the first situation, there was no other person using the stairs or elevator—that is, no confederate. In the second, there was a compliant confederate (one who used the stairs). In the third, there was a noncompliant confederate (one who used the elevator). Suppose that the participants (people who arrived to use the elevator at the time the experiment was going on) were randomly assigned to the three groups. There were significant differences between groups.

a. Can we generalize widely to a large group? Why or why not?

b. Can we infer causality? Why or why not?

(Source: Lary Shaffer and Matthew R. Merrens. 2001. *Research stories for introductory psychology.* Allyn and Bacon. Original Source: M. S. Wogalter, et al. 1987. Effectiveness of warnings. *Human Factors* 29, 599–612.)

10.34 Physiotherapy Suppose a new medicine to help patients suffering from arthritis was developed and tested. Patients voluntarily entered the study and were randomly assigned either the new medicine or physiotherapy. Suppose a larger percentage of those using the new medicine reported relief from joint pain.

a. Can we generalize widely to a large group? Why or why not?

b. Can we infer causality? Why or why not?

10.35 Hospital Rooms When patients are admitted to hospitals, they are sometimes assigned to a single room with one bed and sometimes assigned to a double room, with a roommate. (Some insurance companies will pay only for the less expensive, double rooms.) A researcher was interested in the effect of the type of room on the length of stay in the hospital. Assume that we are not dealing with health issues that require single rooms.

Suppose that upon admission to the hospital, the names of patients who would have been assigned a double room were put onto a list and a systematic random sample was taken; every tenth patient who would have been assigned to a double room was part of the experiment. For each participant, a coin was flipped: If it landed heads up, she or he got a double room, and if it landed tails up, a single room. Then the experimenters observed how many days the patients stayed in the hospital and compared the two groups. The experiment ran for two months. Suppose those who stayed in single rooms stayed (on average) one less day, and suppose the difference was significant.

a. Can you generalize to others from this experiment? If so, to whom can you generalize, and why can you do it?
b. Can you infer causality from this study? Why or why not?

10.36 Museum Visit A random sample of 100 visitors (out of a total of 10,000 visitors) was obtained from museum records using systematic sampling. Half of those visitors had a museum tour assigned with a guide and half had a tour with the curator. The decision was made randomly by a coin flip for each visitor. Suppose that those with the curator rated their experience higher than those with the guide.

a. Can you generalize to other visitors at the museum? Explain.
b. Can you infer causality from this study? Explain.

10.37 A Drug for Platelets Platelets are an important part of the blood because they cause the blood to clot when that is needed to stop bleeding. The drug eltrombopag (which we shall henceforth refer to as EL), was tried on patients with chronic liver disease who were about to be operated on to see whether it would prevent the need for transfusion of platelets during the surgery and shortly thereafter. Read the following excerpt from the abstract that accompanied this study, and answer the questions below.

> "*Methods:* We randomly assigned 292 patients with chronic liver disease of diverse causes and platelet counts of less than 50,000 per cubic millimeter to receive EL, at a dose of 75 mg daily, or placebo for 14 days before a planned elective invasive procedure that was performed within 5 days after the last dose. The primary end point was the avoidance of a platelet transfusion before, during, and up to 7 days after the procedure.
>
> *Results:* A platelet transfusion was avoided in 104 of 145 patients who received EL (72%) and in 28 of 147 who received placebo (19%) ($P < 0.001$)."

(Source: Nezam Afdhal et al. 2012. Eltrombopag before procedures in patients with cirrhosis and thrombocytopenia. *New England Journal of Medicine* 367, 716–724.)

a. Identify the treatment variable and the response variable.
b. Was this a controlled experiment or observational study. Explain.
c. Do the sample percentages suggest that the drug was effective in reducing the chance of needing platelet transfusions?
d. What does the small p-value show?
e. Can you conclude that the use of EL reduces the chance of needing a platelet transfusion? Why or why not?

10.38 Steroids and Height Does the use of inhaled steroids by children affect their height as adults? Excerpts from the abstract of a study about this are given. Read them and then answer the questions that follow.

> "*Methods:* We measured adult height in 943 of 1041 participants (90.6%) in the Childhood Asthma Management Program; adult height was determined at a mean ($\pm$SD) age of 24.9 ± 2.7 years. Starting at the age of 5 to 13 years, the participants had been randomly assigned to receive 400 μg of budesonide, 16 mg of nedocromil, or placebo daily for 4 to 6 years.
>
> *Results: Mean* adult height was 1.2 cm lower (95% confidence interval [CI], -1.9 to -0.5) in the budesonide group than in the placebo group ($P = 0.001$) and was 0.2 cm lower (95% CI, -0.9 to 0.5) in the nedocromil group than in the placebo group ($P = 0.61$)."

a. Identify the treatment variable and the response variable.
b. Was this a controlled experiment or an observational study? Explain.
c. Does the first interval, (-1.9 to -0.5), capture 0? What does that show?
d. From the interval, can you conclude that the use of budesonide in childhood reduces the heights of the children when they become adults? Why or why not?
e. Does the second interval, (-0.9 to 0.5), capture 0? What does that show?
f. From the interval, can you conclude that the use of nedocromil in childhood reduces the heights of the children when they become adults? Why or why not?

(Source: L. H. William Kelly et al. 2012. Effect of inhaled glucocorticoids in childhood on adult height. *New England Journal of Medicine* 367, 904–912.)

10.39 Funds and Returns In May 2016, the *Economic Times* reported that "Growth-oriented funds tend to exhibit strong returns within a short span of time." Is this conclusion likely to be the result of an observational study or a controlled experiment? Is it saying that growth-oriented funds lower the risk of low returns?

10.40 Treatment of Diarrhea A research group compared probiotics with antibiotics to know the best way to treat diarrhea. It showed that consumption of probiotics was a better way to cure diarrhea than antibiotics.

a. What do you need to know to decide whether this was an observational study or a controlled experiment?
b. Why do controlled experiments with randomization allow us to draw conclusions implying cause and effect?

10.41 Iron and Death Rate In February 2012, the magazine *Health After 50* reported on the Iowa Women's Health Study. Experts used a questionnaire to collect data from 39,000 women 55–69 years old concerning their dietary supplements. After 19 years they looked at death rates. Many of the supplements were associated with a higher risk of death, iron being the most notable: The higher the dose of iron taken, the higher the death risk. Does this study show that consumption of iron causes higher death rates?

10.42 Calcium and Death Rate The magazine *Health After 50* reported on the Iowa Women's Health Study. Experts used a questionnaire to collect data from 39,000 women 55–69 years old concerning their dietary supplements. After 19 years they looked at death rates. Many of the supplements were associated with a higher risk of death. However, the consumption of calcium was associated with a *lower* death rate. Does this study show that consumption of calcium causes lower death rates?

10.43 Design with Iron Refer to Exercise 10.41. How could you find out whether iron caused the higher death rates associated in this study with its use? Describe the design of a study assuming you had 200 women to work with. Assume that you do not have to study the women for 19 years but, rather, will look at them for a much shorter time period, perhaps one or two years.

10.44 Design with Calcium Refer to Exercise 10.42. How could you find out whether calcium caused the lower death rates associated in this study with its use? Describe the design of a study assuming you had 300 women to work with. Assume that you do not have to study the women for 19 years but, rather, will look at them for a much shorter time period, such as one or two years.

10.45 Drug for Rheumatoid Arthritis Read the portion of the abstract of a scientific study that appears below, and then answer the questions that follow.

> "*Methods:* In this . . . double-blind, placebo-controlled . . . study, 611 patients were randomly assigned, in a 4:4:1 ratio, to [receive] 5 mg of tofacitinib twice daily, 10 mg of tofacitinib twice daily, or placebo. The primary end point, assessed at month 3, was the percentage of patients with at least a 20% improvement on the American College of Rheumatology scale (ACR 20),
>
> *Results:* At month 3, a higher percentage of patients in the tofacitinib groups than in the placebo groups met the criteria for an ACR 20 response (59.8% in the 5-mg tofacitinib group and 65.7% in the 10-mg tofacitinib group vs. 26.7% in the placebo group, $P < 0.001$ for both comparisons)."

a. Identify the treatment variable and the response variable.
b. Was this a controlled experiment or an observational study? Explain.
c. Do the sample percentages suggest that the drug was effective in achieving a 20% reduction in symptoms?
d. What does the small p-value show?
e. Can you conclude that the use of tofacitinib increases the chances of a 20% improvement in symptoms? Why or why not?

(Source: Fleischmann, Roy et al. 2012. Placebo-controlled trial of tofacitinib tonotherapy in rheumatoid arthritis. *New England Journal of Medicine* 367, 495–507.)

10.46 Tight Control of Blood Sugar "Tight glycemic control" means that the blood sugar is kept within a narrow range. Read the abstract below, and then answer the questions that follow it.

> "*Methods:* In this two-center, prospective, randomized trial, we enrolled 980 children, 0 to 36 months of age, undergoing surgery with cardiopulmonary bypass. Patients were randomly assigned to either tight glycemic control . . . targeting a blood glucose level of 80 to 110 mg per deciliter . . . or standard care in the cardiac intensive care unit.
>
> *Results:* A total of 444 of the 490 children assigned to tight glycemic control (91%) received insulin versus 9 of 490 children assigned to standard care (2%). . . . [T]ight glycemic control was not associated with a significantly decreased rate of health care–associated infections (8.6 vs. 9.9 per 1000 patient-days, $P = 0.67$).
>
> *Conclusions:* Tight glycemic control can be achieved with a low hypoglycemia rate after cardiac surgery in children, but it does not significantly change the infection rate . . . as compared with standard care."

a. Identify the treatment variable and the response variable.
b. Was this a controlled experiment or an observational study? Explain.
c. What does the p-value show?
d. Can you conclude that the use of tight glycemic control affects the rate of infections? Why or why not?

(Source: Agus, Michael et al. 2012. Tight glycemic control versus standard care after pediatric cardiac surgery. *New England Journal of Medicine* 367, 1208–1219.)

* **10.47 Alumni Donations** The alumni office wishes to determine whether students who attend a reception with alumni just before graduation are more likely to donate money within the next two years.

a. Describe a study based on a sample of students that would allow the alumni office to conclude that attending the reception *causes* future donations but that it is *not* possible to generalize this result to all students.
b. Describe a study based on a sample of students that does *not* allow fundraisers to conclude that attending receptions causes future donations but does allow them to generalize to all students.
c. Describe a study based on a sample of students that allows fundraisers to conclude that attending the reception causes future donations and also allows them to generalize to all students.

* **10.48 School Dropout Rates** The school dropout rate in an Indian village is about 25%, which means that 25% of the students leave school without completing their education. There have been many attempts to reduce this rate. One of these attempts is to encourage students to restart education through incentives like mid-day meals and financial assistance. Suppose you want to determine whether the encouragement methods actually help in reducing the dropout rate. Suppose that students who are aided with incentives are observed for a year to see whether they drop out.

a. Describe a study based on a sample of students that would allow the management to conclude that encouragement causes a reduction in dropout rate but would not allow it to generalize this result to students in all villages.
b. Describe a study based on a sample of students that does *not allow* the management to conclude that encouragement causes a reduction in dropout rate but does allow it to generalize to students in all villages.
c. Describe a study based on a sample of students that allows the management to conclude that encouragement causes a reduction in the dropout rate and also allows it to generalize to students in all villages.

TRY g **10.49 Drug for Asthma (Example 8)** Eosinophils are a form of white blood cell that is often present in people suffering from allergies. People with asthma and high levels of eosinophils who used steroid inhalers were given either a new drug or a placebo. Read extracts from the abstract of this study that appear below, and then evaluate the study. *See page 539 for questions and guidance.*

> "*Methods:* We enrolled patients with persistent, moderate-to-severe asthma and a blood eosinophil count of at least 300 cells per microliter . . . who used medium-dose to high-dose inhaled glucocorticoids. . . . We administered dupilumab (300 mg) or placebo subcutaneously once weekly. The primary end point was the occurrence of an asthma exacerbation [worsening].
>
> *Results:* A total of 52 patients were [randomly] assigned to the dupilumab group, and 52 patients were [randomly] assigned to the placebo group. . . . Three patients had an asthma exacerbation with dupilumab (6%) versus 23 with placebo (44%), corresponding to an 87% reduction with dupilumab (odds ratio, 0.08; 95% confidence interval, 0.02 to 0.28; $P < 0.001$).
>
> *Conclusions:* In patients with persistent, moderate-to-severe asthma and elevated eosinophil levels who used inhaled glucocorticoids and LABAs, dupilumab therapy, as compared with placebo, was associated with fewer asthma exacerbations [worsenings]."

(Source: Sally Wenzel et al. 2013. Dupilumab in persistent asthma with elevated eosinophil levels. *New England Journal of Medicine* 368, 2455–2466.)

10.50 Blood Sugar Refer to Exercise 10.46 on tight glycemic control, and answer the questions asked in Exercise 10.49 in the Guided Exercises on page 539.

* **10.51 Colored Vegetables and Stroke** A study of colored vegetables and the risk of stroke was done. Although the investi-

gators did not see any effect on stroke of consumption of green, orange, red, yellow, or purple vegetables, they concluded that "High intake of white fruits and vegetables may protect against stroke."

Suppose that for each pair of color groups, we wished to test whether the stroke risk was different. For instance, is it different for green vs. orange? For green vs. red? Suppose there are 10 different pairs to be compared. Suppose that, for each pair, we perform a hypothesis test with a significance level of 10%. Assume that in truth, there are no differences between any of the pairs. By chance alone, how many of the hypothesis tests would you expect to appear significant (and thus lead us to mistakenly believe that there was a difference)? (Source: Oude Griep, et al . 2011. Colors of fruit and vegetables and 10-year incidence of stroke. *Stroke* 42(11), 3190–3195.)

10.52 Autism and MMR Vaccine An article in the British medical journal *Lancet* claimed that autism was caused by the measles, mumps, and rubella (MMR) vaccine. This vaccine is typically given to children twice, at about the age of 1 and again at about 4 years of age. The article reports a study of 12 children with autism who had all received the vaccines shortly before developing autism. The article was later retracted by *Lancet* because the conclusions were not justified by the design of the study.

Explain why *Lancet* might have felt that the conclusions were not justified by listing potential flaws in the study, as described above. (Source: A. J. Wakefield et al. 1998. Ileal–lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet* 351, 637–641.)

CHAPTER REVIEW EXERCISES

* **10.53 Perry Preschool Arrests** The Perry Preschool Project discussed in Exercises 10.23–10.25 found that 8 of the 58 students who attended preschool had at least one felony arrest by age 40 and that 31 of the 65 students who did not attend preschool had at least one felony arrest (Schweinhart et al. 2005).

a. Compare the percentages descriptively. What does this comparison suggest?
b. Create a two-way table from the data and do a chi-square test on it, using a significance level of 0.05. Test the hypothesis that preschool attendance is associated with being arrested.
c. Do a two-proportion z-test. Your alternative hypothesis should be that preschool attendance lowers the chances of arrest.
d. What advantage does the two-proportion z-test have over the chi-square test?

* **10.54 Parental Training and Criminal Behavior of Children** In Montreal, Canada, an experiment was done with parents of children who were thought to have a high risk of committing crimes when they became teenagers. Some of the families were randomly assigned to receive parental training, and the others were not. Out of 43 children whose parents were randomly assigned to the parental training group, 6 had been arrested by the age of 15. Out of 123 children whose parents were not in the parental training group, 37 had been arrested by age 15.

a. Find and compare the percentages of children arrested by age 15. Is this what researchers might have hoped?
b. Create a two-way table from the data, and test whether the treatment program is associated with arrests. Use a significance level of 0.05.
c. Do a two-proportion z-test, testing whether the parental training lowers the rate of bad results. Use a significance level of 0.05.
d. Explain the difference in the results of the chi-square test and the two-proportion z-test.
e. Can you conclude that the treatment causes the better result? Why or why not?

(Source: R. E. Tremblay et al. 1996. From childhood physical aggression to adolescent maladjustment: The Montreal prevention experiment. In R.D. Peters and R. J. McMahon. *Preventing childhood disorders, substance use and delinquency*. Thousand Oaks, California: Sage, pp. 268–298.)

10.55 Late Registrations of Birth The table shows statistical information on recorded live births in the Republic of South Africa in different years. Some of these registrations happened on time and some were recorded late (Source: Statistics South Africa, 2014). Explain why you should not perform a chi-square test of homogeneity on this data set.

Year	On Time	Late
2004	49.3%	50.7%
2005	57.5%	42.5%
2006	63.9%	36.1%
2007	71.6%	28.4%
2008	71.7%	28.3%
2009	70.1%	29.90%
2010	68.7%	31.3%
2011	75.8%	24.2%
2012	79.3%	20.7%

10.56 Tourists by Month All the tourists visiting the Republic of Maldives from 2009 to 2013 were tallied. Then the average number of tourists per month was calculated and is shown in the table (Source: Ministry of Tourism, 2014). Explain why it would be inappropriate to do a chi-square analysis of this data set.

Month	Average Tourists Per Month
January	85.92
February	90.74
March	82.88
April	77.88
May	64.70
June	53.08
July	65.08
August	72.80
September	67.06
October	80.62
November	78.52
December	77.90

* **10.57 Vaccinations for Diarrhea in Mexico** Diarrhea can kill children and is often caused by rotavirus. Read the abstract below, and answer the questions that follow.

> *"Methods:* We obtained data on deaths from diarrhea, regardless of cause, from January 2003 through May 2009 in Mexican children under 5 years of age. We compared diarrhea-related mortality in 2008 and during the 2008 and 2009 rotavirus seasons with the mortality at baseline (2003–2006), before the introduction of the rotavirus vaccine. Vaccine coverage was estimated from administrative data.
>
> *Results:* Diarrhea-related mortality fell from an annual median of 18.1 deaths per 100,000 children at baseline to 11.8 per 100,000 children in 2008 (rate reduction, 35%; 95% confidence interval [CI], 29 to 39; $P < 0.001$). . . . Mortality among unvaccinated children between the ages of 24 and 59 months was not significantly reduced. The reduction in the number of diarrhea-related deaths persisted through two full rotavirus seasons (2008 and 2009).
>
> *Conclusions:* After the introduction of a rotavirus vaccine, a significant decline in diarrhea-related deaths among Mexican children was observed, suggesting a potential benefit from rotavirus vaccination."

a. State the death rate before vaccine and the death rate after vaccine. What was the change in deaths per 100,000 children? From the given p-value, can you reject the null hypothesis of no change in death rate?
b. Would you conclude that the vaccine was effective? Why or why not?

(Source: R. Vesta et al. 2010. Effect of rotavirus vaccination on death from childhood diarrhea in Mexico. *New England Journal of Medicine* 62, 299–305.)

10.58 Nighttime Physician Staffing in ICU (Intensive Care Unit) Read the following abstract and explain what it shows. A rate ratio of 1 means there is no difference in rates, and a confidence interval for rate ratios that captures 1 means there is no significant difference in rates. (An intensivist is a doctor who specializes in intensive care.)

> We conducted a 1-year randomized trial in an academic medical ICU of the effects of nighttime staffing with in-hospital intensivists (intervention) as compared with nighttime coverage by daytime intensivists who were available for consultation by telephone (control). We randomly assigned blocks of 7 consecutive nights to the intervention or the control strategy. The primary outcome was patients' length of stay in the ICU. Secondary outcomes were patients' length of stay in the hospital, ICU and in-hospital mortality, discharge disposition, and rates of readmission to the ICU.
>
> A total of 1598 patients were included in the analyses. . . . Patients who were admitted on intervention days were exposed to nighttime intensivists on more nights than were patients admitted on control days. Nonetheless, intensivist staffing on the night of admission did not have a significant effect on the length of stay in the ICU (rate ratio for the time to ICU discharge, 0.98; 95% confidence interval [CI], 0.88 to 1.09; $P = 0.72$), on ICU mortality (relative risk, 1.07; 95% CI, 0.90 to 1.28), or on any other end point.

(Source: Meeta Kerlin et al. 2013. A randomized trial of nighttime physician staffing in an intensive care unit. *New England Journal of Medicine* 368, 2201–2209.)

10.59 Prostate Cancer Treatment The following portion of an abstract gives information on the comparison of treatments of men with prostate cancer. Read it and answer the questions about it below. The prostate gland surrounds the neck of the bladder in men.

> "*Methods:* Between 1989 and 1999, we randomly assigned 695 men with early prostate cancer to watchful waiting or radical prostatectomy and followed them through the end of 2012. The primary end points in the Scandinavian Prostate Cancer Group Study Number 4 (SPCG-4) were death from any cause, death from prostate cancer, and the risk of metastases.
>
> *Results:* During 23.2 years of follow-up, 200 of 347 men in the surgery group and 247 of the 348 men in the watchful-waiting group died. Of the deaths, 63 in the surgery group and 99 in the watchful-waiting group were due to prostate cancer; the relative risk was 0.56 (95% confidence interval, 0.41 to 0.77; $P = 0.001$), and the absolute difference was 11.0 percentage points (95% CI, 4.5 to 17.5). The number needed to treat to prevent one death was 8. One man died after surgery in the radical-prostatectomy group. . . . The benefit of surgery with respect to death from prostate cancer was largest in men younger than 65 years of age (relative risk, 0.45) and in those with intermediate-risk prostate cancer (relative risk, 0.38). However, radical prostatectomy was associated with a reduced risk of metastases among older men (relative risk, 0.68; $P = 0.04$)."

(Source: Anna Bill-Axelson et al. 2014. Radical prostatectomy or watchful waiting in early prostate cancer. *New England Journal of Medicine* 370, 932–942.)

a. Compare the percentages of death for the two groups descriptively. Which group had patients who were more likely to live?
b. Find and compare the percentage who died from prostate cancer for each group.
c. Was this an observational study or a controlled experiment?

* **10.60 Multiple Myeloma** One treatment for multiple myeloma (cancer of the blood and bones) is a stem cell transplant. However, in some cases the cancer returns. McCarthy and colleagues reported on a study that randomly assigned 460 patients (100 days after a stem cell transplant) to receive either lenalidomide or placebo. At one point in the study, 46 of the patients who received the real drug had a bad result (had progressive disease or had died), compared to 101 of those who received the placebo. Assume that exactly half were assigned to each group.

a. Find and compare the percentages that had a bad result for the two groups.
b. Test the hypothesis that the drug reduced the chance of a bad result compared to the placebo using a significance level of 0.05.
c. The study started in April 2005 and was "unblinded" in 2009 when an interim analysis showed better results with the group taking the drug. After the unblinding, many of the patients from the placebo group "crossed over" to the drug group. Explain what you think "unblinding" means and why this seems like a reasonable thing to do.

(Source: P. L. McCarthy et al. 2012. Lenalidomide after stem-cell transplantation for multiple myeloma. *New England Journal of Medicine* 366, 1770–1781.)

GUIDED EXERCISES

10.17 Obesity and Marital Status A study reported in the medical journal *Obesity* in 2009 analyzed data from the National Longitudinal Study of Adolescent Health. Obesity was defined as having a body mass index (BMI) of 30 or more. The research subjects were followed from adolescence to adulthood, with all individuals in the sample categorized in terms of whether or not they were obese and whether they were dating, cohabiting, or married.

	Dating	Cohabiting	Married
Obese	81	103	147
Not Obese	359	326	277

QUESTION Test the hypothesis that the variables *Relationship Status* and *Obesity* are associated, using a significance level of 0.05. Also consider whether the study shows causality. The steps will guide you through the process. Minitab output is provided.

Step 1 ▶ Hypothesize

H_0: Relationship status and obesity are independent.

H_a: ?

```
Chi-Square Test: Dating, Cohabiting, Married
Expected counts are printed below observed counts

        Dating   Cohabiting   Married   Total
 Obese      81          103       147     331
        112.64       109.82    108.54

 Not       359          326       277     962
        327.36       319.18    315.46

Total      440          429       424    1293

Chi-Sq = 30.829, DF = 2, P-Value = 0.000
```

Step 2 ▶ Prepare

We choose the chi-square test of independence because the data were from *one* random sample in which the people were classified two different ways. We do not have a random sample or a random assignment, so we will test to see whether these results could easily have occurred by chance. Find the smallest expected count and report it. Is it more than 5? Report the level of significance.

Step 3 ▶ Compute to compare

Refer to the output given.

$X^2 =$ ______

p-value = ______

Step 4 ▶ Interpret

Reject or do not reject the null hypothesis, and state what that means.

Causality

Can we conclude from these data that living with someone is making some people obese and that marrying is making even more people obese? Can we conclude that obesity affects your relationship status? Explain why or why not.

Percentages

Find and compare the percentages obese in the three relationship statuses.

g **10.49 Drug for Asthma (Example 8)** Eosinophils are a form of white blood cell that is often present in people suffering from allergies. People with asthma and high levels of eosinophils who used steroid inhalers were given a new drug or a placebo. Read the extracts from the abstract of this study that appear below, and then evaluate the study.

> "*Methods:* We enrolled patients with persistent, moderate-to-severe asthma and a blood eosinophil count of at least 300 cells per microliter . . . who used medium-dose to high-dose inhaled glucocorticoids. . . . We administered dupilumab (300 mg) or placebo subcutaneously once weekly. The primary end point was the occurrence of an asthma exacerbation [worsening].
>
> *Results:* A total of 52 patients were [randomly] assigned to the dupilumab group, and 52 patients were [randomly] assigned to the placebo group. . . . Three patients had an asthma exacerbation with dupilumab (6%) versus 23 with placebo (44%), corresponding to an 87% reduction with dupilumab (odds ratio, 0.08; 95% confidence interval, 0.02 to 0.28; ($P < 0.001$)."
>
> *Conclusions:* In patients with persistent, moderate-to-severe asthma and elevated eosinophil levels who used inhaled glucocorticoids and LABAs, dupilumab therapy, as compared with placebo, was associated with fewer asthma exacerbations. . . .

Step 1 ▶ What is the research question that these investigators are trying to answer?

Step 2 ▶ What is their answer to the research question?

Step 3 ▶ What were the methods they used to collect data? (controlled experiment or observational study)

Step 4 ▶ Is the conclusion appropriate for the methods used to collect data?

Step 5 ▶ To what population do the conclusions apply?

Step 6 ▶ Have the results been replicated in other articles? Are the results consistent with what other researchers have suggested?

(Source: Sally Wenzel et al. 2013. Dupilumab in persistent asthma with elevated eosinophil levels. *New England Journal of Medicine* 368, 2455-2466.)

TechTips

General Instructions for All Technology

EXAMPLE (CHI-SQUARE TEST FOR TWO-WAY TABLES): PERRY PRESCHOOL AND GRADUATION FROM HIGH SCHOOL ► In the 1960s an experiment was started in which a group of children were randomly assigned to attend preschool or not to attend preschool. They were studied for years, and whether they graduated from high school is shown in Table A.

We will show the chi-square test for two-way tables to see whether the factors are independent or not. For Minitab and StatCrunch, we also show Fisher's Exact Test.

	Preschool	No Preschool
Grad HS	37	29
No Grad HS	20	35

▲ **TABLE A** Two-way table for preschool and graduation from high school

Discussion of Data

Much of technology is set up so that you can use the table summary (such as Table A) and find the calculated results. However, it is also possible to start with a spreadsheet containing the raw data. Table B shows the beginning of the raw data, for which there would be 121 rows for the 121 children.

Preschool	Graduate HS
Yes	No
Yes	Yes
No	Yes
No	Yes
Yes	Yes
No	No

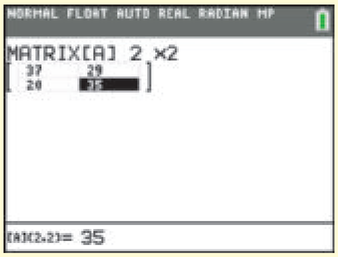
▲ **TABLE B** Some raw data

TI-84

Chi-Square for Two-way Tables

You will not put the data into the lists. You will use a matrix (table), and the data must be in the form of a summary such as Table A.

1. Press **2ND** and **MATRIX** (or **MATRX**).
2. Scroll over to **EDIT** and press **ENTER** when **1:** is highlighted.
3. See Figure 10A. Put in the dimensions. Because the table has two rows and two columns, press **2**, **ENTER**, **2**, **ENTER**. (The first number is the number of rows, and the second number is the number of columns.)



▲ **FIGURE 10A** TI-84 Input for Two-way Table

4. Enter each of the four numbers in the table, as shown in Figure 10A. Press **ENTER** after typing each number.
5. Press **STAT**, and scroll over to **TESTS**.
6. Scroll down (or up) to **C: χ^2-Test** and press **ENTER**.
7. Leave the **Observed** as **A** and the **Expected** as **B**. Scroll down to **Calculate** and press **ENTER**.

   You should get the output shown in Figure 10B.

▲ **FIGURE 10B** TI-84 Output for a Chi-Square Test

8. To see the expected counts, click **2ND**, **MATRIX**, scroll over to **EDIT**, scroll down to **2: [B]**, and press **ENTER**. You may have to scroll to the right to see some of the numbers. They will be arranged in the same order as the table of observed values. Check these numbers for the required minimum value of 5.

MINITAB

Two-way Tables

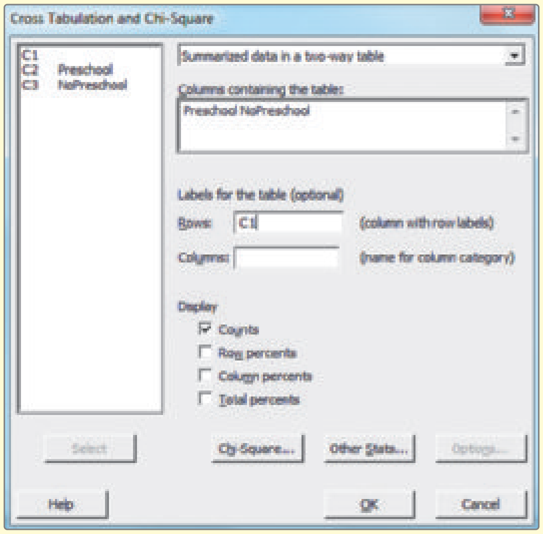
For Minitab you may have your data as a table summary (as shown in Table A) or as raw data (as shown in Table B).

TABLE SUMMARY

1. Type in the data summary, together with the optional row and column labels, as shown below.

↓	C1-T	C2	C3
		Preschool	No Preschool
1	GradHS	37	29
2	NoGradHS	20	35

2. **Stat > Tables > Cross Tabulation and Chi-Square**
3. See Figure 10C. Choose **Summarized data in a two-way table**. Select both columns (by double clicking them) For **Rows**: select **C1**.



▲ **FIGURE 10C** Minitab Input for Two-way Tables

4. Click **Chi-square...**, check **Chi-square test** and **Expected cell counts**; click **OK**. Click **Other Stats...** and check **Fisher's exact test**; click **OK**. Click **OK**.

   Figure 10D shows the output.

```
          Preschool  NoPreschool  All

GradHS           37           29   66
              31.09        34.91

NoGrad           20           35   55
              25.91        29.09

All              57           64  121

Cell Contents:      Count
                    Expected count

Pearson Chi-Square = 4.671, DF = 1, P-Value = 0.031
Likelihood Ratio Chi-Square = 4.710, DF = 1, P-Value = 0.030

Fisher's exact test: P-Value =  0.0439073
```

▲ **FIGURE 10D** Minitab Output for Two-way Tables.

RAW DATA

1. Make sure your raw data are in the columns. See Table B.
2. **Stat > Tables > Cross Tabulation and Chi-square**.
3. Choose **Raw data (categorical variables)**. Click either C1 **For rows** and C2 **For columns** (or vice versa).
4. Do step 4 above.

EXCEL

Two-way Tables

1. Type a summary of your data into two (or more) columns, as shown in columns A and B in Figure 10E.

B2 ▾ f_x 35

	A	B	C	D	E	F	G
1	37	29					
2	20	35					
3							
4							

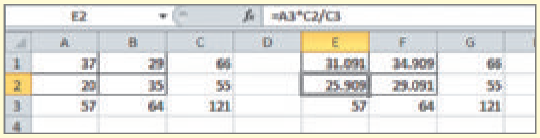
▲ **FIGURE 10E** Excel Input for Two-way Table

2. To get the total of 66 (from 37 + 29), click in the box to the right of 29, and then click ***fx***, double click **SUM**, and click **OK**. You can do the same thing for the other sums, each time starting from the cell in which you want to put the sum. Save the grand total for last. Your table with the totals should look like columns A, B, and C in Figure 10F on the next page. Alternatively, you could simply add to get the totals.

3. To get the expected counts, you will be using the formula

$$\text{Expected count} = \frac{(\text{row marginal total}) \times (\text{column marginal total})}{(\text{grand total})}$$

   To get the first expected count, 31.09091, click in the cell where you want the expected count to be placed (here, cell E1). (An empty column, such as column D in Figure 10F, improves the clarity.) Then type = and click on the **66** in the table, type ***** (for multiplication), click on the **57**, type **/** (for division), click on the **121**, and press **Enter**. The input for getting the expected count of 25.909 in cell E2 is shown in the formula bar at the top of Figure 10F. For each of the expected counts, you start from the cell you want filled, and you click on the row total, ***** (for multiply), the column total, **/** (for divide), and the grand total and press **Enter**. Alternatively, you could figure out the expected counts by hand.



▲ **FIGURE 10F** Excel, Including Totals and Expected Counts

After you have all four expected counts, be sure they are arranged in the same order as the original data:

4. Click *fx*.
5. Select a category: **Statistical** or **All**.
6. Choose **CHISQ.TEST**. For the **Actual_range**, highlight the table containing the observed counts, but do *not* include the row and column totals or the grand total. For the **Expected range**, highlight the table with the expected counts.

   You will see the p-value (0.030671). Press **OK** and it will show up in the active cell in the worksheet.

The previous steps for Excel will give you the p-value but not the value for chi-square. If you want the numerical value for chi-square, continue with the steps that follow.

7. Click in an empty cell.
8. Click *fx*.
9. Select a category: **Statistical** or **All**.
10. Choose **CHISQ.INV.RT** (for inverse, right tail).
11. For the **Probability**, click on the cell from step 6 that shows the p-value of **0.030671**. For **Deg_freedom**, put in the degrees of freedom (df). For two-way tables,

$$df = (\text{number of rows} - 1)(\text{number of columns} - 1).$$

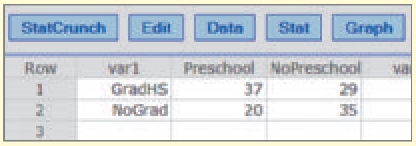
For these data, df is 1. Click **OK**. You should get a chi-square of 4.67.

STATCRUNCH

Two-way Tables

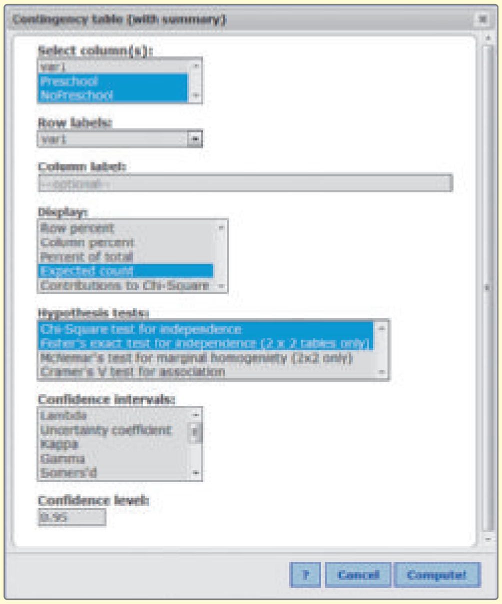
TABLE SUMMARY

1. Enter your data summary as shown in Figure 10G. Note that you can have column labels (Preschool or No Preschool) and also row labels (GradHS or NoGrad).



▲ **FIGURE 10G** StatCrunch Input for Two-way Table

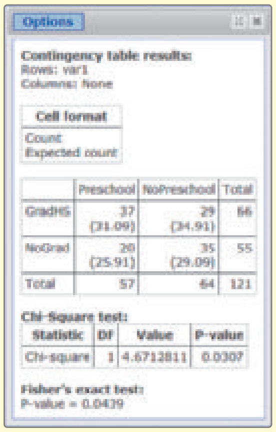
2. **Stat > Tables > Contingency > With Summary**
3. See Figure 10H. Select the columns that contain the summary counts (press keyboard **Ctrl** when selecting the second column), and then select the column that contains the **Row labels**, here **var1**. Select **Expected Count.**



▲ **FIGURE 10H** StatCrunch Two-way-Table Options

4. If Fisher's Exact Test is also wanted, press **Ctrl** on keyboard while selecting **Fisher's exact test for independence**…
5. Click **Compute!**

   Figure 10I shows the well-labeled output.



▲ **FIGURE 10I** StatCrunch Output for Two-way Table

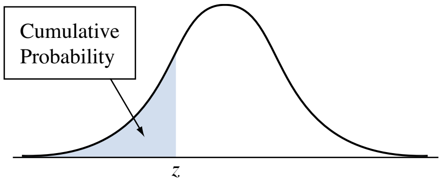
RAW DATA

1. Be sure you have raw data in the columns; see Table B.
2. **Stat > Tables > Contingency > With Data**
3. Select both columns.
4. Select **Expected Count**.
5. Click **Compute!**

Appendix A: Tables

Table 1: Random Numbers

Line						
01	21033	32522	19305	90633	80873	19167
02	17516	69328	88389	19770	33197	27336
03	26427	40650	70251	84413	30896	21490
04	45506	44716	02498	15327	79149	28409
05	55185	74834	81172	89281	48134	71185
06	87964	43751	80971	50613	81441	30505
07	09106	73117	57952	04393	93402	50753
08	88797	07440	69213	33593	42134	24168
09	34685	46775	32139	22787	28783	39481
10	07104	43091	14311	69671	01536	02673
11	27583	01866	58250	38103	35825	94513
12	60801	04439	58621	09840	35119	60372
13	62708	04888	37221	49537	96024	24004
14	21169	14082	65865	29690	00280	35738
15	13893	00626	11773	14897	37119	29729
16	19872	41310	65041	61105	31028	80297
17	29331	36997	05601	09785	18100	44164
18	76846	74048	08496	22599	29379	11114
19	11848	80809	25818	38857	23811	80902
20	85757	33963	93076	39950	29658	07530
21	71141	00618	48403	46083	40368	33990
22	47371	36443	41894	62134	86876	18548
23	46633	10669	95848	69055	49044	75595
24	79118	21098	63279	26834	43443	38267
25	91874	87217	11503	47925	13289	42106
26	85337	08882	68429	61767	18930	37688
27	88513	05437	22776	17562	03820	44785
28	31498	85304	22393	21634	34560	77404
29	93074	27086	62559	86590	18420	33290
30	90549	53094	76282	53105	45531	90061
31	11373	96871	38157	98368	39536	08079
32	52022	59093	30647	33241	16027	70336
33	14709	93220	89547	95320	39134	07646
34	57584	28114	91168	16320	81609	60807
35	31867	85872	91430	45554	21567	15082
36	07033	75250	34546	75298	33893	64487
37	02779	72645	32699	86009	73729	44206
38	24512	01116	49826	50882	44086	87757
39	52463	30164	80073	55917	60995	38655
40	82588	59267	13570	56434	66413	99518
41	20999	05039	87835	63010	82980	66193
42	09084	98948	09541	80623	15915	71042

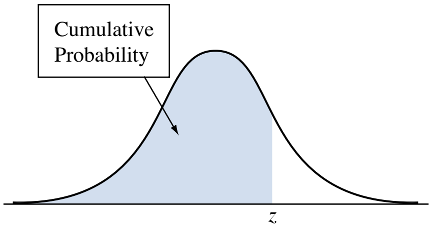


Cumulative probability for z is the area under the standard Normal curve to the left of z.

Table 2: Standard Normal Cumulative Probabilities

z	.00
−5.0	.000000287
−4.5	.00000340
−4.0	.0000317
−3.5	.000233

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641



Cumulative probability for z is the area under the standard Normal curve to the left of z.

Standard Normal Cumulative Probabilities (*continued*)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

z	.00
3.5	.999767
4.0	.9999683
4.5	.9999966
5.0	.999999713

Table 3: Binomial Probabilities

n	x	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	x
2	0	.980	.902	.810	.640	.490	.360	.250	.160	.090	.040	.010	.002	0+	0
	1	.020	.095	.180	.320	.420	.480	.500	.480	.420	.320	.180	.095	.020	1.
	2	0+	.002	.010	.040	.090	.160	.250	.360	.490	.640	.810	.902	.980	2
3	0	.970	.857	.729	.512	.343	.216	.125	.064	.027	.008	.001	0+	0+	0
	1	.029	.135	.243	.384	.441	.432	.375	.288	.189	.096	.027	.007	0+	1
	2	0+	.007	.027	.096	.189	.288	.375	.432	.441	.384	.243	.135	.029	2
	3	0+	0+	.001	.008	.027	.064	.125	.216	.343	.512	.729	.857	.970	3
4	0	.961	.815	.656	.410	.240	.130	.062	.026	.008	.002	0+	0+	0+	0
	1	.039	.171	.292	.410	.412	.346	.250	.154	.076	.026	.004	0+	0+	1
	2	.001	.014	.049	.154	.265	.346	.375	.346	.265	.154	.049	.014	.001	2
	3	0+	0+	.004	.026	.076	.154	.250	.346	.412	.410	.292	.171	.039	3
	4	0+	0+	0+	.002	.008	.026	.062	.130	.240	.410	.656	.815	.961	4
5	0	.951	.774	.590	.328	.168	.078	.031	.010	.002	0+	0+	0+	0+	0
	1	.048	.204	.328	.410	.360	.259	.156	.077	.028	.006	0+	0+	0+	1
	2	.001	.021	.073	.205	.309	.346	.312	.230	.132	.051	.008	.001	0+	2
	3	0+	.001	.008	.051	.132	.230	.312	.346	.309	.205	.073	.021	.001	3
	4	0+	0+	0+	.006	.028	.077	.156	.259	.360	.410	.328	.204	.048	4
	5	0+	0+	0+	0+	.002	.010	.031	.078	.168	.328	.590	.774	.951	5
6	0	.941	.735	.531	.262	.118	.047	.016	.004	.001	0+	0+	0+	0+	0
	1	.057	.232	.354	.393	.303	.187	.094	.037	.010	.002	0+	0+	0+	1
	2	.001	.031	.098	.246	.324	.311	.234	.138	.060	.015	.001	0+	0+	2
	3	0+	.002	.015	.082	.185	.276	.312	.276	.185	.082	.015	.002	0+	3
	4	0+	0+	.001	.015	.060	.138	.234	.311	.324	.246	.098	.031	.001	4
	5	0+	0+	0+	.002	.010	.037	.094	.187	.303	.393	.354	.232	.057	5
	6	0+	0+	0+	0+	.001	.004	.016	.047	.118	.262	.531	.735	.941	6
7	0	.932	.698	.478	.210	.082	.028	.008	.002	0+	0+	0+	0+	0+	0
	1	.066	.257	.372	.367	.247	.131	.055	.017	.004	0+	0+	0+	0+	1
	2	.002	.041	.124	.275	.318	.261	.164	.077	.025	.004	0+	0+	0+	2
	3	0+	.004	.023	.115	.227	.290	.273	.194	.097	.029	.003	0+	0+	3
	4	0+	0+	.003	.029	.097	.194	.273	.290	.227	.115	.023	.004	0+	4
	5	0+	0+	0+	.004	.025	.077	.164	.261	.318	.275	.124	.041	.002	5
	6	0+	0+	0+	0+	.004	.017	.055	.131	.247	.367	.372	.257	.066	6
	7	0+	0+	0+	0+	0+	.002	.008	.028	.082	.210	.478	.698	.932	7
8	0	.923	.663	.430	.168	.058	.017	.004	.001	0+	0+	0+	0+	0+	0
	1	.075	.279	.383	.336	.198	.090	.031	.008	.001	0+	0+	0+	0+	1
	2	.003	.051	.149	.294	.296	.209	.109	.041	.010	.001	0+	0+	0+	2
	3	0+	.005	.033	.147	.254	.279	.219	.124	.047	.009	0+	0+	0+	3
	4	0+	0+	.005	.046	.136	.232	.273	.232	.136	.046	.005	0+	0+	4
	5	0+	0+	0+	.009	.047	.124	.219	.279	.254	.147	.033	.005	0+	5
	6	0+	0+	0+	.001	.010	.041	.109	.209	.296	.294	.149	.051	.003	6
	7	0+	0+	0+	0+	.001	.008	.031	.090	.198	.336	.383	.279	.075	7
	8	0+	0+	0+	0+	0+	.001	.004	.017	.058	.168	.430	.663	.923	8

NOTE: 0+ represents a positive probability less than 0.0005.

(*continued*)

Binomial Probabilities (*continued*)

		p													
n	x	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	x
9	0	.914	.630	.387	.134	.040	.010	.002	0+	0+	0+	0+	0+	0+	0
	1	.083	.299	.387	.302	.156	.060	.018	.004	0+	0+	0+	0+	0+	1
	2	.003	.063	.172	.302	.267	.161	.070	.021	.004	0+	0+	0+	0+	2
	3	0+	.008	.045	.176	.267	.251	.164	.074	.021	.003	0+	0+	0+	3
	4	0+	.001	.007	.066	.172	.251	.246	.167	.074	.017	.001	0+	0+	4
	5	0+	0+	.001	.017	.074	.167	.246	.251	.172	.066	.007	.001	0+	5
	6	0+	0+	0+	.003	.021	.074	.164	.251	.267	.176	.045	.008	0+	6
	7	0+	0+	0+	0+	.004	.021	.070	.161	.267	.302	.172	.063	.003	7
	8	0+	0+	0+	0+	0+	.004	.018	.060	.156	.302	.387	.299	.083	8
	9	0+	0+	0+	0+	0+	0+	.002	.010	.040	.134	.387	.630	.914	9
10	0	.904	.599	.349	.107	.028	.006	.001	0+	0+	0+	0+	0+	0+	0
	1	.091	.315	.387	.268	.121	.040	.010	.002	0+	0+	0+	0+	0+	1
	2	.004	.075	.194	.302	.233	.121	.044	.011	.001	0+	0+	0+	0+	2
	3	0+	.010	.057	.201	.267	.215	.117	.042	.009	.001	0+	0+	0+	3
	4	0+	.001	.011	.088	.200	.251	.205	.111	.037	.006	0+	0+	0+	4
	5	0+	0+	.001	.026	.103	.201	.246	.201	.103	.026	.001	0+	0+	5
	6	0+	0+	0+	.006	.037	.111	.205	.251	.200	.088	.011	.001	0+	6
	7	0+	0+	0+	.001	.009	.042	.117	.215	.267	.201	.057	.010	0+	7
	8	0+	0+	0+	0+	.001	.011	.044	.121	.233	.302	.194	.075	.004	8
	9	0+	0+	0+	0+	0+	.002	.010	.040	.121	.268	.387	.315	.091	9
	10	0+	0+	0+	0+	0+	0+	.001	.006	.028	.107	.349	.599	.904	10
11	0	.895	.569	.314	.086	.020	.004	0+	0+	0+	0+	0+	0+	0+	0
	1	.099	.329	.384	.236	.093	.027	.005	.001	0+	0+	0+	0+	0+	1
	2	.005	.087	.213	.295	.200	.089	.027	.005	.001	0+	0+	0+	0+	2
	3	0+	.014	.071	.221	.257	.177	.081	.023	.004	0+	0+	0+	0+	3
	4	0+	.001	.016	.111	.220	.236	.161	.070	.017	.002	0+	0+	0+	4
	5	0+	0+	.002	.039	.132	.221	.226	.147	.057	.010	0+	0+	0+	5
	6	0+	0+	0+	.010	.057	.147	.226	.221	.132	.039	.002	0+	0+	6
	7	0+	0+	0+	.002	.017	.070	.161	.236	.220	.111	.016	.001	0+	7
	8	0+	0+	0+	0+	.004	.023	.081	.177	.257	.221	.071	.014	0+	8
	9	0+	0+	0+	0+	.001	.005	.027	.089	.200	.295	.213	.087	.005	9
	10	0+	0+	0+	0+	0+	.001	.005	.027	.093	.236	.384	.329	.099	10
	11	0+	0+	0+	0+	0+	0+	0+	.004	.020	.086	.314	.569	.895	11
12	0	.886	.540	.282	.069	.014	.002	0+	0+	0+	0+	0+	0+	0+	0
	1	.107	.341	.377	.206	.071	.017	.003	0+	0+	0+	0+	0+	0+	1
	2	.006	.099	.230	.283	.168	.064	.016	.002	0+	0+	0+	0+	0+	2
	3	0+	.017	.085	.236	.240	.142	.054	.012	.001	0+	0+	0+	0+	3
	4	0+	.002	.021	.133	.231	.213	.121	.042	.008	.001	0+	0+	0+	4
	5	0+	0+	.004	.053	.158	.227	.193	.101	.029	.003	0+	0+	0+	5
	6	0+	0+	0+	.016	.079	.177	.226	.177	.079	.016	0+	0+	0+	6
	7	0+	0+	0+	.003	.029	.101	.193	.227	.158	.053	.004	0+	0+	7
	8	0+	0+	0+	.001	.008	.042	.121	.213	.231	.133	.021	.002	0+	8
	9	0+	0+	0+	0+	.001	.012	.054	.142	.240	.236	.085	.017	0+	9
	10	0+	0+	0+	0+	0+	.002	.016	.064	.168	.283	.230	.099	.006	10
	11	0+	0+	0+	0+	0+	0+	.003	.017	.071	.206	.377	.341	.107	11
	12	0+	0+	0+	0+	0+	0+	0+	.002	.014	.069	.282	.540	.886	12

NOTE: 0+ represents a positive probability less than 0.0005.

(*continued*)

Binomial Probabilities (*continued*)

		p													
n	***x***	**.01**	**.05**	**.10**	**.20**	**.30**	**.40**	**.50**	**.60**	**.70**	**.80**	**.90**	**.95**	**.99**	***x***
13	0	.878	.513	.254	.055	.010	.001	0+	0+	0+	0+	0+	0+	0+	0
	1	.115	.351	.367	.179	.054	.011	.002	0+	0+	0+	0+	0+	0+	1
	2	.007	.111	.245	.268	.139	.045	.010	.001	0+	0+	0+	0+	0+	2
	3	0+	.021	.100	.246	.218	.111	.035	.006	.001	0+	0+	0+	0+	3
	4	0+	.003	.028	.154	.234	.184	.087	.024	.003	0+	0+	0+	0+	4
	5	0+	0+	.006	.069	.180	.221	.157	.066	.014	.001	0+	0+	0+	5
	6	0+	0+	.001	.023	.103	.197	.209	.131	.044	.006	0+	0+	0+	6
	7	0+	0+	0+	.006	.044	.131	.209	.197	.103	.023	.001	0+	0+	7
	8	0+	0+	0+	.001	.014	.066	.157	.221	.180	.069	.006	0+	0+	8
	9	0+	0+	0+	0+	.003	.024	.087	.184	.234	.154	.028	.003	0+	9
	10	0+	0+	0+	0+	.001	.006	.035	.111	.218	.246	.100	.021	0+	10
	11	0+	0+	0+	0+	0+	.001	.010	.045	.139	.268	.245	.111	.007	11
	12	0+	0+	0+	0+	0+	0+	.002	.011	.054	.179	.367	.351	.115	12
	13	0+	0+	0+	0+	0+	0+	0+	.001	.010	.055	.254	.513	.878	13
14	0	.869	.488	.229	.044	.007	.001	0+	0+	0+	0+	0+	0+	0+	0
	1	.123	.359	.356	.154	.041	.007	.001	0+	0+	0+	0+	0+	0+	1
	2	.008	.123	.257	.250	.113	.032	.006	.001	0+	0+	0+	0+	0+	2
	3	0+	.026	.114	.250	.194	.085	.022	.003	0+	0+	0+	0+	0+	3
	4	0+	.004	.035	.172	.229	.155	.061	.014	.001	0+	0+	0+	0+	4
	5	0+	0+	.008	.086	.196	.207	.122	.041	.007	0+	0+	0+	0+	5
	6	0+	0+	.001	.032	.126	.207	.183	.092	.023	.002	0+	0+	0+	6
	7	0+	0+	0+	.009	.062	.157	.209	.157	.062	.009	0+	0+	0+	7
	8	0+	0+	0+	.002	.023	.092	.183	.207	.126	.032	.001	0+	0+	8
	9	0+	0+	0+	0+	.007	.041	.122	.207	.196	.086	.008	0+	0+	9
	10	0+	0+	0+	0+	.001	.014	.061	.155	.229	.172	.035	.004	0+	10
	11	0+	0+	0+	0+	0+	.003	.022	.085	.194	.250	.114	.026	0+	11
	12	0+	0+	0+	0+	0+	.001	.006	.032	.113	.250	.257	.123	.008	12
	13	0+	0+	0+	0+	0+	0+	.001	.007	.041	.154	.356	.359	.123	13
	14	0+	0+	0+	0+	0+	0+	0+	.001	.007	.044	.229	.488	.869	14
15	0	.860	.463	.206	.035	.005	0+	0+	0+	0+	0+	0+	0+	0+	0
	1	.130	.366	.343	.132	.031	.005	0+	0+	0+	0+	0+	0+	0+	1
	2	.009	.135	.267	.231	.092	.022	.003	0+	0+	0+	0+	0+	0+	2
	3	0+	.031	.129	.250	.170	.063	.014	.002	0+	0+	0+	0+	0+	3
	4	0+	.005	.043	.188	.219	.127	.042	.007	.001	0+	0+	0+	0+	4
	5	0+	.001	.010	.103	.206	.186	.092	.024	.003	0+	0+	0+	0+	5
	6	0+	0+	.002	.043	.147	.207	.153	.061	.012	.001	0+	0+	0+	6
	7	0+	0+	0+	.014	.081	.177	.196	.118	.035	.003	0+	0+	0+	7
	8	0+	0+	0+	.003	.035	.118	.196	.177	.081	.014	0+	0+	0+	8
	9	0+	0+	0+	.001	.012	.061	.153	.207	.147	.043	.002	0+	0+	9
	10	0+	0+	0+	0+	.003	.024	.092	.186	.206	.103	.010	.001	0+	10
	11	0+	0+	0+	0+	.001	.007	.042	.127	.219	.188	.043	.005	0+	11
	12	0+	0+	0+	0+	0+	.002	.014	.063	.170	.250	.129	.031	0+	12
	13	0+	0+	0+	0+	0+	0+	.003	.022	.092	.231	.267	.135	.009	13
	14	0+	0+	0+	0+	0+	0+	0+	.005	.031	.132	.343	.366	.130	14
	15	0+	0+	0+	0+	0+	0+	0+	0+	.005	.035	.206	.463	.860	15

NOTE: 0+ represents a positive probability less than 0.0005.

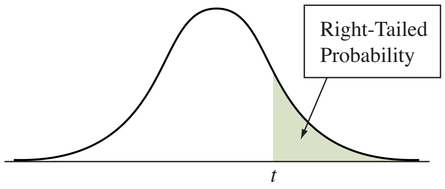


Table 4: *t*-Distribution Critical Values

	Confidence Level					
	80%	**90%**	**95%**	**98%**	**99%**	**99.8%**
	Right-Tailed Probability					
df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	$t_{.001}$
1	3.078	6.314	12.706	31.821	63.656	318.289
2	1.886	2.920	4.303	6.965	9.925	22.328
3	1.638	2.353	3.182	4.541	5.841	10.214
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.894
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.611
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.704	3.307
50	1.299	1.676	2.009	2.403	2.678	3.261
60	1.296	1.671	2.000	2.390	2.660	3.232
80	1.292	1.664	1.990	2.374	2.639	3.195
100	1.290	1.660	1.984	2.364	2.626	3.174
∞	1.282	1.645	1.960	2.326	2.576	3.091

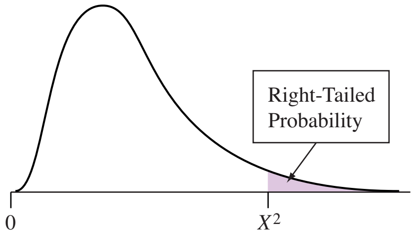


Table 5: Chi-Square Distribution for Values of Various Right-Tailed Probabilities

	Right-Tailed Probability						
df	**0.250**	**0.100**	**0.050**	**0.025**	**0.010**	**0.005**	**0.001**
1	1.32	2.71	3.84	5.02	6.63	7.88	10.83
2	2.77	4.61	5.99	7.38	9.21	10.60	13.82
3	4.11	6.25	7.81	9.35	11.34	12.84	16.27
4	5.39	7.78	9.49	11.14	13.28	14.86	18.47
5	6.63	9.24	11.07	12.83	15.09	16.75	20.52
6	7.84	10.64	12.59	14.45	16.81	18.55	22.46
7	9.04	12.02	14.07	16.01	18.48	20.28	24.32
8	10.22	13.36	15.51	17.53	20.09	21.96	26.12
9	11.39	14.68	16.92	19.02	21.67	23.59	27.88
10	12.55	15.99	18.31	20.48	23.21	25.19	29.59
11	13.70	17.28	19.68	21.92	24.72	26.76	31.26
12	14.85	18.55	21.03	23.34	26.22	28.30	32.91
13	15.98	19.81	22.36	24.74	27.69	29.82	34.53
14	17.12	21.06	23.68	26.12	29.14	31.32	36.12
15	18.25	22.31	25.00	27.49	30.58	32.80	37.70
16	19.37	23.54	26.30	28.85	32.00	34.27	39.25
17	20.49	24.77	27.59	30.19	33.41	35.72	40.79
18	21.60	25.99	28.87	31.53	34.81	37.16	42.31
19	22.72	27.20	30.14	32.85	36.19	38.58	43.82
20	23.83	28.41	31.41	34.17	37.57	40.00	45.32
25	29.34	34.38	37.65	40.65	44.31	46.93	52.62
30	34.80	40.26	43.77	46.98	50.89	53.67	59.70
40	45.62	51.80	55.76	59.34	63.69	66.77	73.40
50	56.33	63.17	67.50	71.42	76.15	79.49	86.66
60	66.98	74.40	79.08	83.30	88.38	91.95	99.61
70	77.58	85.53	90.53	95.02	100.43	104.21	112.32
80	88.13	96.58	101.88	106.63	112.33	116.32	124.84
90	98.65	107.57	113.15	118.14	124.12	128.30	137.21
100	109.14	118.50	124.34	129.56	135.81	140.17	149.45

Appendix B: Answers to Check Your Tech

CHAPTER 3

Finding the Standard Deviation of Vacation Days for Several Countries

1: $180/6 = 30$

2:

x	$x - \bar{x}$	$(x - \bar{x})^2$
13	$13 - 30 = -17$	$(-17)^2 = 289$
25	$25 - 30 = -5$	$(-5)^2 = 25$
42	$42 - 30 = 12$	$12^2 = 144$
37	$37 - 30 = 7$	$7^2 = 49$
35	$35 - 30 = 5$	$5^2 = 25$
28	$28 - 30 = -2$	$(-2)^2 = 4$

3: $289 + 25 + 144 + 49 + 25 + 4 = 536$

4: $536/5 = 107.2$

5: $\sqrt{107.2} = 10.3537$, which rounds to 10.35. This is the same value shown in Figure A.

Making a Boxplot of the Area of Western States

1: The axis is shown in the graph.
2: The box goes from 84.5 on the left to 134.5 on the right.
3: The median line is at 104, which is Colorado.
4: $\text{IQR} = 134.5 - 84.5 = 50$
5: Lower limit $= \text{Q1} - (1.5 \times \text{IQR}) = 84.5 - (1.5 \times 50) = 84.5 - 75 = 9.5$
The smallest area is 11, which is not below 9.5, so there are no low-end potential outliers.
6: The low whisker goes down to 11 (Hawaii).
7: Upper limit $= \text{Q3} + (1.5 \times \text{IQR}) = 134.5 + (1.5 \times 50) = 134.5 + 75 = 209.5$
The only area larger than 209.5 is 656 (Alaska), which is an outlier and should have a separate mark.
8: The upper whisker goes to 164 (California), which is the largest area that is not a potential outlier.

CHAPTER 4

1:

x	$x - \bar{x}$	z_x	y	$y - \bar{y}$	z_y	$z_x z_y$
20	$20 - 30 = -10$	$-10/10 = -1$	20	$20 - 25 = -5$	$-5/5 = -1$	$(-1) \times (-1) = 1$
30	$30 - 30 = 0$	$0/10 = 0$	30	$30 - 25 = 5$	$5/5 = 1$	$0 \times (1) = 0$
40	$40 - 30 = 10$	$10/10 = 1$	25	$25 - 25 = 0$	$0/5 = 0$	$1 \times 0 = 0$

2: Add the last column to get $\sum z_x z_y = 1 + 0 + 0 = 1$.

3: Find the correlation and check it with the output.

$$r = \frac{\sum z_x z_y}{n - 1} = \frac{1}{3 - 1} = \frac{1}{2} = 0.5$$

This is the same correlation as the output.

4: Find the slope.

$$b = r\frac{s_y}{s_x} = r \times \frac{5}{10} = 0.50 \times (0.50) = 0.25$$

5: Find the y-intercept.

$$a = \bar{y} - b\bar{x} = 25 - 0.25 \times 30 = 25 - 7.5 = 17.5$$

6: Finally, put together the equation:

$$y = a + bx$$

$$\text{Predicted Wife} = a + b\ \text{Husband}$$

$$\text{Predicted Wife} = 17.5 + 0.25\ \text{Husband}$$

The equation is the same as the Minitab output.

CHAPTER 9

1: $92.5143 - 73.6 = 18.9143$ (which is also shown in the Minitab Output)

2: $SE_{\text{diff}} = \dfrac{S_{\text{diff}}}{\sqrt{n}} = \dfrac{15.0497}{\sqrt{35}} = \dfrac{15.0497}{5.9161} = 2.5439$

3: $t = \dfrac{\overline{X}_{\text{diff}} - 0}{SE_{\text{diff}}} = \dfrac{18.9143}{2.5439} = 7.44$

4: a. If the means were farther apart, that would cause t to be larger. Because the difference between means is in the numerator, a bigger difference in the numerator means a value that is larger. b. If the standard deviation (S_{diff}) were larger, that would cause t to be smaller, because the standard deviation is in the denominator, and a larger denominator results in a smaller value. c. If the sample size were larger, that would cause t to be larger. The larger sample size would cause SE_{diff} to be smaller (because the sample size is in the denominator of SE_{diff}), and the smaller SE_{diff} would cause t to be larger (because SE_{diff} is in the denominator of t).

Appendix C: Answers to Odd-Numbered Exercises

CHAPTER 1

Section 1.2

1.1 Nine (9)

1.3 a. Handedness: categorical **b.** Age: numerical

1.5 Answers will vary but could include such things as number of friends on Facebook or foot length. *Don't copy these answers.*

1.7 The label would be "Brown Eyes" and there would be 8 ones and 3 zeros.

1.9 Male is categorical with two categories. The 1's represent males, and the 0's represent females. If you added the numbers, you would get the number of males, so it makes sense here.

1.11 a. Stacked **b.** 1 means male, and 0 means female.

c.

Female	Male
9.5	9.4
9.5	9.5
9.9	9.5
	9.7

1.13 a. Stacked and coded

Duration (in minutes)	Morning
45	1
90	1
75	1
35	1
60	1
80	1
50	0
65	0
45	0
70	0

Alternatively, you could label the second column above "Afternoon" and then the 1's would become 0's and the 0's would become 1's.

b. Unstacked

Morning	Afternoon
45	50
90	65
75	45
35	70
60	
80	

Section 1.3

1.15 a.

	Men	Women	Total
Yes, Older S	12	55	67
No Older S	11	39	50
	23	94	117

b. $12/23 = 52.2\%$

c. $11/23 = 47.8\%$

d. $55/94 = 58.5\%$

e. $67/117 = 57.3\%$

f. $55/67 = 82.1\%$

g. $0.585 \times 600 = 351$

1.17 a. $17/42 = 40.5\%$ of the marbles are blue **b.** $0.63(430) = 270.9$, or about 271 blue marbles in the jar.

c. $0.45(x) = 90$
$90/0.45 = 200$ marbles in the jar

1.19 The frequency of women is 7, the proportion is $7/11$, and the percentage is 63.6%.

1.21 The answers follow the guidance given on page 57.

a. and **b.**

	Male	Female	Total
Right	4	5	9
Left	0	2	2
Total	4	7	11

c. $5/7 = 71.4\%$

d. $5/9 = 55.6\%$

e. $9/11 = 81.8\%$

f. $0.714(70) = 50$

1.23 $0.202x = 88{,}547{,}000$
$x = 438{,}351{,}485$ or a rounded version of this

1.25 The answers follow the guidance starting on page 57. Steps 1–3 are shown in the accompanying table.

State	AIDS/ HIV	Rank Cases	Population	Population (thousands)	AIDS/ HIV per 1000	Rank Rate
New York	192,753	1	19,421,055	19,421	9.92	2
California	160,293	2	37,341,989	37,342	4.29	5
Florida	117,612	3	18,900,773	18,901	6.22	3
Texas	77,070	4	25,258,418	25,258	3.05	6
New Jersey	54,557	5	8,807,501	8,808	6.19	4
District of Columbia	9,257	6	601,723	602	15.38	1

4: No, the ranks are not the same. The District of Columbia had the highest rate and had the lowest number of cases. (Also, the rate for Florida puts its rank above California, and the rate for New Jersey puts it above Texas in ranking.)

5: The District of Columbia is the place (among these six regions) where you would be most likely to meet a person diagnosed with AIDS/HIV, and Texas is the place (among these six regions) where you would be least likely to do so.

1.27 1990: 58.7%, 1997: 56.4%, 2000: 56.2%, 2007: 55.1%
The percentage of married people is decreasing over time (at least with these dates).

1.29 We don't know the percentage of female students in the two classes. The larger number of women at 8 a.m. may just result from a larger number of students at 8 a.m., which may be because the class can accommodate more students because perhaps it is in a large lecture hall.

Section 1.4

1.31 Controlled experiment

1.33 Observational study

1.35 Observational study

1.37 Controlled experiment

1.39 This is an observational study and, therefore, a cause-effect relationship between higher grades and use of audio-visual tools cannot be established. Possible confounders (answers will vary): 1. Those who chose the audio-visual class could have already been serious about their grades. 2. A larger group of students might have chosen to attend the audio-visual class, resulting in better students moving to the group.

1.41 **a.** It was a controlled experiment, as you can tell by the random assignment. This tells us that the researchers determined who received which treatment. **b.** We can conclude that the early surgery caused the better outcomes, because it was a randomized controlled experiment.

1.43 Written answers will vary. However, they should all mention randomly dividing the 100 people into two groups and giving one group the copper bracelets. The other group could be given (as a placebo) bracelets that look like copper but are made of some other material. Then the pain levels after treatment could be compared.

1.45 No. This was an observational study, because researchers could not have deliberately exposed people to weed killers. There was no random assignment, and no one would randomly assign a person to be exposed to pesticides. From an observational study, you cannot conclude causation. This is why the report was careful to use the phrase *associated with* rather than the word *caused*.

1.47 Ask whether the patients were randomly assigned the full or the half dose. Without randomization there could be bias, and we cannot infer causation. With randomization we can infer causation.

1.49 This was an observational study: vitamin C and breast milk. We cannot conclude cause and effect from observational studies.

1.51 **a.** LD: 8% tumors; LL: 28% tumors **b.** A controlled experiment. You can tell by the random assignment. **c.** Yes, we can conclude cause and effect because it was a controlled experiment, and random assignment will balance out potential confounding variables.

Chapter Review Exercises

1.53 **a.** Dating: $81/440 = 18.4\%$ **b.** Cohabiting: $103/429 = 24.0\%$ **c.** Married: $147/424 = 34.7\%$ **d.** No, this was an observational study. Confounding variables may vary. Perhaps married people are likely to be older, and older people are more likely to be obese.

1.55 **a.** The two-way table follows.

	Boy	**Girl**
Violent	10	11
Nonviolent	19	4

b. For the boys, 10/29, or 34.5%, were on probation for violent crime. For the girls, 11/15, or 73.3%, were on probation for violent crime. **c.** The girls were more likely to be on probation for violent crime.

1.57 Answers will vary. *Students should not copy the words they see in these answers.* Randomly divide the group in half, using a coin flip for each woman: Heads she gets the vitamin D, and tails she gets the placebo (or vice versa). Make sure that neither the women themselves nor any of the people who come in contact with them know whether they got the treatment or the placebo ("double-blind"). Over a given length of time (such as three years), note which women had broken bones and which did not. Compare the percentage of women with broken bones in the vitamin D group with the percentage of women with broken bones in the placebo group.

1.59 **a.** The treatment variable was Medicaid expansion or not and the response variables were the death rate and the rate of people who reported their health as excellent or very good. **b.** This was observational. Researchers did not assign people either to receive or not to receive Medicaid. **c.** No, this was an observational study. From an observational study, you cannot conclude causation. It is possible that other factors that differed between the states caused the change.

1.61 No, we cannot conclude causation. There was no control group for comparison, and the sample size was very small.

CHAPTER 2

Sections 2.1 and 2.2

2.1 **a.** 11 are morbidly obese. **b.** 11/134 is about 8%, which is much more than 3%.

2.3 New vertical axis labels: 0.06, 0.12, 0.18, 0.24, 0.30. Note that 0.06 comes from 1/17.

2.5 **a.** 4 have no paper shredders **b.** 7 paper shredders **c.** Between 40 and 45 **d.** Around 10 **e.** 10/120 or, 0.083

2.7 **a.** Both dotplots are right-skewed. The dotplot for the females is also multimodal. **b.** The females tend to have more pairs of shoes. **c.** The numbers of pairs for the females are more spread out. The males' responses tend to be clustered at about 10 pairs or fewer.

2.9 Left-skewed. There will be a lot of players who practiced for 4 or more hours and few who practiced for 0 or 1 hour.

2.11 It would be bimodal because the men and women tend to have different heights and therefore different armspans.

2.13 About 28 years (between 24 and 32)

2.15 Typing the letter using a typewriter shows a larger typical value and also more variation.

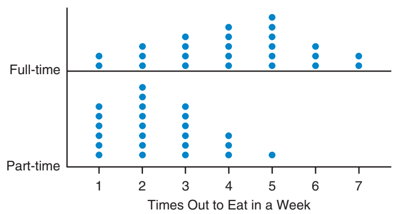
2.17 **a.** Multimodal with modes at 12 years (high school), 14 years (junior college), 16 years (bachelor's degree), and 18 years (possible master's degree). It is also left-skewed with numbers as low as 0. **b.** Estimate: 300 + 50 + 100 + 40 + 50, or about 500 to 600, had 16 or more years. **c.** This is between 25% (from 500/2018) and 30% (from 600/2018) have a bachelor's degree or higher. This is very similar to the 27% given.

2.19 Both graphs go from about 0 to about 20, but the data for years of formal education for the respondents (compared to their mothers) include more with education above 12 years. For example, the bar at 16 (college bachelor's degree) is higher for the respondents than for the mothers, which shows that the respondents tend to have a bit more education than their mothers. Also, the bar at 12 is taller for the mothers, showing that the mothers were more likely to get only a high school diploma. Furthermore, the bar graph for the mothers includes more people (taller bars) at lower numbers of years, such as 0 and 3 and 6.

2.21 1 is B, 2 is A, and 3 is C.

2.23 **1.** C **2.** B **3.** A

2.25 The answers follow the guidance given on page 100. **1.** See the dotplots. Histograms would also be good for visualizing the distributions. Stemplots would not work with these data sets because all the observed values have only one digit.



2. Full-time is a bit left-skewed, and part-time is a bit right-skewed.

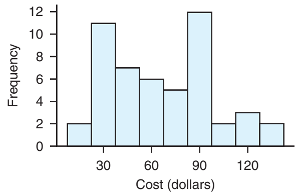
3. Those with full-time jobs tend to go out to eat more than those with part-time jobs.

4. The full-time workers have a distribution that is more spread out; full-time goes from 1 to 7, whereas part-time goes only from 1 to 5.

5. There are no outliers—that is, no dots detached from the main group with an empty space between.

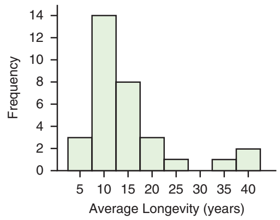
6. For the full-time workers the distribution is a little left-skewed, and for the part-time workers it is a little right-skewed. The full-time workers tend to go out to eat more, and their distribution is more spread out.

2.27 See histogram.

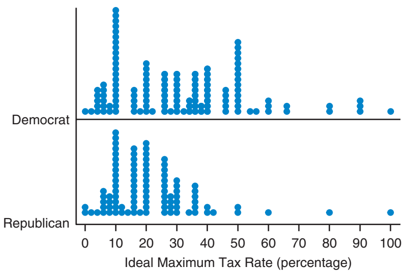


The histogram is bimodal with modes at about \$30 and about \$90.

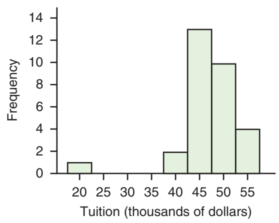
2.29 See histogram. The histogram is right-skewed. The typical value is around 12 (between 10 and 15) years, and there are three outliers: Asian elephant (40), African elephant (35), and hippo (41). Humans (75 years) would be way off to the right; they live much longer than other mammals.



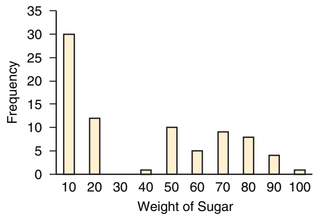
2.31 Both graphs are multimodal and right-skewed. The Democrats have a higher typical value, as shown by the fact that the center is roughly around 35 or 40%, while the center value for the Republicans is closer to 20 to 30%. Also note the much larger proportion of Democrats who think the rate should be 50% or higher. The distribution for the Democrats appears more spread out because the Democrats have a greater proportion of people responding with both lower and higher percentages.



2.33 The distribution appears left-skewed because of the low-end outlier at about \$20,000 (Brigham Young University).



2.35 With this grouping, the distribution appears unimodal with modes at about 0–10 g of sugar. (With fewer—that is wider—bins, it may not appear unimodal). There are high-end outliers at 90–100 g. The distribution is right-skewed.



Sections 2.3 and 2.4

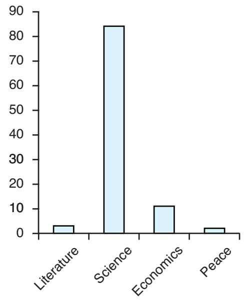
2.37 No, the largest category is Wrong to Right, which suggests that changes tend to make the answers more likely to be right.

2.39 **a.** 80 to 82% **b.** Truth. Almost all observations are in the Top Fifth category. **c.** Ideal. These are almost uniformly spread across the five groups. **d.** They underestimate the proportion of wealth held by the top 20%.

2.41 **a.** Dem (not strong) **b.** Other. It is easier to pick out the second tallest bar in the bar chart. (Answers may vary.)

2.43 **a.** The percentage of old people is increasing, the percentage of those 25–64 is decreasing, and the percentage of those 24 and below is relatively constant. **b.** The money for Social Security normally comes from those in a working age range (which includes those 25–64), and that group is decreasing in percentage. Also, the group receiving Social Security (those 65 and older) is becoming larger. This suggests that in the future, Social Security might not get enough money from the workers to support the old people.

2.45 A Pareto chart or pie chart would also be appropriate.



Note that the mode is Science and that there is substantial variation. (Of course, individual categories such as Physics, Medicine, and Chemistry were grouped into Science.)

Section 2.5

2.47 The graph is a histogram (the bars touch) and histograms are used for numerical data. However, this data set is categorical and the numbers 0 and 1 represent categories. A more appropriate graph would be a bar graph or a pie chart.

2.49 Hours of sleep is a numerical variable. A histogram or dotplot would better enable us to see the distribution of values. Because there are so many possible numerical values, this pie chart has so many "slices" that it is difficult to tell which is which.

2.51 Those who still play tended to have practiced more as teenagers, which we can see because the center of the distribution for those who still play is about 2 or 2.5 hours, compared to only about 1 or 1.5 hours for those who do not. The distribution could be displayed as a pair of histograms or a pair of dotplots.

Chapter Review Exercises

2.53 Histograms, dotplots, and stemplots would be appropriate. Two separately colored bins or dots each for the boys and the girls would work for this numerical data set.

2.55 **a.** The diseases with higher rates for HRT were heart disease, stroke, pulmonary embolism, and breast cancer. The diseases with lower rates for HRT were endometrial cancer, colorectal cancer, and hip fracture.
b. Comparing the rates makes more sense than comparing just the numbers, in case there were more women in one group than in the other.

2.57 The vertical axis does not start at zero and exaggerates the differences. Make a graph for which the vertical axis starts at zero.

2.59 The shapes are roughly bell-shaped and symmetric; the later period is warmer, but the spread is similar. This is consistent with theories on global warming. The difference is $57.9 - 56.7 = 1.2$, so the difference is only a bit more than 1 degree Fahrenheit.

2.61 **a.** The graph shows that a greater percentage of people survived when lying prone (on their stomachs) than when lying supine (on their backs). This suggests that we should recommend that doctors ask these patients to lie prone. **b.** Both variables (*Position* and *Outcome*) are categorical, so a bar chart is appropriate.

2.63 The created 10-point dotplots will vary. The dotplot should have skew.

2.65 Graphs will vary. Histograms and dotplots are both appropriate. For the group without a camera the distribution is roughly symmetrical, and for the group with a camera it is right-skewed. Both are unimodal. The number of cars going through a yellow light tends to be less at intersections with cameras. Also, there is more variation in the intersections without cameras.

2.67 Both distributions are right-skewed. The typical speed for the men (a little above 100 mph) is a bit higher than the typical speed for the women (which appears to be closer to 90 mph). The spread for the men is larger primarily because of the outlier of 200 mph for the men.

2.69 It should be left-skewed.

2.71 **a.** The tallest bar is Wrong to Right, which suggests that the instruction was correct. **b.** For both instructors, the largest group is Wrong to Right, so it appears that changes made tend to raise the grades of the students.

CHAPTER 3

IQRs vary with different technology.

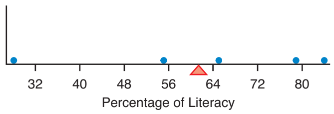
Section 3.1

3.1 a

3.3 The typical age of the CEOs is between about 56 and 60 (or any number from 56 to 60). (The distribution is symmetric, so the mean should be about in the middle.)

3.5 **a.** The mean percentage of literacy in the five countries is 62.2.

b.



c. The standard deviation of the five percentages of literacy is 22.29.
d. The percentage farthest from the mean is 28, which is the lowest rate of literacy.

3.7 **a.** The typical number of vacation days is about 17.3, the mean. One could also say that number of vacation days tends to be about 17.3, on average. **b.** The standard deviation of paid vacation days for the six countries is 10.7 **c.** The United States, at 0, is farthest from the mean and contributes most to the standard deviation.

3.9 **a.** 76.6; More power: 476.6 (the mean power of top 5 cars) is more than 402.4 for the next 5 cars. **b.** More: 214.26 (the standard deviation of power of top 5 cars) reveals much more variability than the 162.07 for the next 5 cars.

3.11 **a.** $120.04 - 140.88 = -20.84$. The negative value means desired weight is less than actual weight, and so women want to lose on average 21 pounds. **b.** $169.62 - 187.6 = -17.98$, and so on average men want to lose about 19 pounds. **c.** The women wanted to lose more weight on average. **d.** The standard deviation for the men (38.9) was larger than the standard deviation for the women (22.29).

3.13 **a.** The mean for September is 4.8, which is less than the mean for October, which is 4.9. So more students tend to be absent in October.
b. The standard deviation of 3.316 for September was less than the standard deviation of 3.614 for October. So the variation in number of students absent is greater in October.

3.15 The inflation rates in New Zealand (Figure B) have a larger standard deviation than the inflation rates in Australia, because the data from Australia show that a lot of rates are near the center of the graph, and the rates in New Zealand show rates away from the center.

3.17 **a.** Top: $3462 + 500 = 3962$
Bottom: $3462 - 500 = 2962$
b. Yes, a birth weight of 2800 grams is more than one SD below the mean because it is less than 2962.

3.19 **a.** The standard deviation is 3.05 years. **b.** The same. The standard deviation in 15 years is still 1.36. Adding 15 to each number does not affect the standard deviation. Standard deviation does not depend on the size of numbers, only on how far apart they are. **c.** The mean is 5.4 years. **d.** Larger: The mean is 20.40 years of age. When 15 is added to each number, the mean increases by 15.

3.21 The standard deviation for the players in a T-20 would be less. All the players have to score quickly in a T-20 match. In a Test match, the players are at leisure to score and the standard deviation can be quite widely spread out.

3.23 **a.** Neither data set shows strong skew.

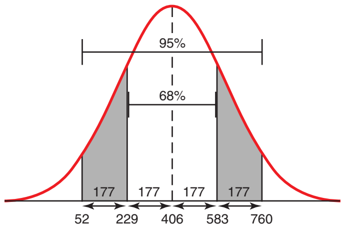
```
Female
7   99
8   01233355667789
9   3559

Male
8   899
9   012344555667
10  00468
```

b. The mean for the females is 8.625, and the mean for the males is 9.56, which shows that the typical brain size is larger for the males. **c.** The standard deviation for the females is 0.560, and for the males it is 0.561, showing that the variation in brain size is about the same for men and women.

3.25 **a.** The mean for the urban areas was 20.25, and for the rural areas was 28.58, showing that the children from rural areas are typically more malnourished than those from urban areas. **b.** The standard deviation for the urban areas was 10.28, and for the rural areas was 14.03, showing more variation in the number of malnourished children from rural areas. **c.** The mean for the urban areas is now 20.70 and the mean for the rural areas is now 28.58. Thus, the mean for the rural areas is still larger than the urban areas, but not as much larger. **d.** The standard deviation would be smaller without the two outliers. This is because the contribution to the standard deviation from these two outliers is large since they are farthest from the mean, and that contribution would be removed.

Section 3.2

3.27 Answers correspond to the guided steps.
1: 95% (See the accompanying curve.)
2: 583 is from $406 + 177$, because it is one standard deviation above the mean.
3: As shown on the curve, A is 52, B is 229, and C is 760.
4: Answer a. About 95% between 52 and 760.
5: Answer b: About 68% between 229 and 583.
6: Answer c: Most would not consider 584 unusual because it is between 52 and 760.
7: Answer d: 30 is unusually small because it is less than 52, which means it is more than two standard deviations below the mean, so less than 2.5% of the population have values lower than this.



3.29 $z = \dfrac{1010 - 978}{32} = \dfrac{32}{32} = 1$

a. About 95% according to the Empirical Rule, because the z-scores are -2 and 2. **b.** About 68% according to the Empirical Rule.

c. Nearly all the data should be within three standard deviations of mean. Three standard deviations above the mean is

$$978 + 3(32) = 1074$$

Because 2031 is much greater than 1074, you might think it unlikely that any of the state's rates would be 2031.

3.31 a. -2 **b.** 67 inches (or 5 feet 7 inches)

3.33 The strike rate of 192 has a z-score of 1 and the strike rate of 156 has a z-score of -2, so the strike rate of 156 is more unusual because the z-score is farther from 0.

3.35 a. $z = \dfrac{2500 - 3462}{500} = \dfrac{-962}{500} = -1.92$

b. $z = \dfrac{2500 - 2622}{500} = \dfrac{-122}{500} = -0.24$

c. A birth weight of 2500 grams is more common (the z-score is closer to 0) for babies born one month early. In other words, there is a higher percentages of babies with low birth weight among those born one month early. This makes sense because babies gain weight during gestation, and babies born one month early have had less time to gain weight.

3.37 a. $x = \bar{x} + zs = 7 + 1(1.5) = 8.5$
b. $x = \bar{x} + zs = 7 - 1.50(1.5) = 4.75$

Section 3.3

3.39 Two measures of the center of data are the mean and the median. The median is preferred for data that are strongly skewed or have outliers. If the data are relatively symmetric, the mean is preferred but the median is also OK.

3.41

224	237	244	246	255		256	261	293	340	415
		Q1			Med			Q3		

a. The median of 255.5 million is the typical income for the top ten grossing Pixar animated movies. **b.** Interquartile range $= 293 - 244 = 49$ million. This is the range of the middle 50% of the sorted incomes in the top ten grossing Pixar animated movies.

3.43

246	255	256	261	293	340	415
	Q1		Med		Q3	

The median of the top seven is 261 million.
Interquartile range $=$ Q3 $-$ Q1 $= 340 - 255 = 85$ million

3.45 a. The median for the men was 73 and the median for the women was 81, showing that the women were typically a bit happier than the men. **b.** The interquartile range for the men was about 51.5 and the interquartile range for the women was 44, showing more variation in happiness level for the men. Remember that IQRs vary with different technology.

Section 3.4

3.47 a. Outliers are observed values that are far from the main group of data. In a histogram they are separated from the others by space. If they are mistakes, they should be removed. If they are not mistakes, do the analysis twice: once with and once without outliers. **b.** The median is more resistant, which implies that it changes less than the mean (when the data with and without outliers are compared).

3.49 The corrected value will give a different mean as well as a different median. Removing a 0 will reorder the scores by changing the value of the median entry, that is, the third entry.

3.51 a. You could use the mean and standard deviation for the men because the data set is roughly symmetric (although the median and interquartile range are also appropriate). **b.** You should use the median and interquartile range for the women because of the large outlier. The median and IQR are not affected by the outlier. **c.** When comparing the men and women, you should use the median and interquartile range for both to make a fair comparison. This will reduce the effect of the outlier on the comparison. **d.** In the men's data, the mean and median would be close because the distribution is relatively symmetric. **e.** In the women's data, the mean and median would be farther apart. Also, the mean would be larger than the median because the large outlier pulls the mean toward it.

3.53 a. The data sets are right-skewed, as shown by the potential outliers and longer right whiskers, so the medians and interquartile ranges are appropriate.

Medians: P between 15,000 and 20,000, AD about 40,000, S about 25,000, and A almost 20,000. Interquartile range (smallest to largest): S, P, A, AD

Section 3.5

3.55 The data sets are right-skewed, as shown by the potential outliers and longer right whiskers, so the medians and interquartile ranges are appropriate.

Medians: P between 15,000 and 20,000, AD about 40,000, S about 25,000, and A almost 20,000. Interquartile range (smallest to largest): S, P, A, AD

3.57 Answers will vary: Mawsynram tends to be wettest since median rainfall is high, although Cherrapunji and Tutendo have approximately the same medians. The most varied rainfall levels are in Sant Antonio de Ureca. Both Cherrapunji and Tutendo have little variation, as determined by interquartile range.

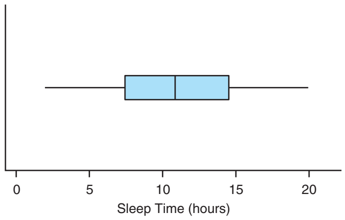
The choice of place will vary.

3.59 a. Histogram 1 is left-skewed, histogram 2 is roughly bell shaped and symmetric (not very skewed), and histogram 3 is right-skewed.

b. Histogram 1 goes with Boxplot C.
Histogram 2 goes with Boxplot B.
Histogram 3 goes with Boxplot A.

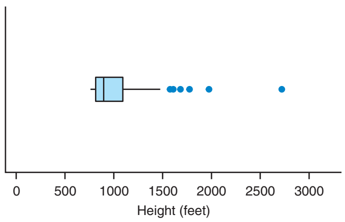
A long left tail on a histogram goes with observations going down on a boxplot, because smaller numbers are to the left in a histogram and on the bottom in boxplots like these.

3.61 The median of 10.8 hours is more than the 8 hours for people.



3.63 $1.5 \times \text{IQR} = 1.5(1093 - 810) = 1.5 \times 283 = 424.5$
Thus any building more than 424.5 feet below Q1 is a potential outlier: $810 - 424.5 = 385.5$. There are no potential outliers on the left side of the box, so the whisker will stop at the minimum of 745.

Potential outliers on the right are more than 424.5 feet above Q3, or $1093 + 424.5 = 1517.5$ feet. The right-side whisker will stop at the tallest building that is less than 1517.5 feet, and (from the dotplot) that appears to be a bit less than 1500 feet. From the boxplot there appear to be six outliers on the right side. However, (from the dotplot) two are the same, so there are really seven outliers. The distribution is skewed right.



3.65 The IQR is $90 - 78 = 12$ and $1.5 \times 12 = 18$, so any score below $78 - 18 = 60$ is a potential outlier. We can see that there is at least one potential outlier (the minimum score of 40), but we don't know how many other potential outliers there are between 40 and 60. Therefore, we don't know which point to draw the left-side whisker to.

Chapter Review Exercises

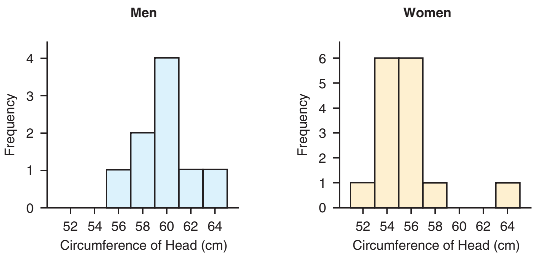
3.67 **a.** The median number of death row prisoners in the states in the South is 60 **b.** IQR = Q3 − Q1 = 161 − 34 = 127

The range of the middle 50% of the state data on death row prisoners in the South is 127.

c. The mean number of death row prisoners in the states in the South is 106.2 **d.** The mean is pulled up by the really large numbers, such as the numbers from Texas and Florida. The median is not affected by the size of these large numbers. **e.** The median is more stable with outliers.

3.69 The answers given follow the steps in the Guided Exercises.

1:



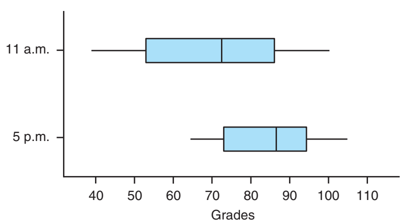
2: The men's histogram is roughly symmetric (bell-shaped). The women's histogram is bell-shaped except for the high-end outlier, so it may be called right-skewed.
3: Compare the medians and interquartile ranges because of the outlier for the women.
4: median, 59.5, median, 55, men
5: Men's IQR = 3.75. Women's IQR = 2.5. Men, IQR may vary with different technology. These values were obtained from Minitab.
6: The measurement of 63 for the women was an outlier.
7: Both data sets are unimodal and roughly symmetric except for one large outlier for the women. The men tended to have larger heads with more variation in size than the women. However, there was a large outlier (63 cm) for the women.

3.71 Summary statistics are shown below. The 5 p.m. class did better, typically; both the mean and the median are higher**.** Also, the spread (as reflected in both the standard deviation and the IQR) is larger for the 11 a.m. class, so the 5 p.m. class has less variation.

The visual comparison is shown by the boxplots. Both distributions are slightly left-skewed. Therefore, you can compare the means and standard deviations *or* the medians and IQRs.

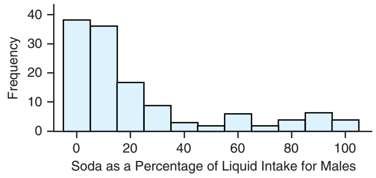
Minitab Statistics

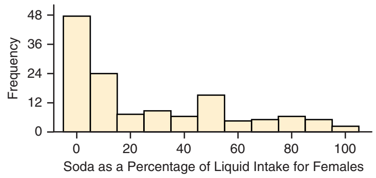
Variable	N	Mean	Median	StDev	Min	Max	Q1	Q3
11am	15	70.73	72.5	19.84	39	100	53	86
5pm	19	84.78	86.5	11.95	64.5	104.5	73	94



3.73 The graph is bimodal, with modes around 65 inches (5 feet 5 inches) and around 69 inches (5 feet 9 inches). There are two modes because men tend to be taller than women.

3.75





Both histograms are strongly right-skewed. The median for the males is 5% which is larger than the median for the females which is 3%. The interquartile range for the females is 40 percentage points, which is much larger than the interquartile range for the males, which is 20 percentage points, showing that there is more variation among these females than among these males.

3.77 **a.** The mean is about 68 inches. **b.** $(78 - 58)/6 = 3.3$, so the standard deviation should be roughly 3 inches.

3.79 and 3.81 Constructed numbers will vary.

3.83 Answers will vary but should be supported with appropriate graphs and suitable variations.

3.85 **a.** 11.5, 9.5, longboards **b.** 8, 5, longboards (Answers for interquartile range may vary with different technology.)

3.87 **a.** Both data sets are right-skewed and have outliers that represent large numbers of hours of study, so the medians (and interquartile ranges) should be compared. **b.** The median of 7 was larger for the women; the men's median was only 4. The interquartile range was 4.5 for the women and 3 for the men, so the IQR was larger for the women. Both data sets are right-skewed with outliers at around 15 or 20 hours. Summary: The women tended to study more and had more variation as measured by the interquartile range.

3.89 $70 + 1.5(10) = 85$

3.91 The z-score for the qualified person's salary of $80,000 is 1.5, and the z-score for the semi-qualified person's salary of $36,000 is 2. The salary of $36,000 for semi-qualified is more unusual, because its z-score is farther from 0.

3.93 **a.** The shape is right-skewed, the median is 28.50, and the interquartile range is $31.75 - 24.75 = 7$. There is an outlier at 55. **b.** The mean is 29.64 and the median is 28.50. The mean is larger because of the outlier, 55. The mean and median should be marked on the histogram of the data (not shown).

CHAPTER 4

Section 4.1

4.1 **a.** Acres: The number of acres has a stronger relationship with the value of the land, as shown by the fact that the points are less scattered in a vertical direction. **b.** Acreage. The value of land is more strongly associated with the acreage than with the number of rooms because the vertical spread is less.

4.3 Very little trend.

4.5 The more people weigh, the more weight they tend to want to lose.

4.7 The trend is positive. Students with more sisters tended to have more brothers. This trend makes sense, because large families are likely to have a large number of sons and a large number of daughters.

4.9 There is a slightly negative trend. The negative trend suggests that the more hours of work a student has, the fewer hours of TV the student tends to watch. The person who works 70 hours appears to be an outlier, because that point is separated from the other points by a large amount.

4.11 There is a slight negative trend that suggests that older adults tend to sleep a bit less than younger adults. Some may say there is no trend.

Section 4.2

4.13 a. You may find the correlation because the trend is linear. **b.** You should not find the correlation because the trend is not linear.

4.15 There is no correlation between the mean age of working mothers when they give birth for the first time and the mean number of children.

4.17 0.767 A
0.299 B
−0.980 C

4.19 0.87 A
−0.47 B
0.67 C

4.21 a. $r = 0.950$
b. Same
c. Same

4.23 The correlation would not change. The correlation does not depend on which variable is the predictor and which is the response.

4.25 The correlation is −0.272. The hotels that tend to have the best reservation process tend not to help the traveler for booking.

4.27 The correlation is 0.704. The positive sign suggests that the more time spent on video games, the higher the BMI tends to be.

Section 4.3

4.29 a. The independent variable is the volume of used cars sold, and the dependent variable is the value of used cars sold. **b.** Car value distributions are usually skewed. Medians are therefore a more meaningful measure of the center. **c.** Around £22 billion **d.** £28.18 billion **e.** Answers will vary. The condition of car, accident history, service and maintenance, insurance record, and distance traveled are some of the factors that might influence the value of used cars.

4.31 a. Between 8 and 8.5. **b.** 8.5

4.33 a. Predicted Armspan = 16.8 + 2.25 Height
b. $b = 0.948(8.10/3.41) = 2.25$ **c.** $a = 159.86 - 2.25(63.59) = 16.8$
d. Armspan = 16.8 + 2.25(64) = 160.8, or about 161 cm

4.35 a. Predicted Armspan = 6.24 + 2.515 Height (Rounding may vary.)
b. Minitab: slope = 2.51, intercept = 6.2
StatCrunch: slope = 2.514674, intercept = 6.2408333
Excel: slope = 2.514674, intercept = 6.240833
TI-84: slope = 2.514673913, intercept = 6.240833333

4.37 The association for the teachers is stronger because correlation (r-value) is further from 0.

4.39 a. The slope would be near 0. **b.** r is 0 **c.** Monthly shopping is not associated with weight.

4.41 Explanations will vary.

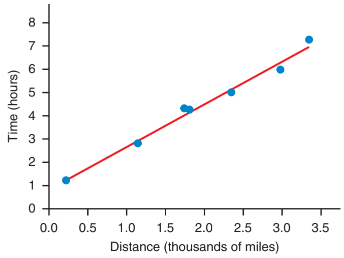
	x	y
a.	gas	miles
b.	years	salary
c.	weight	belt size

4.43 a. The higher the percentage of smoke-free homes in a state, the lower the percentage of high school students who smoke tends to be.
b. 56.32 − 0.464(70) = 23.84, or about 24%

4.45 a. The graph shows that the fastest athletes are between 30 and 40 years of age and that the young and old athletes have slower speed. **b.** Since the trend is not linear, it would not be appropriate for linear regression.

4.47 The answers follow the guided steps.

1: Graph.



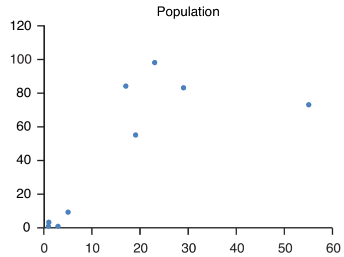
2: The linear model is appropriate. The points suggest a straight line.

3: In the formula, the time is in hours and the distance in thousands of miles.

Predicted Time = 0.8394 + 1.838 Distance

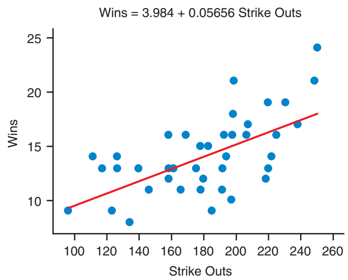
4: Each additional thousand miles takes, on average, about 1.84 more hours to arrive.
5: The additional time shown by the intercept might be due to the time it takes for the plane to taxi to the runway, delays, the slower initial speed, and similar delays in the landing as well. The time for this appears to be about 0.84 hours (or 50 minutes).
6: Predicted Time = 0.8394 + 1.838(3) = 0.8394 + 5.514 = 6.35 hours. The predicted time from Boston to Seattle is 6.35 hours.

4.49 a. Positive. The greater the population, the greater is the number of billionaires. **b.** See scatterplot **c.** $r = 0.763$ **d.** For each additional hundred thousand in the population, there is an additional 0.3205 billionaires.



e. It does not make sense to look for billionaires in countries with no population.

4.51 a. The slope is positive. **b.** Predicted Wins = 3.98 + 0.0566 Strike-outs **c.** For each additional strike-out, there is an average of 0.057 more wins. Or, for each 100 more strike-outs, there are an average of 5.7 more wins. **d.** With 0 strike-outs, there should be about 4 wins, but there are no data near this region, so this extrapolation may mislead.



Section 4.4

4.53 a. An influential point is a point that changes the regression equation by a large amount. **b.** Going to church may not cause lower blood pressure; the mere fact that two variables are related does not show that one caused the other. It could be that healthy people are more likely to go to church, or there could be other confounding factors.

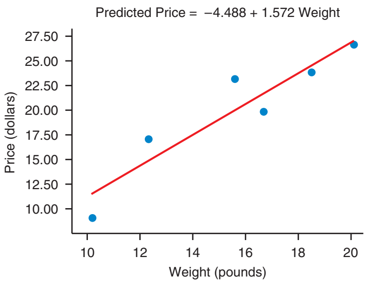
4.55 Older children have larger shoes and have studied math longer. Large shoes do not cause more years of studying. Both are affected by age.

4.57 Square 0.67 and you get 0.4489, so the coefficient of determination is about 45%. Therefore, 45% of the variation in weight can be explained by the regression line.

4.59 Apart from the fact that increasing social awareness will make people support causes like this, regression toward the mean predicts that these raises should tend to improve and move closer to the mean in the future.

4.61 **a.** The salary is \$2.099 thousand less for each year later that the person was hired, or \$2.099 thousand more for each year earlier. **b.** The intercept (\$4,255,000) would be the salary for a person who started in the year 0, which is ridiculous.

4.63 **a.** See graph.



b. $r = 0.933$
A positive correlation suggests that larger turkeys tend to have a higher prices. **c.** Predicted Price = −4.49 + 1.57 Weight **d.** The slope: For each additional pound, the price goes up by \$1.57. The interpretation of the intercept is inappropriate, because it is not possible to have a turkey that weighs 0 pounds. **e.** The 30-pound free turkey changes the correlation to −0.375 and changes the equation to

Predicted Cost = 26.87 − 0.553 Weight

This implies that the bigger the turkey, the less it costs! The 30-pound free turkey was an influential point, which really changed the results.

4.65 **a.** Positive correlation **b.** For each additional dollar spent on teachers' pay, the expenditure per pupil goes up about 23 cents on average. **c.** The intercept might represent the mean cost of education excluding the cost of paying teachers. However, doing so requires extrapolation, which may not be valid.

4.67 **a.** Correlation is negative because the trend is downhill, or because the slope's sign is negative. **b.** For each additional hour of work, the score tended to go down by 0.48 points. **c.** A student who did not work would expect to get about 87 on average.

4.69 Linear regression is not appropriate, because the points do not follow a straight line and tend to follow a curve.

4.71 **a.** Predicted Mother's Education = 3.12 + 0.637(12) = 312 + 7.64 = 10.76 (12, 10.76)

Predicted Mother's Education = 3.12 + 0.637 (4.0) = 3.12 + 2.548 = 5.67 (4, 5.67)

b. The new equation is given below.
Predicted Father's Education = 2.78 + 0.637 (Mother's Education)
Predicted Father's Education = 2.78 + 0.637 (12) = 10.42 (12, 10.42)
Predicted Father's Education = 2.78 + 0.637 (4) = 5.33 (4, 5.33)
c. Regression toward the mean: Values for the predictor variable that are far from the mean lead to responses that are closer to the mean.

4.73 The answer follows the guided steps.

1: **a.** Slope: $b = r\dfrac{s_y}{s_x} = 0.7\dfrac{10}{10} = 0.7$

b. Intercept: $a = \bar{y} - b\bar{x}$
$a = 75 - 0.7(75) = 22.5$

c. Equation

Predicted Final = 22.5 + 0.7 Midterm

2: Predicted Final = 22.5 + 0.7 Midterm
= 22.5 + 0.7(95)
= 89

3: The score of 89 is lower than 95 because of regression toward the mean.

Chapter Review Exercises

4.75 **a.** $r = 0.941$

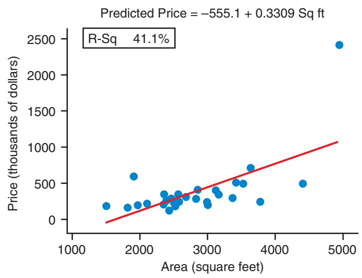
Predicted Weight = −245 + 5.80 Height

b.

Height	Weight
60(2.54) = 152.4	105/2.205 = 47.6190 kg
66(2.54) = 167.64	140/2.205 = 63.4921 kg
72(2.54) = 182.88	185/2.205 = 83.9002 kg
70(2.54) = 177.8	145/2.205 = 65.7596 kg
63(2.54) = 160.02	120/2.205 = 54.4218 kg

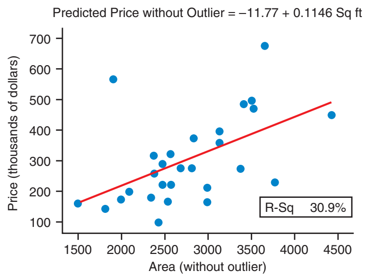
c. The correlation between height and weight is 0.941. It does not matter whether you use inches and pounds or centimeters and kilograms. A change of units does not affect the correlation because it has no units.
d. The equations are different.
Predicted Weight Pounds = −245 + 5.80 Height (in inches)
Predicted Weight Kilograms = −111 + 1.03 Height (in centimeters)

4.77 **a.**



The slope is 0.33: For each additional square foot, the average price is 0.33 thousand dollars (\$330) more. The intercept (which is negative) is not appropriate to interpret, because the concept of a house with 0 square feet does not make sense. The coefficient of determination is 41%, showing that 41% of the variation in price can be explained by the regression.

b.



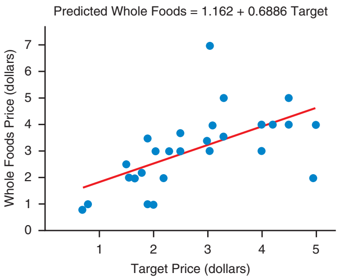
The slope is 0.11: For each additional square foot, the average price is 0.11 thousand dollars (\$110) more. The intercept (which is negative) is not appropriate to interpret, because houses of 0 square feet cannot exist. The coefficient of determination is 31%, showing that 31% of the variation in price can be explained by the regression.

c. The outlier is the house costing \$2.4 million with square footage of about 4900 square feet. The outlier pulls the line up on the far right, increasing the slope and increasing the coefficient of determination.

4.79 **a.** There were no women taller than 69 inches, so the line should stop at 69 inches to avoid extrapolating. **b.** Men who are the same height as women wear shoes that are, on average, larger sizes. **c.** The mean increase in shoe size based on height is the same for men and women.

4.81 Among those who exercise, the effect of age on weight is less. An additional year of age does not lead to as great an increase in the average weight for exercisers as it does for non-exercisers.

4.83 a. See graph. There is a slight curvature to the trend; in general, items that cost more at Target tend to cost more at Whole Foods, at least for items below 3 dollars at Target. However, more expensive items at Target tend to be less expensive at Whole Foods, relative to other Whole Food prices.
b. Since the trend is not linear, do not go on to do the analysis.

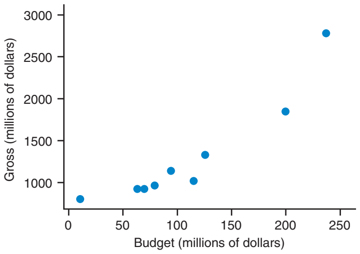


4.85 Correlation table:

	Diameter	Height
Height	0.519	
Volume	0.967	0.598

The diameter is a better predictor of volume than the height, because there is a larger correlation between diameter and volume than between height and volume.

4.87 It is not appropriate to fit a linear regression model, because the trend is not linear. However, we can see that big-budget films tend to gross more. One point in the upper right corner jumps out at you. *Avatar* had the largest budget (of these) and also the largest gross. *Titanic* is the other outlier.



4.89 a. The positive trend shows that the more stories there are, the taller the building tends to be.
b. 115.4 + 12.85(100) = 115.4 + 1285 = 1400.4 feet, or about 1400 feet
c. Buildings that have one additional story tend to have an average height of 12.85 additional feet.
d. Because there are no building with 0 stories, the interpretation of the intercept is not appropriate.
e. About 71% of the variation in height can be explained by the regression, and about 29% is not explained.

4.91 and 4.93 will vary.

4.95 The trend is positive. In general, if one twin has a higher-than-average level of education, so does the other twin. The point that shows one twin with 1 year of education and the other twin with 12 years is an outlier. (Another point showing one twin with 15 years and the other with 8 years is a bit unusual, as well.)

4.97 There appears to be a positive trend. It appears that the number of hours of homework tends to increase slightly with enrollment in more units.

4.99 Linear regression is not appropriate because the trend is not linear, it is curved.

4.101 The cholesterol going down might be partly caused by regression toward the mean.

CHAPTER 5

Section 5.1

5.1 a. 2 6 4 2 7   4 0 6 5 0
b. T T T T H   T T T H T
c. No. We got 2 heads.

5.3 Empirical probability, because it is based on an experiment.

5.5 Empirical probability, because it is based on an experiment.

Section 5.2

5.7 a. The nine equally likely outcomes are Arnold, Rai, Brown, Callaway, Cooper, Gillibrand, Morgan, and Parsons. **b.** The outcomes that make up event A are Arnold, Rai, Brown, Cooper, Gillibrand, and Parsons. **c.** 6/9, or about 66.7%. **d.** Callaway, Goozee, and Morgan.

5.9 The number −0.85 could not be a probability because it is negative, 8.50 could not be a probability because it is greater than 1, and 850% could not be a probability because it is more than 100%.

5.11 a. A heart: 13/52 or 1/4 **b.** A red card: 26/52 or 1/2 **c.** An ace: 4/52 or 1/13 **d.** A face card: 12/52 or 3/13 **e.** A three: 4/52 or 1/13 Answers may also be in decimal or percentage form.

5.13 a. P(guessing correctly) = 1/2 **b.** P(guessing incorrectly) = 1/2

5.15

	Number of Girls	Probability
a.	0	1/16
b.	1	4/16 = 1/4
c.	2	6/16 = 3/8
d.	3	4/16 = 1/4
e.	4	1/16

5.17 The probability of celebrating an anniversary on a weekday is 5/7, or about 71.43%.

5.19 a. 553/1275, or about 43.4%
b. 978/1275, or about 76.7%

5.21 577/1275, or about 45.3%

5.23 a. (978 + 101)/1275 = 1079/1275, or about 84.6%
b. Saying Yes is not the complement of saying No because there were some people who said they were Unsure. You can tell that from the fact that the probabilities do not add to 100%.

5.25 The answer follows the guided steps.
1: 553/1275.
2: 978/1275.
3: No, they are not mutually exclusive because there are males who said Yes.
4: 401/1275.
5: If you don't subtract that probability, you will have counted the number of males who said Yes twice.
6: 553/1275 + 978/1275 − 401/1275 = 1130/1275, or about 88.6%.
7: The probability that a person is male OR said Yes is about 88.6%.

5.27 Answers will vary. Being a man and being a woman are mutually exclusive. Saying Yes and saying No are mutually exclusive. Naming any two rows (or two columns) gives you mutually exclusive events.

5.29 a. Not mutually exclusive **b.** Mutually exclusive

5.31 No, the two probabilities cannot be simply added. You need to know the percentage of farmers who use chemical fertilizers AND practice crop rotation and subtract it from the sum.

5.33 a. 4/6, or 66.7% **b.** 3/6 = 1/2, or 50%

5.35 a. A OR B: 0.18 + 0.25 = 0.43 **b.** A OR B OR C: 0.18 + 0.25 + 0.37 = 0.80 **c.** Lower than a C: 1 − 0.80 = 0.20

5.37 8% held a learner's license. The sum must be 100%.

5.39 a. Most: Category 4: either bat OR bowl **b.** Fewest: Category 1: both bat AND bowl

5.41 Students who can sing OR dance is the larger group.

5.43 a. $0.6(0.6) = 0.36$ **b.** $0.6(0.4) + 0.4(0.6) = 0.24 + 0.24 = 0.48$ **c.** $0.36 + 0.48 = 0.84$ **d.** $1 - 0.36 = 0.64$

5.45 a. More than 12 correct: $1 - 0.42 - 0.38 = 0.20$ **b.** At most 12 correct: $0.42 + 0.38 = 0.80$ **c.** 5 or more correct: $0.38 + 0.20 = 0.58$ **d.** The events in parts a and b are complementary because "at most 12 correct" means from 0 correct up to 12 correct. "More than 12 correct" means 13 to 20 correct. Together, these mutually exclusive events include the entire sample space.

Section 5.3

5.47 P(Yes | Male)

5.49 a. 577/722, or 79.9% **b.** 401/553, or 72.5% **c.** The females were slightly more likely to say Yes.

5.51 The events are associated. Only persons holding a valid commercial driving license can drive a cab in most parts of the world. Put another way, a person driving a cab is likely to hold a valid commercial driving license.

5.53 Wearing spectacles and gender are independent because wearing spectacles does not depend on gender.

5.55 They are not independent because the probability of saying Yes if the person is female is 577/722, or 79.9%, whereas the overall probability of saying Yes is 978/1275, or 76.7%, and these are not exactly the same.

5.57 The answers follow the guided steps.

	M	W	
Right	18	42	60
Left	12	28	40
	30	70	100

1: See table.

2: 60/100, or 60%

3: $18/30 = 60\%$

4: The variables are independent because the probability of having the right thumb on top given that a person is a man is equal to the probability that a person has the right thumb on top (for the whole data set).

5.59

	Agree	**Don't Know**	**Disagree**	**Total**
Happy	242	65	684	991
Unhappy	45	30	80	155
Total	287	95	764	1146

$$P(\text{Happy}/\text{Agree}) = 242/287 = 0.843$$
$$P(\text{Happy}) = 991/1146 = 0.895 = 0.865$$

These percentages are very close, so you might think the factors are independent. But we said for this chapter that independence requires precisely equal probabilities, and so they are *not independent* by that standard.

5.61 a. 1/8 **b.** 1/8

5.63 They are the same. Both probabilities are $\left(\frac{1}{6}\right)^5$, or $\frac{1}{7776}$

5.65 a. $(0.20)(0.20)(0.20) = 0.008$, or about 1%. We must assume that the prisoners are independent with regard to recidivism.
b. $0.80(0.80)(0.80) = 0.512$, or about 51.20%.
c. About 48.8% (from $100\% - 51.2\%$)

Section 5.4

5.67 a. HTHTH HHTTT THHTH HHHTT
b. 11/20

5.69 Histogram B was for 10,000 rolls because it has nearly a flat top. In theory, there should be the same number of each outcome, and the Law of Large Numbers says that the one with the largest sample should be closest to the theory.

5.71 The proportion should get closer to 0.5 as the number of flips increases.

5.73 Alfred and David are betting more times (100 times), so they are more likely to end up with each having about half of the wins, compared to Laura and June. The Law of Large Numbers says that the more times an experiment is repeated, the closer the experimental proportion comes to the theoretical proportion (50%). The graph shows that the proportion of wins settles down to about 50% by 100 trials. But at 20 trials, the percentage of wins has not settled down and will vary quite a bit.

5.75 You are equally likely to get a red or a black card (assuming the deck is fair) because the deck's results are independent of each other—that is, the deck does not "keep track" of its past.

5.77 The probability of selecting a digit from 0 to 25 is 26/51, or 51%. So, it does not represent the probability we wish to simulate.

5.79 a. You could use the numbers 1, 2, 3, and 4 to represent the outcomes and ignore 0 and 5–9, but answers to this will vary. **b.** The empirical probabilities will vary. The theoretical probability of getting a 1 is 1/4; remember that the die is four-sided.

Chapter Review Exercises

5.81 228/380, or 60%

5.83 a. Number of visitors visiting an art exhibition and the nationality of artists are independent, because it is the name and works of the artist that draw in more visitors. **b.** Breed of horse and weight are associated because certain breeds (such as Shire) have more weight and other breeds (such as Appaloosa) tend to weigh less.

5.85 a. $(0.63)(0.37) = 0.2331$, or about 23%
b. $(0.37)(0.63) = 0.2331$, or about 23%
c. 53% (from $1.00 - 0.2331 - 0.2331 = 0.5338$)
d. 77% (from $1.00 - 0.2331 = 0.7669$)

5.87 a. $0.93(0.93) = 0.8649$ **b.** The married couple might have Internet access at home, in which case, if one of them has Internet access, then the other also does (unless one of them prohibits the other's use).

5.89 a. Both married in August $= \left(\frac{1}{12}\right)\left(\frac{1}{12}\right) = \frac{1}{144}$

b. Glen OR Shahid were married in August:

$$P(\text{A OR B}) = P(\text{A}) + P(\text{B}) - P(\text{A AND B})$$
$$= \frac{1}{12} + \frac{1}{12} - \frac{1}{12}\left(\frac{1}{12}\right)$$
$$= \frac{1}{12} + \frac{1}{12} - \frac{1}{144}$$
$$= \frac{23}{144}$$

5.91 a. $0.27(1500) = 405$ **b.** 165 **c.** 855

5.93 a.

Age	Consumed Coffee	Did Not Consume Coffee	Total
9–18	49	251	300
19–30	165	335	500
31–50	63	37	100
51+	134	66	200
Total	411	689	1100

b. The older the people are, the more likely they consume coffee and coffee substitutes. This makes sense because coffee contains riboflavin and magnesium, which can lower the risk of diabetes and heart disease and protect drinkers from Alzheimer's disease and Parkinson's disease.
c. 165/1100, or about 15%
d. $165/411 = 40.5\%$
e. 165/500, or 33% (which was given)

f. In part c, the divisor is the whole group (1100). That large divisor results in a smaller percentage.

5.95 **a.** $(0.28)(0.28) = 0.0784$, or about 8% **b.** $(0.72)(0.72) = 0.5184$ **c.** $1 - 0.5184 = 0.4816$, or about 48%

5.97 The recidivism rates are not the same for men and women. So, recidivism and gender are not independent.

5.99 Answers will vary.

5.101 The smaller hospital will have more than 60% girls born more often, because, according to the Law of Large Numbers, there's more variability in proportions for small sample sizes. For the larger sample size ($n = 45$), the proportion will be more "settled" and will vary less from day to day. Over half of the subjects in Tversky and Kahneman's study said that "both hospitals will be the same." But *you* didn't, did you?

5.103 **a.** The probability that the student will correctly guess an answer is 0.20 (1 out of 5). The probability of randomly selecting a 0 or a 1 is $2/10 = 1/5 = 0.20$.

b.

1	1	3	7	3	9	6	8	7	1
R	R	W	W	W	W	W	W	W	R

c. Yes. There were 3 correct.

d. WWRWW WWRWW. No. The student scored only 2 correct.

e. RWWRW WWWWR. 3 correct. Yes.
WRWWW WWWWW 1 correct. No.

f. There were four trials, and two had a successful outcome. Thus the empirical probability is 2/4, or 0.50.

5.105 **a.** The action is to arrive at a light, which is either green or not. The probability of a success is 60%. **b.** Answers will vary. Our method: Let the digits from 0 to 5 represent a green light, and let the digits from 6 to 9 represent yellow or red. (Any assignment that gives six digits to green and four digits to non-green will work.) **c.** The event of interest is "get three out of three green." **d.** A single trial consists of reading off three digits. **e.** Outcomes (non-green are labeled R). Three greens in a row are underlined.

2 7 5	8 3 0	1 8 6	6 5 8	2 5 0	3 8 1	0 3 3	5 8 2	5 9 4	5 1 3
G R G	R G G	G R R	R G R	G G G	G R G	G G G	G R G	G R G	G G G

6 0 8	0 1 0	4 4 3	9 5 8	6 2 1	0 9 8	4 0 3	5 1 1	9 6 0	3 7 2
R G R	G G G	G G G	R G R	R G G	G R R	G G G	G G G	R R G	G R G

Number of greens (with successful events in red)

2 2 1 1 3 2 3 2 2 3

1 3 3 1 2 1 3 3 1 2

f. P(all three green) is estimated as 7/20.

5.107 **a.** $530/1858 = 28.5\%$ **b.** $689/1858 = 37.1\%$

5.109 $104/1858 = 5.6\%$

5.111 $(689 + 469)/1858 = 62.3\%$

5.113 $\frac{530}{1858} + \frac{689}{1858} - \frac{306}{1858} = \frac{913}{1858} = 49.1\%$
The probability of being liberal OR a Democrat is 49.1%.

5.115 Any two column headings or any two row headings are mutually exclusive. For example, you cannot be both a Republican AND a Democrat, so those are mutually exclusive. Likewise, you cannot be both liberal AND moderate, so those are mutually exclusive.

5.117 **ii.** $D|C$

5.119 **a.** $306/530 = 57.7\%$ **b.** $104/593 = 17.5\%$ **c.** Liberals

5.121 Two coin flips
a. 0 heads: 1/4 (Decimals and percentages are also acceptable.) **b.** 1 head: 1/2 (from 2/4) **c.** 2 heads: 1/4 **d.** At least 1 head: 3/4 **e.** Not more than 2 heads (which means 2 or fewer heads): 1 (from 4/4)

5.123 **a.** Mutually exclusive **b.** Not mutually exclusive

5.125 You don't know what percentage of urban households have at least one four-wheeler AND at least one two-wheeler. You cannot simply add the percentages, because the events are not mutually exclusive and you would count households with at least one four-wheeler AND at least one two-wheeler twice.

5.127 **a.** Both believe: $(0.62)(0.50) = 0.31$ **b.** Neither believes: $(0.38)(0.50) = 0.19$ **c.** Same beliefs: $0.31 + 0.19 = 0.50$ **d.** Different beliefs: $1 - 0.50 = 0.50$

5.129 Answers will vary. Red die is 1, blue die is 1.

5.131 $0.40(10000) = 4000$

CHAPTER 6

Answers may vary slightly due to rounding or type of technology used.

Section 6.1

6.1 **a.** Continuous **b.** Discrete

6.3 **a.** Continuous **b.** Continuous

6.5

Number of Spots	1	2	3	4	5	6
Probability	0.1	0.2	0.2	0.2	0.2	0.1

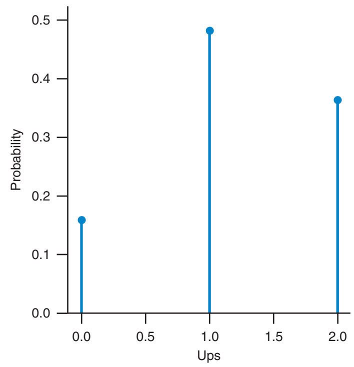
The table may have a different orientation.

6.7

UU	UD	DU	DD
0.6(0.6)	0.6(0.4)	0.4(0.6)	0.4(0.4)
0.36	0.24	0.24	0.16

6.9

0	1	2
0.16	0.48	0.36



6.11 $(6 - 3)(0.2) = 0.6$, or 60%, and the area between 3 and 6 should shaded.

Section 6.2

6.13 **a.** ii., 95% **b.** i., almost all **c.** iii., 68% **d.** iv., 50% **e.** ii., 13.5%

6.15 **a.** iii., 50% **b.** iii., 68% **c.** v., about 0% **d.** v., about 0% **e.** ii., 95% **f.** v., 2.5%

6.17 Use the output from (A), and the percentage of college women with heights of less than 63 inches is about 21%.

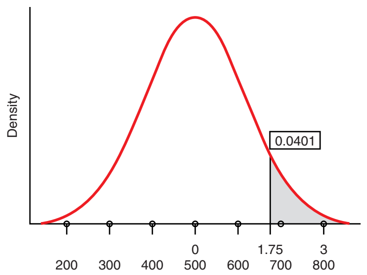
6.19 **a.** 0.8461, or about 85% **b.** $1 - 0.8461 = 0.1539$, or about 15%

6.21 **a.** 0.9608, or about 96% **b.** $1 - 0.9608 = 0.0392$, or about 4% **c.** $0.1515 - 0.0968 = 0.0547$, or about 5%

6.23 **a, b,** and **c** are all 0.000. **d.** The proportion to the right of 4.00 would be the largest of the three, and the proportion to the right of 50.00 would be the smallest. **e.** below -10.00

6.25 The answers use the steps given in the Guided Exercises.

1. $z = \dfrac{x - \mu}{\sigma} = \dfrac{675 - 500}{100} = \dfrac{175}{100} = 1.75$
2. 500 is the mean, and it belongs right below 0 because 0 is the mean of the standard Normal or the mean z-score.
3. and 4. are shown in the sketch.
5. 0.9599
6. $1 - 0.9599 = 0.0401$
7. The percentage of female college-bound seniors taking the SAT who scored 675 or more is about 4.0%.



6.27 0.8413, or about 84%

6.29 **a.** 15.9% **b.** 1.7% **c.** 90.4%

6.31 **a.** 56. The mean is right in the middle, which is the average of the boundary values. **b.** 1. According to the Empirical Rule, the middle 95% of the data fall within two standard deviations of the mean. The upper boundary is two standard deviations above the mean, and the lower boundary is two standard deviations below the mean. That spans four standard deviations. The range is 4, and if you divide by 4 you get 1. (There are other ways to do this as well.)

6.33 **a.** 0.5000, or about 50% **b.** 0.3643, or about 36.43%

6.35 **a.** 0.5538, or about 55.38% **b.** 0.4208, or about 42.08%

6.37 About 80% of the days in February have minimums of 32°F or less.

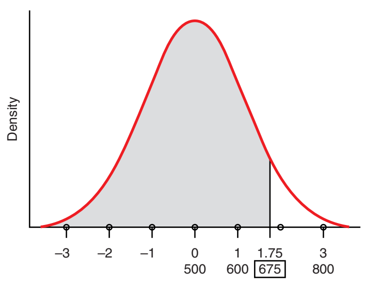
6.39 **a.** Measurement (inverse) **b.** Probability

6.41 $z = 0.52$

6.43 **a.** $z = 1.07$ **b.** $z = -2.18$

6.45 The answers follow the steps given in the Guided Exercises.

1. The test score will be above the mean, because 96% of students score worse, so it must be a very high grade.
2. See the figure.
3. $z = 1.75$
4. See the figure.
5. $500 + 1.75(100) = 500 + 175 = 675$
6. See the figure.
7. The SAT score at the 96th percentile is **675.**



6.47 69 inches (5 feet 9 inches)

6.49 62.9 inches (5 feet 2.9 inches)

6.51 **a.** 567 **b.** 433 **c.** $567 - 433 = 134$ **d.** The IQR is 134 and the standard deviation is 100, so the interquartile range is larger.

6.53 **a.** 77th percentile **b.** 66 inches (5 feet 6 inches)

6.55 **a.** 3.5 ounces **b.** 2.5 ounces

Section 6.3

6.57 Conditions

Two complementary outcomes: boy or girl

Fixed number of trials: 4 children

Same probability of success on each trial: 1/2 probability of a boy

All trials independent: Because there are no twins, the gender of each child is independent of the gender of the others.

6.59 There is not a fixed number of trials, because the members keep rolling dice *until* the 10 minutes are over.

6.61 It is not a binomial experiment, because the two people (joint account holders) are not independent with respect to closing an account. If one closes the account, then the other person's account is also closed.

6.63 **a.** $b(25, 0.603, 14)$ **b.** $b(25, 0.397, 17)$

6.65 $b(6, 0.40, 2) = 0.311$

6.67 **a.** 0.164 **b.** 0.164

6.69 **a.** 0.387 **b.** 0.264 **c.** 0.736

6.71 **a.** 0.251 **b.** It will be larger because 6 or more includes the probability for 6 and also for 7, 8, and so forth. **c.** 0.633

6.73 **a.** $0.7(15) = 10.5$; expect 10 or 11 **b.** $b(15, 0.70, 11) = 0.219$ **c.** $b(15, 0.70, 11$ or fewer$) = 0.703$

6.75 **a.** DD DN ND NN
b. DD 1/4

DN 1/4

ND 1/4

NN 1/4

c. Neither was drunk: 1/4 **d.** Exactly one was drunk (from 1/4 + 1/4): 1/2 **e.** Both were drunk: 1/4

6.77 **a.** In 60 rolls, expect the odd numbers to turn up 30 times. **b.** $\sigma = \sqrt{60(0.5)(1 - 0.5)} = \sqrt{15} = 3.87$, or about 4. **c.** You should expect the odd numbers to turn up between 26 and 34 times.

6.79 **a.** 160 **b.** $\sqrt{200(0.80)(0.20)} = 6$ **c.** 148, 172 **d.** No, 164 would not be surprising because it is within the interval given in part c.

Chapter Review Exercises

6.81 95.2%, or about 95%

6.83 **a.** 0.7625, or about 76% **b.** 98.6°F

6.85 **a.** 280 **b.** $\sqrt{400(0.7)(0.3)}$ = about 9 **c.** $400 - 280 = 120$ **d.** 9 **e.** They are the same because one is $\sqrt{400(0.7)(0.3)}$ and the other is $\sqrt{400(0.3)(0.7)}$

6.87 **a.** 3.45% **b.** $0.0345(0.0345) = 0.0012$
c. b(200, 0.0345, 7 or fewer) = 0.614. **d.** $200(0.0345)$ = about 7
e. $\sqrt{200(0.0345)(0.9655)} = 2.58$ **f.** $7 + 2(2.58)$ = about 12, and $7 - 2(2.58)$ = about 2; about 2 to 12 **g.** Yes, 45 would be surprising because it is so far from the interval.

6.89 **a.** 19.7 inches **b.** 20.5 inches **c.** They are the same. The distribution is symmetric and so the mean is right in the middle at the 50th percentile.

CHAPTER 7

Answers may vary slightly due to rounding or type of technology used.

Section 7.1

7.1 A parameter is a measure of the population, and a statistic is a measure of a sample.

7.3 **a.** $\bar{x}$ is a statistic, and μ is a parameter. **b.** $\bar{x}$

7.5 Since it is only a random sample, the mean GPA of 100 students can be used to infer the approximate GPA of the 500 students.

7.7 You want to test a sample of batteries. If you tested them all until they burned out, no usable batteries would be left.

7.9 First, one card is drawn from the deck of cards and noted.
"Without Replacement": After the first card is drawn out, it is not replaced, so the second draw must be a different card.

"With Replacement": The card that is selected is replaced in the deck, and a second card is drawn. It is possible that the same card is drawn twice.

7.11 Chosen: 7, 3, 5, 2

7.13 Assign each student a pair of digits 00–29 (or 01–30). Read off pairs of digits from the random number table. The students whose digits are called are in the sample. Skip repeats. Stop after the first 10 are selected.

7.15 The administrator might dismiss the negative findings by saying the results could be biased because the small percentage who chose to return the survey might be very different from the majority who did not return the survey.

7.17

	With Persuasion	**Without Persuasion**
Support Cap	6 + 2	13 + 5
Oppose	9 + 8	2 + 5

Support capital punishment
a. With persuasion: $8/25 = 32\%$
b. Without persuasion: $18/25 = 72\%$
c. Yes, she spoke against it, and fewer who heard her statements against it (32%) supported capital punishment, compared with those who did not hear her persuasion (72%).

Section 7.2

7.19 a. The sketch should show bullet holes consistently to the left of the target and close to each other. If the bullets go consistently to the left, then there is bias, not lack of precision. **b.** The sketch should show bullet holes that are all near the center of the target.

7.21 Yes, this would be a biased sample from just the in-store shoppers. You have eliminated the possibility of selecting customers who do only online shopping from the store and may have a different purchase volume. You have also eliminated the online purchase volume of customers who shop both online and offline.

7.23 a. 11/30, or about 36.67%, multiples of 3. **b.** 11/30 is $\hat{p}$ (p-hat), the sample proportion. **c.** The error is 11/30 minus 10/30, (or 1/30), or about 3.33%.

7.25 a. We should expect 0.35, or 35% mixed fruit juice buyers.

b. $$\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.35 \times 0.65}{20}} = \sqrt{\frac{0.2275}{200}} = \sqrt{0.0011} = 0.03, \text{ or } 3\%$$

c. We expect about 35% mixed fruit juice buyers, give or take 3%.

7.27 The largest sample is the narrowest (the graph on the bottom), and the smallest sample is the widest (the graph on the top). Increasing the sample size makes the graph narrower.

7.29 The top dotplot has the largest standard error because it is widest, and the bottom dotplot has the smallest standard error because it is narrowest.

7.31 Graph A is for the fair coin, because it is centered at 0.50.

7.33 a. 50% seniors **b.** 0% seniors **c.** 50% seniors **d.** 100% seniors

Section 7.3

7.35 $$z = \frac{0.38 - 0.35}{0.03} = \frac{0.03}{0.03} = 1$$

The area to the right of a z-score of 1.00 is 0.1587.

The probability that the percentage of mixed fruit juice buyers will be 38% or more is about 16%.

7.37 Because the sampling distribution for the sample proportion is approximately Normal, we know the probability of falling within one standard error is about 0.68. Therefore, the probability of falling more than one standard error away from the mean is about 0.32.

7.39 Answers are given in the order shown in the Guided Exercises.

1: $p = 0.65$

2: $np = 200(0.65) = 130$ which is more than 10

$n(1 - p) = 200(0.35) = 70$ which is more than 10

The other assumptions were given.

3:

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.65(0.35)}{200}} = \sqrt{\frac{0.2275}{200}} = \sqrt{0.0011375} = 0.03373$$

$$z = \frac{\hat{p} - p}{SE} = \frac{0.67 - 0.65}{0.03373} = 0.59$$

The area to the right of a z-value of 0.59 is 0.2776 (or 0.2766 with technology and no rounding of intermediate steps).

4: The probability is represented by the area to the right of 0.59, because the question asks for the probability that the sample proportion will be "at least" 0.67 (134 out of 200), which translates to a z-score of 0.59. This means we are asked to find the probability that the z-score will be 0.59 or greater.

5: The probability that at least 67% of 200 pass the exam is about 28% because the area to the left of a z-score of 0.59 is about 72%, and so the area to the right is about 28%.

7.41 a. We expect the proportion of overweight or obese children in the sample to be about 0.53. Because 0.50 is smaller than this, and because the sampling distribution is approximately Normal, there would be a greater than 50% chance of seeing a sample proportion greater than 0.50.

b. $$z = \frac{0.50 - 0.53}{\sqrt{\frac{0.53(0.47)}{200}}} = \frac{-0.03}{0.03529} = -0.85,$$ and the area to the right of a z-score of −0.85 is 0.802.

Thus the probability that more than half in the sample will be overweight or obese is about 80%.

7.43 a. 0.5 **b.** 0.5 **c.** 0.103 **d.** Lower, because the probability of guessing correctly on each question is lower when there are four options.

Section 7.4

7.45 a. I am 95% confident that the population proportion favoring the death penalty for those convicted of murder is between 0.599 and 0.641. **b.** Yes, it is plausible to claim that a majority favor the death penalty, because the interval includes values above 50%.

7.47 a. $\hat{p} = 261/435 = 0.60$ **b.** 1: *Random and Independent*: Gallup Polls are random samples, and the people selected are independent. 2: *Large Sample*: Because we do not know p, we will use $\hat{p}$ for these calculations; $n\hat{p} = 435(0.60) = 261$ and $n(1 - \hat{p}) = 435(0.40) = 174$, and both of these are more than 10. 3: *Big Population*: The population of East Germany is more than 10(435). **c.** I am 95% confident that the population proportion who think they are struggling in East Germany is between 0.554 and 0.646.

7.49 a. I am 95% confident that the population percentage of students choosing economics is between 23.5% and 32.5%. **b.** Yes, there is strong evidence that there will not be enough students for economics, because some values are below 25%. **c.** A sample from one section out of 12 sections would not be representative of the entire school and would be worthless in this context.

7.51 a. $$\sqrt{\frac{0.77 \times 0.23}{2000}} = 0.0094, \text{ or } 0.94\%$$ **b.** 1.84%

c. (75.16%, 78.84%) **d.** The interval supports this claim. Because 60% is not in the interval and all values are above 60%, we are confident that the current percentage is higher than 60%.

7.53 a. I am 99% confident that the population proportion of adults having a great deal of confidence or quite a lot of confidence in the public schools is between 0.264 and 0.316. **b.** I am 90% confident that the population proportion of adults having a great deal of confidence or quite a lot of confidence in the public schools is between 0.273 and 0.307. **c.** 99%

$(0.316 - 0.264 = 0.052)$ and 90% $(0.307 - 0.273 = 0.034)$. The 99% interval is wider. **d.** A 95% interval would be wider than the 90% interval but narrower than the 99% interval.

7.55 a. 4000 **b.** Yes, it is large enough because $np = 5000(0.8) = 4000$ and $n(1 - p) = 5000(0.20) = 1000$, and both are more than 10. **c.** (0.789, 0.811) **d.** $0.811 - 0.789 = 0.22$, or about 2% **e.** 16000 **f.** (0.7945, 0.8055) **g.** $0.8055 - 0.7945 = 0.0110$, or about 1%. **h.** No, when the sample size is multiplied by 4, the margin of error is divided by 2 because 1% is about half of 2%.

7.57 a. He or she should find 15, because 15 is half of 30. **b.** You would expect about 4 out of 40 not to capture 50%, because with a 90% confidence interval about 10% should not capture, and 10% of 40 is 4.

7.59 a. 45,513,001/70,686,784, or 64.4%, of the voters voted for Putin. **b.** Yes, it would be inappropriate to find a confidence interval. This is because confidence interval is needed only in case of a sample statistic and we want to generalize about the population from which the sample was drawn. In the present case, we have the population proportion.

7.61 a. $101/1974 = 5.12\%$ **b.** 95% CI (0.041, 0.061) **c.** 5% is plausible because it is inside the interval.

Section 7.5

7.63 This interval contains 0, which means it is plausible that both groups feel the same. A positive value means that a higher proportion of low-income people feel this way. It is plausible that the proportion of low-income people who feel this way is as much as 2 percentage points greater than the proportion of middle-income people who do. A negative value means a higher proportion in the middle-income population feel this way. It is plausible that the proportion of middle-income people who feel this way is as much as 8 percentage points greater than the proportion of low-income people who do.

7.65 a. No, we cannot conclude this. Although a greater proportion of the *sample* of stay-at-home moms report being stressed compared to the *sample* of employed moms, in the *population* of all moms these proportions might be the same or different—or even reversed. **b.** Sample 1: number of successes, 540; number of observations, 1000. Sample 2: number of successes, 490; number of observations, 1000. **c.** 1. *Random and Independent*: Gallup polls take random samples, and the respondents are independent. 2. *Large Samples*: $n_1\hat{p}_1 = 1000(0.54) = 540$, and $n_1(1 - \hat{p}_1) = 1000(0.46) = 460$, and $n_2\hat{p}_2 = 1000(0.49) = 490$, and $n_2(1 - \hat{p}_2) = 1000(0.51) = 510$; and all four numbers are more than 10. 3. *Big Populations*: There are far more than 10(1000) of each type of mom in the United States. 4. *Independent Samples*: The stay-at-home moms are not linked with the working moms in any way. **d.** (0.00625, 0.09375) I am 95% confident that the difference in population proportions (stay-at-home minus working moms) is between 0.006 and 0.094 (between 0.6% and 9.4%). Because this interval does not capture 0, we can conclude that the population proportions are not the same. Because both boundaries are positive, we can conclude that the stay-at-home moms experience more stress than the working moms.

7.67 Answers are given in the order shown in the Guided Exercises.
1: 29/64, or 45.3%, of the children who did not go to preschool graduated from high school.
2: In this sample, the children who attended preschool were more likely to graduate from high school.
3: $n_2\hat{p}_2 = 64(0.453) = 29$
$n_2\hat{p}_2 = 64(0.547) = 35$
4: I am 95% confident that the difference in proportions graduating (Preschool rate minus No Preschool rate) is between 0.022 and 0.370.
5: Statement ii is correct.
6: We cannot generalize to a larger population because this was not a random sample.
7: We can conclude that the Perry Preschool caused the higher graduation rate because of the random assignment.

7.69 a. $817/922 = 88.6\%$ recovered with the amoxicillin, and $785/922 = 85.1\%$ recovered with the placebo. Thus the group receiving the antibiotic did better. **b.** 1. *Random and Independent*: Although we do not have random samples, we have random assignment to groups. 2. *Large Samples*: $n_1\hat{p}_1 = 922(0.886) = 817$, and $n_1(1 - \hat{p}_1) = 922(0.114) = 105$, and $n_2\hat{p}_2 = 922(0.851) = 785$, and $n_2(1 - \hat{p}_2) = 922(0.149) = 137$; and all four numbers are more than 10. 3. *Big Populations*: There are far more than 10(922) children with severe acute malnutrition in Malawi. 4. *Independent Samples*: The children in the group that received the drug were unrelated to the children in the group that received the placebo. **c.** I am 95% confident that the difference in proportions (antibiotic minus placebo) is between 0.0039 and 0.0655. The interval does not capture 0. It is not plausible that the recovery rates are the same. The antibiotic with food caused a better recovery rate than the placebo with food. We can conclude causation because the study was a randomized experiment, but we cannot generalize widely because we did not have a random sample from a larger population.

7.71 a. 62.5% of men and 74.5% of women used turn signals. **b.** $(-0.166, -0.073)$ I am 95% confident that the population percentage of men using turn signals minus the population percentage of women using turn signals is between −16.6% and −7.3%. The interval does not capture 0, so we are confident that the percentages are different. This shows that women are more likely than men to use turn signals. **c.** 62.8% of the men and 73.8% of the women used turn signals. CI: (−0.2575, 0.03739). I am 95% confident that the population percentage of men using turn signals minus the population percentage of women using turn signals is between −25.8% and 3.7%. The interval captures 0, showing it is plausible that the percentages are the same in the population. **d.** With more data (part b), we had a narrower margin of error and thus a more precise estimate of the true difference in proportions. We could be confident that the percentages were different in the population. In part d, the very wide interval did not allow us to make this call.

7.73 a. No, the rate of miscarriages was higher for the unexposed women, which is the opposite of what was feared. **b.** There was not a random sample, and also there was not random assignment.

Chapter Review Exercises

7.75 a. I am 95% confident that the population percentage opposed to banning super-size sugary soft drinks is between 62% and 68%. **b.** Narrower **c.** Wider **d.** No, the size of the population is not a factor to consider as long as the population is much larger than the sample.

7.77 The sample proportion must be 44% because the interval is symmetric around the sample proportion, which is in the middle.

7.79 The margin of error must be 4%. From the sample proportion to find the upper boundary you go up one margin of error and to find the lower boundary you go down one margin of error. Therefore, the boundaries are separated by two margins of error, and half of 8% is 4%.

7.81 a. 28/50 is 56% and 112/200 is also 56%, so the percentages are the same. **b.** $SE = \sqrt{\dfrac{p(1-P)}{n}} = \sqrt{\dfrac{0.5(1-0.5)}{50}} = \sqrt{0.005} = 0.07071$, which is the standard error for a sample size of 50.

$$SE = \sqrt{\frac{p(1-P)}{n}} = \sqrt{\frac{0.5(1-0.5)}{200}} = \sqrt{0.00125} = 0.03535,$$ which is the standard error for a sample size of 200.

c. When using the CLT for one sample proportion, if you increase the sample size, the standard error will <u>decrease</u> which makes the z-score <u>farther from 0</u>, and that makes the tail area(s) (areas at the left or right edge of the curve) <u>smaller</u>.

7.83 1: *Random and Independent*: given. 2: *Large Sample*: $np = 200(0.29) = 58$ and $n(1 - p) = 200(0.71) = 142$, and both of these are more than 10. 3: *Big Population*: The population of dreamers is more than 10(200).

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.29(0.71)}{200}} = \sqrt{\frac{0.2059}{200}} = \sqrt{0.00103} = 0.0320858$$

$$z = \frac{0.50 - 0.29}{0.0320858} = \frac{0.21}{0.0320858} = 6.54$$

The area of the normal curve to the right of a z-value of 6.54 is less than 0.001. Therefore, the probability that a sample of 200 will contain 50% or more dreaming in color is less than 0.001.

7.85 Because of sampling variability, it is possible that the actual number of persons exercising daily is slightly lower and those meditating is slightly higher. This difference between actual number of persons and the number in this particular sample could be small enough that in the population, both options actually have the same number of people. In other words, a 95% confidence interval for the difference in the proportion of all persons who exercise minus the proportion of all persons who meditate would include 0.

7.87 **a.** $105/188 = 55.9\%$ have career goals in 1997. Also, $403/610 = 66.1\%$ have career goals in 2011. Thus the rate of career goals went up between 1997 and 2011. (The change is 10.2 percentage points.)
b. $(-0.1825, -0.0218)$ or $(0.0218, 0.1825)$ I am 95% confident that the change in population proportions with high-paying career goals (2011 proportion minus 1997 proportion) is between 0.0218 and 0.1825. The interval does not capture 0, showing there is a difference in the population proportions. Women in 2011 preferred a high-paying career in greater proportions than women in 1997.

7.89 $n = \dfrac{1}{m^2} = \dfrac{1}{(0.03)^2} = \dfrac{1}{0.0009} = 1111.11$

Thus we would need a sample size of 1111 or 1112 to get a margin of error of 3 percentage points.

7.91 Marco took a convenience sample. The students may not be representative of the voting population, so the proposition may not pass.

7.93 No, the people you met would not be a random sample but a convenience sample.

7.95 The small mean might have occurred by chance.

CHAPTER 8

Answers may vary slightly due to type of technology or rounding.

Section 8.1

8.1 population parameter

8.3 H_a: The proportion of criminals who attend boot camp who return to prison is less than 0.40.

H_a: $p < 0.40$

8.5 **a.** iii **b.** iii

8.7 i.

8.9 ii.

8.11 does not, 0.05

8.13 **a.** H_0: $p = 0.30$, H_a: $p \neq 0.30$ **b.** $z = 0.77$

8.15 **a.** 0.44 **b.** 0.40

c. $z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \dfrac{0.44 - 0.40}{\sqrt{\dfrac{0.4(0.6)}{100}}} = \dfrac{0.04}{\sqrt{0.0024}} = \dfrac{0.04}{0.04899} = 0.82$

The value of the test statistic tells us that the observed proportion was 0.82 standard errors above the null hypothesis value of 0.40.

8.17 **a.** You would expect about 10 right out of 20 if the person is guessing.
b. The smaller p-value will come from person B. The p-value measures how unusual an event is, assuming the null hypothesis is true. Getting 18 right out of 20 is more unusual than getting 13 right out of 20 when you are expecting 10 right under the null hypothesis. The larger difference in proportions ($\hat{p} - p$) results in a smaller p-value.

8.19 The probability that a person will get 13 or more right, if the person is truly guessing, is about 9%.

8.21 It would be low. An outcome, 19/20, has occurred that would be very surprising if the student were just guessing.

Section 8.2

8.23 Random sample was mentioned. Independence is assumed.
Large sample: $np_0 = 113(0.29) = 32.77 > 10$ and $n(1 - p_0) = 113(0.71) = 80.23 > 10$.
Large population: There are more than 1130 people in the population of dreamers.
So the conditions are met.

8.25 Figure B is correct because it is one-tailed. The p-value is 0.173. If the population proportion that believe marriage is obsolete is 0.38, there is about a 17% chance of getting 782 or more out of 2004 randomly selected adults who believe that.

8.27 Figure B is correct because the alternative hypothesis should be one-sided, because the person should get better than half right if she or he can tell the difference.

8.29 *Step 1:* H_0: $p = 0.33$, H_a: $p \neq 0.33$. *Step 2:* One-proportion z-test: $0.33(200) = 66 > 10$, other product is larger, sample random and assumed independent, $\alpha = 0.05$. Entries: p_0: 0.33, x: 42, n: 200, two-sided ($\neq$).

8.31 *Step 3:* $z = -3.61$, p-value < 0.001. *Step 4:* Reject H_0. The proportion that sleep walk is significantly different from 0.33.

8.33 In Figure (A), the shaded area could be a p-value because it includes tail areas only; it would be for a two-sided alternative because both tails are shaded. In Figure (B) the shaded area would not be a p-value because it is the area between two z-values.

8.35 *Step 1:* H_0: $p = 0.52$, H_a: $p \neq 0.52$, where p is the population proportion favoring stricter gun control. *Step 2:* One-proportion z-test, random sample with independent measurements, sample size large (1011 times 0.48 = about 485, which is more than 10), and population large, $\alpha = 0.05$. *Step 3:* $\hat{p} = 0.4896$, $SE = 0.01571$, $z = -1.93$, p-value $= 0.053$. *Step 4:* Do not reject H_0. The percentage is not significantly different from 52%. (Choose conclusion i.)

8.37 **a.** $527/1014 = 52.0\%$. This is less than the 58%.
b. *Step 1:* H_0: $p = 0.58$, H_a: $p \neq 0.58$, where p is the population proportion of people who believe there is global warming. *Step 2:* One-proportion z-test, $0.58(1014) = 588 > 10$ and $0.42(1014) = 426 > 10$, sampling random and independent, population more than 10 times sample, $\alpha = 0.05$. *Step 3:* $z = -3.89$, p-value < 0.001. *Step 4:* Reject H_0.
c. Choose ii.

8.39 *Step 1:* H_0: $p = 0.20$, H_a: $p \neq 0.20$, where p is the population proportion of dangerous fish. *Step 2:* One-proportion z-test, $0.2(250) = 50 > 10$ and $0.8(250) = 200 > 10$, population large, assume a random and independent sample, $\alpha = 0.05$. *Step 3:* $z = 1.58$, p-value $= 0.114$. *Step 4:* Do not reject H_0. We are not saying the percentage is 20%. We are only saying that we cannot reject 20%. (We might have been able to reject the value of 20% if we had had a larger sample.)

8.41 *Step 1:* H_0: $p = 0.09$, H_a: $p \neq 0.09$, where p is the population proportion of t's in the English language. *Step 2:* One-proportion z-test, the sample is independent, random, and $0.09(600) = 54 > 10$ and $0.91(600) = 546 > 10$, $\alpha = 0.10$. *Step 3:* $z = -0.86$, p-value $= 0.392$. *Step 4:* Do not reject H_0. We cannot reject 9% as the current proportion of t's because 0.392 is more than 0.10.

Section 8.3

8.43 iv. $z = 3.00$. It is farthest from 0 and therefore has the smallest tail area.

8.45 The null hypothesis is that the penny is not biased. The first kind of error is saying the penny is biased (rejecting the null hypothesis of no bias) when in fact the penny is not biased. The second kind of error is saying the penny is not biased (not rejecting the null hypothesis) when in fact it is biased.

8.47 The first type of error is having the innocent person suffer (convicting an innocent person). The second type of error is "ten guilty persons escape" (letting guilty persons go free).

8.49 We don't use "prove" with inferential statistics because we are not 100% sure. She could say she could reject the hypothesis that the coin is unbiased. Or she could say the data are consistent with a biased coin, or the data demonstrate bias.

8.51 Choose hypothesis testing and the one-proportion z-test, because he only wants to know whether or not it will pass; he is not interested in knowing the proportion who will vote for it. Suppose p is the population proportion supporting the proposition.
H_0: $p = 0.50$
H_a: $p > 0.50$
$z = 5.06$
p-value < 0.001
Reject H_0. The proposition is likely to pass.

8.53 Interpretation iv.

8.55 No; we don't use "prove" because we cannot be 100% sure of conclusions based on chance processes.

8.57 It is a null hypothesis.

8.59 Interpretations b and d are valid. Interpretations a and c are both "accepting" the null hypothesis claim, which is an incorrect way of expressing the outcome.

Section 8.4

8.61 Far apart. Assuming the standard errors are the same, the farther apart the two proportions are, the larger the absolute value of the numerator of z, and therefore the larger the absolute value of z and the smaller the p-value.

8.63 The answers follow the guidance.
Step 0: The survival rate for those taking idel was about 92% and for those taking the placebo was 80%. So the group taking idel appears to have done better. However, this difference might have been due to chance.
Step 1: H_0: $p_{idel} = p_{plac}$
Step 2: Because the two sample sizes are equal ($n_1 = n_2$), the numbers below are the same as the numbers calculated using n_1.
$n_2 \times \hat{p} = 110(0.8591) = 94.5$
$n_2 \times (1 - \hat{p}) = 110(0.1409) = 15.5$
All four numbers are more than 10.
Step 3: $z = 2.52$
p-value $= 0.006$
Step 4: Reject H_0. Choose ii.
Causality: Yes, because of the random assignment to groups, as well as the double-blind nature of the study, we can say that idel caused the better result.

8.65 **a.** For nicotine gum, the proportion quitting was 0.106. For the placebo, it was 0.040. This was what was hoped for—that the drug was helpful compared to the placebo. **b.** $z = 7.23$ (or -7.23).

8.67 87/211 (or 41.2%) of the therapy group were convicted, and 74/198 (or 37.4%) of the control group were convicted. Thus, contrary to expectations, the experimental group had a higher conviction rate. It is not necessary to do a complete analysis, because the sample rate is higher in the MST group than in the therapy group, so the hypothesis test cannot conclude that the population rate for the therapy is lower. If you do a hypothesis test, you will get a p-value of 0.212 and reach the same conclusion.

8.69 **a.** Men: 46.2% smiling. Women: 51.1% smiling. **b.** *Step 1:* H_0: $p_{men} = p_{women}$, H_a: $p_{men} \neq p_{women}$, where p is the proportion smiling. *Step 2:* Two-proportion z-test, expected counts (3461, 4279, 3614, 4470) are all larger than 10, assume random and independent sample, population large, $\alpha = 0.05$. *Step 3:* $z = 6.13$ (or -6.13), p-value < 0.001. *Step 4:* Reject H_0. There is a significant difference between smiling rates for men and women. **c.** The difference is significant because of the large sample size.

Chapter Review Exercises

8.71 **a.** Two-proportional z-test. One population is all cricketers, and the other population is all soccer players. **b.** One-proportion z-test. The population is all students of the state.

8.73 **a.** p_b is the population portion of boys who devote equal time to play and homework. P_g is the population portion of girls who devote equal time to play and studies.
H_0: $p_b = p_g$
H_a: $p_b \neq p_g$
Two-proportion z-test.
b. $p =$ the population proportion of correct answers.
H_0: $p = 0.50$ (she is just guessing), H_a: $p > 0.50$ (she can distinguish the taste).
One-proportion z-test.

8.75 6 right out of 20 is less than half, so he cannot tell the difference. (Or: H_0: $p = 0.50$, H_a: $p > 0.50$, $z = -1.79$, p-value $= 0.963$, do not reject H_0.)

8.77 0.05 (because $1 - 0.95 = 0.05$).

8.79 10% of 500, or 50.

8.81 The board declares on *all* students, the entire population, so inference is inappropriate or not needed. Also, rates have been given instead of counts.

8.83 **a.** 336 (from 0.60 times 560) said more strict in February 1999, and 370 said more strict in late April 1999. **b.** *Step 1:* H_0: $p_{\text{more strict Feb}} = p_{\text{more strict April}}$, H_a: $p_{\text{more strict Feb}} \neq p_{\text{more strict April}}$. *Step 2:* Two-proportion z-test, samples are random, expected counts (353, 207, 353, 207) are all larger than 10, population large, independence within samples and independence between samples, $\alpha = 0.01$. *Step 3:* $z = 2.10$ (or -2.10), p-value $= 0.035$. *Step 4:* Do not reject H_0, because $\alpha = 0.01$. **c.** 672 out of 1120 in February vs. 739 out of 1120 in late April. *Step 1:* H_0: $p_{\text{more strict Feb}} = p_{\text{more strict April}}$, H_a: $p_{\text{more strict Feb}} \neq p_{\text{more strict April}}$. *Step 2:* Two-proportion z-test, samples are random, expected counts (706, 414, 706, 414) are all larger than 10, population large, independence within samples and independence between samples, $\alpha = 0.01$. *Step 3:* $z = 2.93$ (or -2.93), p-value $= 0.003$. *Step 4:* Reject H_0. The results of the polls were significantly different from each other. **d.** With a larger sample size (more evidence), we got a smaller p-value and were able to reject H_0.

8.85 The data were not from a sample but the entire population of Zimbabwe. The inference will not be reasonable, so it would not be appropriate to do such a test.

8.87 *Step 1:* H_0: $p = 0.63$, H_a: $p \neq 0.63$, where p is the population proportion of employers who allow workers to work from home sometimes. *Step 2:* One-proportion z-test, $0.63(400) = 252 > 10$ and $0.37(400) = 148 > 10$, and sample is random and independent, $\alpha = 0.10$. *Step 3:* $z = -1.76$, p-value $= 0.078$. *Step 4:* Reject the null hypothesis because the p-value is less than the significance level of 0.10. The population proportion of employers who allow employees to work from home sometimes is significantly different from 0.63 at the 0.10 significance level.

8.89 **a.** $751/1138 = 66.0\%$, and $523/1138 = 46.0\%$. Both questions are similar in nature (both ask about increasing environmental controls or regulations), but the question that did not mention the expense got a more positive result for regulation. **b.** *Step 1:* H_0: $p_1 = p_2$, H_a: $p_1 \neq p_2$, where p_1 is the population proportion supporting regulation which had the expense mentioned, and p_2 is the population proportion supporting regulation with no mention of expense. *Step 2:* Two-proportion z-test, expected counts (637, 501, 637, 501) are all larger than 10 and the sample is presumed random, observations independent, and samples independent, $\alpha = 0.05$. *Step 3:* $z = -9.63$ or 9.63, p-value < 0.001. *Step 4:* Reject H_0. The proportions are significantly different. The wording seems to be associated with a difference in responses. **c.** (0.160, 0.240). I am 95% confident that the population difference in percentages is between 16% and 24%. Because this does not capture 0, we conclude that there is a real difference in the population proportions. (Note that the interval is centered at a difference of 20% (from 66% – 46%) and that the margin of error is 4 percentage points.)

8.91 **a.** The misconduct rate was higher for those in the sample who did *not* have three strikes (30.6%) than for those in the sample who had three strikes (22.2%). This was not what was expected. **b.** *Step 1:* H_0: $p_{\text{three-strikers}} = p_{\text{others}}$, H_a: $p_{\text{three-strikers}} > p_{\text{others}}$. *Step 2:* Two-proportion z-test, expected counts (213, 924, 521, 2264) are all larger than 10, assume random samples and assume independence, $\alpha = 0.05$. *Step 3:* $z = -4.49$ (or 4.49), p-value > 0.999. *Step 4:* Do not reject H_0. The three-strikers do not have a greater rate of misconduct than the other prisoners. (If a two-sided test had been done, the p-value would have been < 0.001, and we would have rejected the null hypothesis because the three-strikers had *less* misconduct.)

8.93 a. p_{Gallup} is the population proportion with a gun in the house according to the Gallup Poll.

p_{Pew} is the population proportion with a gun in the house according to the Pew Poll.

Step 1: H_0: $p_{\text{Gallup}} = p_{\text{Pew}}$, H_a: $p_{\text{Gallup}} \neq p_{\text{Pew}}$. *Step 2*: Two-proportion z-test, expected counts (365, 635, 365, 635) are all larger than 10, the polling agencies say they use independent, random samples, samples are assumed independent and population is large, $\alpha = 0.05$. *Step 3:* $z = 3.25$ (or -3.25), p-value $= 0.001$. *Step 4:* Reject H_0. The proportions are significantly different. **b.** (0.028, 0.112). I am 95% confident that the difference in population proportions (Gallup minus Pew) is between 2.8% and 11.2%. Because this does not capture 0, we can reject the hypothesis that the percentages are the same. (Note the difference in percentages; 40% – 33% is about 7%, which is the center of the interval, and the margin of error is about 4.2 percentage points.)

8.95 a. iii **b.** iv

8.97 $z = 1.22$.

8.99 The p-value tells us that if the true proportion of those who text while driving is 0.25, then there is only a 0.034 probability that one would get a sample proportion of 0.125 or smaller with a sample size of 40.

8.101 He has not demonstrated ESP; 10 right out of 20 is only 50% right, which you should expect from guessing.

8.103 H_0: The death rate after starting hand washing is still 9.9%, or $p = 0.099$ (p is the proportion of all deaths at the clinic.)
H_a: The death rate after starting hand washing is less than 9.9%, or $p < 0.099$.

8.105 *Step 3:* $z = 8.33$, p-value $= 0.000$. *Step 4:* Reject H_0. The probability of doing this well by chance alone is so small that we must conclude that the student is not guessing.

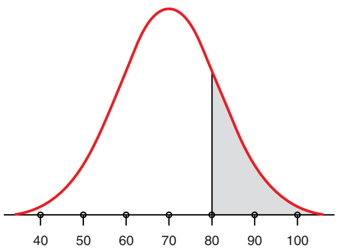
CHAPTER 9

Answers may vary due to rounding or type of technology used.

Section 9.1

9.1 a. They are statistics because they come from a sample.
b. $\bar{x} = 7.9$, $s = 0.9$

9.3 a. See the accompanying figure.



b. 16% (By the Empirical Rule, 68% of the observed values are between 60 and 80, which leaves 32% outside of those boundaries. But we want only the right half.)

9.5 To have an "unbiased" estimator does not mean that the sample mean will be exactly equal to the population mean. It only means that if we take all possible samples of rodents, the mean of all of the sample means would be the same as the population mean. In this case, the sample mean is based on a random sample, so different samples will have different means.

9.7 The sampling distribution of means.

9.9 a. 3600 is the expected sample mean, assuming an unbiased sample. The sample mean is an unbiased estimator of the population mean.

b. $\dfrac{1404}{\sqrt{81}} = 156$

Section 9.2

9.11 a. 68% from the Empirical Rule. 6.4 is one standard deviation below the mean, and 7.6 is one standard deviation above the mean.

b. $\dfrac{0.6}{\sqrt{4}} = 0.3$, so the standard error for the mean is 0.3. Now 7.6 is two standard errors above the mean, and 6.4 is two standard errors below the mean. The answer is 95% from the Empirical Rule. **c.** The distribution of means is taller and narrower than the original distribution and will have more data in the central area.

9.13 a. Yes, the sample is large (more than 25), and so the distribution of means will be normal. **b.** The mean is 3600, and the standard error is 156 $\left(\text{from } \dfrac{1404}{\sqrt{81}}\right)$. **c.** 5% from the Empirical Rule (if 95% of the data lie within two standard deviations (errors) of the mean, then 5% lie beyond two standard deviations (errors) of the mean.

9.15 Figure B is the original distribution; it is the least Normal and widest. Figure A is from samples of 5. Figure C is from samples of 25; it is the narrowest and the least skewed. The larger the sample size, the narrower and more Normal is the sampling distribution.

9.17 a. 7.9 and 7.7 are parameters; 8.2 and 6.0 are statistics. **b.** $\mu = 7.9$, $\sigma = 7.7$, $\bar{x} = 8.2$, $s = 6.0$ **c.** Random: OK. Sample size: $40 > 25$ OK. The shape would be Normal.

Section 9.3

9.19 a. Only ii. is correct. We are 95% confident that the population mean is between 37.5% and 49.5%. **b.** Yes we can reject 50% because it is not in the interval.

9.21 a. i. is correct: (10.125, 10.525). Both ii. and iii. are incorrect. **b.** No, it does not capture 10. Reject the claim of 10 pounds, because 10 is not in the interval.

9.23 a. C-interval **b.** C-level

9.25 Use $t^* = 2.056$.

9.27 a. Sample size: 30; Sample mean: 0.51; Standard deviation: 0.2 **b.** I am 95% confident that the population mean weight of the hamburgers is between 0.44 and 0.58 pounds.

9.29 a. $m = t^*\dfrac{s}{\sqrt{n}} = 2.064\dfrac{13}{\sqrt{25}} = 5.37$

$72 + 5.27 = 77.37$
$72 - 5.37 = 66.63$

I am 95% confident that the mean is between 67 and 77 beats per minute.

b. $m = t^*\dfrac{s}{\sqrt{n}} = 2.797\dfrac{13}{\sqrt{25}} = 7.27$

I am 99% confident that the mean is between 65 and 79 beats per minute.

c. The 99% interval is wider because a greater level of confidence requires a bigger value for t^*.

9.31 a. The first one (2.60, 3.20) is wider ($3.2 - 2.6 = 0.6$), so it would be for the 95% confidence. The second interval is narrower ($3.15 - 2.65 = 0.5$), and it would be for the 90% confidence. **b.** Narrower: A larger sample size produces a narrower interval.

9.33 a. Wider **b.** Narrower **c.** Wider

9.35 a. I am 95% confident that the population mean is between 173.74 and 207.26 cm. **b.** The interval does not capture 170 cm. There is enough evidence to reject 170 cm as the population mean.

Section 9.4

9.37 Answers follow the guided steps given. *Step 1:* H_0: $\mu = 98.6$, H_a: $\mu \neq 98.6$. *Step 2:* One-sample t-test: random and independent sample, not strongly skewed, $\alpha = 0.05$. *Step 3:* $t = -1.65$, p-value $= 0.133$. *Step 4:* Do not reject H_0. We cannot reject 98.6 as the population mean at the 0.05 level (choice i).

9.39 **a.** You should be able to reject 20 pounds because the confidence interval (20.4 to 21.7) did not capture 20 pounds. **b.** *Step 1*: H_0: $\mu = 20$, H_a: $\mu \neq 20$. *Step 2*: One-sample *t*-test: Normal, random, and independent, $\alpha = 0.05$. *Step 3*: $t = 5.00$, p-value $= 0.015$. *Step 4*: Yes, reject H_0. Choose ii.

9.41 *Step 1*: H_0: $\mu = 2.81$, H_a: $\mu > 2.81$ *Step 2*: One-sample *t*-test: conditions are met, $\alpha = 0.05$. *Step 3*: $t = 4.91$, p-value $= 0.000$. *Step 4*: Do not reject H_0. We have *not* shown that the mean is significantly more than 2.81.

9.43 **a.** *Step 1:* H_0: $\mu = 38$, H_a: $\mu \neq 38$. *Step 2:* One-sample *t*-test: Normal and random, $\alpha = 0.05$. *Step 3:* $t = -1.03$, p-value $= 0.319$. *Step 4:* Do not reject H_0. The mean for non-U.S. boys is not significantly different from 38. **b.** *Step 3:* $t = -1.46$, p-value $= 0.155$. *Step 4:* Do not reject H_0. The mean for non-U.S. boys is not significantly different from 38. **c.** Larger *n*, smaller standard error (narrower sampling distribution) with less area in the tails, as shown by the smaller p-value.

9.45 **a.** 3.25, higher **b.** *Step 1:* H_0: $\mu = 2.81$, H_a: $\mu > 2.81$. *Step 2:* One-sample *t*-test: random and $30 > 25$, $\alpha = 0.05$. *Step 3:* $t = 4.91$, p-value < 0.001. *Step 4:* Reject H_0. The mean GPA for Oxnard College statistics students is significantly higher than 2.81.

9.47 I am 95% confident that the population mean GPA is between 3.07 and 3.43. Yes, we would reject 2.81 because it is not in the interval.

9.49 *Step 1:* H_0: $\mu = 0$, H_a: $\mu > 0$. *Step 2:* One-sample *t*-test: Normal, not random (don't generalize), $\alpha = 0.05$. *Step 3:* $t = 3.60$, p-value $= 0.003$. Reject H_0. There was a significant weight loss.

9.51 Expect $0.80(500) = 400$ to capture and 100 to miss.

Section 9.5

9.53 **a.** Paired **b.** Independent

9.55 **a.** The samples are random, independent, and large, so the conditions are met. **b.** I am 95% confident that the mean difference (OC minus MC) is between -0.40 and 1.14 TVs. **c.** The interval for the difference captures 0, which implies that it is plausible that the means are the same.

9.57 Answers follow the guided steps given. *Step 1:* H_0: $\mu_{oc} = \mu_{mc}$, H_a: $\mu_{oc} \neq \mu_{mc}$, where μ is the population mean number of TVs. *Step 2:* Two-sample *t*-test: samples large ($n = 30$), independent, and random, $\alpha = 0.05$. *Step 3:* $t = 0.95$, p-value $= 0.345$. *Step 4:* Do not reject H_0. Choose i. Confidence interval: $(-0.404, 1.138)$. Because the interval for the difference captures 0, we cannot reject the hypothesis that the mean difference in number of TVs is 0.

9.59 **a.** The men's sample mean triglyceride level of 139.5 was higher than the women's sample mean of 84.4. **b.** *Step 1:* H_0: $\mu_{men} = \mu_{women}$, H_a: $\mu_{men} > \mu_{women}$, where μ is the population mean triglyceride level. *Step 2:* Two-sample *t*-test: assume the conditions are met, $\alpha = 0.05$. *Step 3:* $t = 4.02$ or -4.02, p-value < 0.001. *Step 4:* Reject H_0. The mean triglyceride level is significantly higher for men than for women. Choose output B: Difference $= \mu_{female} - \mu_{male}$, which tests whether this difference is less than 0, and that is the one-sided hypothesis that we want.

9.61 $(-82.5, -27.7)$; because the difference of 0 is not captured, it shows there is a significant difference. Also, the difference $\mu_{female} - \mu_{male}$ is negative, which shows that the men's mean (triglyceride level) is significantly higher than the women's mean.

9.63 *Step 1:* H_0: $\mu_{men} = \mu_{women}$, H_a: $\mu_{men} \neq \mu_{women}$, where μ is the population mean clothing expense for one month. *Step 2:* Two-sample *t*-test: assume Normal and random, $\alpha = 0.05$. *Step 3:* $t = 1.42$ or -1.42, p-value $= 0.171$. *Step 4:* Do not reject H_0. The mean clothing expense is not significantly different for men and women.

9.65 **a.** The 95% interval would capture 0, because we could not reject the hypothesis that the mean amounts spent on clothing are the same. **b.** A 99% interval would also capture 0, because it is wider than the 95% interval and centered at the same place. **c.** $(-18.4, 97.7)$ because the interval captures 0, we cannot reject the hypothesis that the mean difference in spending on clothing is 0, which shows we cannot reject the hypothesis that the means are the same.

9.67 **a.** $\bar{x}_{UCSB} = \$61.01$ and $\bar{x}_{CSUN} = \$75.55$, so the sample mean at CSUN was larger. **b.** *Step 1:* H_a: $\mu_{UCSB} \neq \mu_{CSUN}$ where μ is the population mean book price. *Step 2:* Paired *t*-test, matched pairs, assume random and Normal (given). *Step 3:* $t = -3.21$ or 3.21, p-value $= 0.004$. *Step 4:* You can reject H_0. The means are significantly different.

9.69 The answers follow the guided steps. *Step 1:* H_0: $\mu_{before} = \mu_{after}$, H_a: $\mu_{before} < \mu_{after}$, where μ is the population mean pulse rate. *Step 2:* Paired *t*-test: each woman is measured twice (repeated measures), so a measurement in the first column is coupled with a measurement of the same person in the second column, assume random and Normal, $\alpha = 0.05$. *Step 3:* $t = 4.90$ or -4.90, p-value < 0.001. *Step 4:* Reject H_0. The sample mean before was 74.8, and the sample mean after was 83.7. The pulse rates of women go up significantly after they hear a scream.

9.71 Choose Figure B. The items are paired because the same items were priced at each store. *Step 1:* H_0: $\mu_{target} = \mu_{wholefoods}$, H_a: $\mu_{target} \neq \mu_{wholefoods}$. *Step 2:* Paired *t*-test: assume random, sample size large, $\alpha = 0.05$. *Step 3:* $t = -1.26$ or 1.26, p-value $= 0.217$. *Step 4:* Do not reject H_0. The means are not significantly different.

9.73 **a.** $\bar{x}_{groom} = 27.3$ was larger ($\bar{x}_{bride} = 25.9$). **b.** *Step 1:* H_0: $\mu_{bride} = \mu_{groom}$, H_a: $\mu_{bride} \neq \mu_{groom}$, where μ is the population mean age at marriage. *Step 2:* Paired *t*-test: random, large samples, $\alpha = 0.05$. *Step 3:* $t = 2.24$ or -2.24, p-value $= 0.033$. *Step 4:* Reject H_0. The mean ages of brides and grooms are significantly different. **c.** H_a: $\mu_{bride} < \mu_{groom.}$ The new p-value would be half of 0.033, or about 0.017.

9.75 **a.** 95% CI $(-1.44, 0.25)$ captures 0, so the hypothesis that the means are equal cannot be rejected. **b.** *Step 1:* H_0: $\mu_{measured} = \mu_{reported}$, H_a: $\mu_{measured} \neq \mu_{reported}$, where μ is population mean height of men. *Step 2:* Paired *t*-test: each person is the source of two numbers, assume conditions for *t*-tests hold, $\alpha = 0.05$. *Step 3:* $t = 1.50$ or -1.50, p-value $= 0.155$. *Step 4:* Do not reject H_0. The mean measured and reported heights are not significantly different for men or there is not enough evidence to support the claim that the typical self-reported height differs from the typical measured height for men.

9.77 **a.** I am 95% confident that the difference in population means (female minus male) is between –3.72 and 0.95. Because the interval captures 0, we cannot reject the hypothesis that the population means are the same. **b.** *Step 1*: H_0: $\mu_{men} = \mu_{women}$, H_a: $\mu_{men} \neq \mu_{women}$, where μ is the population mean weight for backpacks. *Step 2*: Two-sample *t*-test: random and large samples, $\alpha = 0.05$. *Step 3*: $t = 1.18$ or -1.18 p-value $= 0.242$. *Step 4:* Do not reject H_0. We do not have enough evidence to show that the means are significantly different.

9.79 **a.** I am 95% confident that the difference in population means (weekday minus weeknight) is between –1.53 and –1.09. Because the interval does not capture 0, we can reject the hypothesis that the population means are the same. Because all the values are negative, we can conclude that people tend to get more sleep on the weekends. **b.** *Step 1:* H_0: $\mu_{weekday} = \mu_{weekend}$, H_a: $\mu_{weekday} \neq \mu_{weekend}$, where μ is the population mean number of sleep hours. *Step 2:* Paired *t*-test: random and large sample size. *Step 3:* $t = -11.84$, p-value < 0.001. *Step 4:* Reject H_0. We have shown that the population means are significantly different.

Chapter Review Exercises

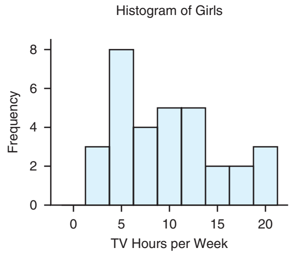
9.81 **a.** No *t*-test (two categorical variables) **b.** Paired *t*-test (each restaurant visited twice) **c.** Two sample *t*-test

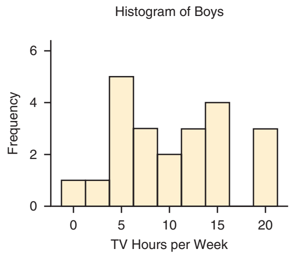
9.83 **a.** H_a: $\mu \neq 3.18$, $t = 3.66$, p-value $= 0.035$, reject H_0. The mean is significantly different from 3.18. **b.** H_a: $\mu < 3.18$, $t = 3.66$, p-value $= 0.982$, do not reject H_0. The mean is not significantly less than 3.18 ounces. **c.** H_a: $\mu > 3.18$, $t = 3.66$, p-value $= 0.018$, reject H_0. The mean is significantly more than 3.18 ounces.

9.85 *Step 1:* H_0: $\mu_{men} = \mu_{women}$, H_a: $\mu_{men} > \mu_{women}$. *Step 2:* Two-sample *t*-test: near Normal, random, $\alpha = 0.05$. *Step 3:* $t = 5.27$ (or -5.27), p-value < 0.001. *Step 4:* Reject H_0. The mean for brain size for men is significantly more than the mean for women.

9.87 *Step 1:* H_0: $\mu_{before} = \mu_{after}$, H_a: $\mu_{before} < \mu_{after}$, where μ is the population mean pulse rate (before and after coffee). *Step 2:* Paired t-test (repeated measures): assume conditions hold, $\alpha = 0.05$. *Step 3:* $t = 2.96$ or -2.96, p-value $= 0.005$. *Step 4:* Reject H_0. Heart rates increase significantly after coffee. (The average rate before coffee was 82.4, and the average rate after coffee was 87.5.)

9.89 The typical number of hours was a little higher for the boys, and the variation was almost the same. $\bar{x}_{girls} = 9.8$, $\bar{x}_{boys} = 10.3$, $s_{girls} = 5.4$ and $s_{boys} = 5.5$. See the histograms.





Step 1: H_0: $\mu_{girls} = \mu_{boys}$, H_a: $\mu_{girls} \neq \mu_{boys}$, where μ is the population mean number of TV viewing hours. *Step 2:* Two-sample t-test: random samples, assume the sample sizes of 32 girls and 22 boys is large enough that slight non-Normality is not a problem, $\alpha = 0.05$. *Step 3:* $t = -0.38$ or 0.38, p-value $= 0.706$. *Step 4:* You cannot reject the null hypothesis. There is not enough evidence to conclude that boys and girls differ in the typical hours of TV watched.

9.91 a. The mean for the day shift was 6.67 hours of sleep, and the mean for the night shift was 5.95 hours, showing that the people on the day shift tended to get more sleep for these samples. **b.** *Step 1:* H_0: $\mu_{day} = \mu_{night}$, H_a: $\mu_{day} \neq \mu_{night}$, where μ is the mean number of hours of sleep per night. *Step 2:* Two-sample t-test: large samples and assume random, $\alpha = 0.05$. *Step 3:* $t = -2.71$, p-value $= 0.0079$, or 0.008. *Step 4:* Reject the null hypothesis. The difference in means is significant. **c.** It would not capture 0, showing a significant difference in means, because that is what we found in part b.

9.93 *Step 1:* H_0: $\mu_{Dem} = \mu_{Rep}$, H_a: $\mu_{Dem} > \mu_{Rep}$. *Step 2:* Two-sample t-test: random and Normal, $\alpha = 0.05$. *Step 3:* $t = 1.57$ or -1.57, p-value $= 0.072$. *Step 4:* Do not reject H_0. The means are not significantly different. **b.** *Step 1:* H_0: $\mu_{Dem} = \mu_{Rep}$, H_a: $\mu_{Dem} > \mu_{Rep}$. *Step 2:* Two-sample t-test: random and Normal, $\alpha = 0.05$. *Step 3:* $t = 2.28$ or -2.28, p-value $= 0.016$. *Step 4:* Reject H_0. The mean for the Democrats is significantly higher than the mean for the Republicans. **c.** With an increase in sample size (more data), the t-value increased and the p-value decreased, allowing us to reject the null hypothesis.

9.95 a. *Step 1:* H_0: $\mu_{7\text{-}Eleven} = \mu_{Vons}$, H_a: $\mu_{7\text{-}Eleven} > \mu_{Vons}$. *Step 2:* Paired t-test: assume random and Normal, $\alpha = 0.05$. *Step 3:* $t = 2.17$, p-value $= 0.033$. *Step 4:* Reject H_0. The mean price at 7-Eleven is significantly more than the mean price at Vons. **b.** *Step 1:* H_0: $\mu_{7\text{-}Eleven} = \mu_{Vons}$, H_a: $\mu_{7\text{-}Eleven} > \mu_{Vons}$. *Step 2:* Two-sample t-test (although not appropriate): assume random and Normal, $\alpha = 0.05$. *Step 3:* $t = 0.57$ or -0.57, p-value $= 0.289$. *Step 4:* Do not reject H_0. The mean price at 7-Eleven is not significantly more than the mean price at Vons when using the two-sample t-test. **c.** We found a larger t-value and a smaller p-value for the appropriate paired t-test. This occurs because finding the differences (for the paired t-test) reduces the variation, making the denominator of t smaller and so t is larger.

9.97 The table shows the results. The average of s^2 in the table is 2.8889 (or about 2.89), and if you take the square root, you get about 1.6997 (or about 1.70), which is the value for sigma (σ) given in the TI-84 output shown in the exercise. This demonstrates that s^2 is an unbiased estimator of σ^2, sigma squared.

Sample	s	s^2
1, 1	0	0
1, 2	0.7071	0.5
1, 5	2.8284	8.0
2, 1	0.7071	0.5
2, 2	0	0
2, 5	2.1213	4.5
5, 1	2.8284	8.0
5, 2	2.1213	4.5
5, 5	0	0
		Sum 26.0

$26/9 = 2.8889$

9.99 Answers will vary.

CHAPTER 10

Answers may vary slightly due to type of technology or rounding.

Section 10.1

10.1 a. Proportions are used for categorical data. **b.** Chi-square tests are used for categorical data.

10.3

	Boys	Girls
Stuffed Toys	8	12
Mechanical Toys	14	11

The table may have a different orientation.

10.5 *Mean Salary*: numerical and continuous. *Department*: categorical

10.7 a.

	Educated Parents	Uneducated Parents	Total
Studying	42	23	65
Not studying	14	8	22
Total	56	31	87

b. $65/87 = 74.7\%$ **c.** $0.747126(56) = 41.84$

d. Expected counts are shown in the table.

	Educated Parents	Uneducated Parents	Total
Studying	41.84	23.16	65
Not studying	14.16	7.84	22
Total	56	31	87

10.9

$$X^2 = \frac{(42 - 41.82)^2}{41.84} + \frac{(23 - 23.16)^2}{23.16} + \frac{(14 - 14.16)^2}{14.16} + \frac{(8 - 7.84)^2}{7.84}$$
$$= 0.0006 + 0.0011 + 0.0018 + 0.0033 = 0.0068$$

From technology to avoid rounding: Chi-square $= 0.007$

Section 10.2

10.11 Independence: one sample.

10.13 Homogeneity: random assignment (to four groups).

10.15 There is no need to draw an inference from the data as they cover the entire population of crude oil producing countries and not a sample. The data are given in the form of rates (percentages), not frequencies (counts), and there is not enough information to convert these percentages to counts.

10.17 The answers follow the guided steps. *Step 1:* H_a: The variables *Relationship Status* and *Obesity* are not independent (are associated). *Step 2:* Chi-square test of independence (given). The smallest expected count is 108.54, which is much more than 5, $\alpha = 0.05$. *Step 3:* $X^2 = 30.83$, p-value < 0.001. *Step 4:* Reject H_0. There is a connection between obesity and marital status; they are not independent. However, we should not generalize. Causality? No, it is an observational study. Percentage Obese: Dating, $81/440 = 18.4\%$; Cohabiting, $103/429 = 24.0\%$; Married, $147/424 = 34.7\%$.

10.19 *Step 1:* H_0: For men, watching violent TV is independent of abusiveness, H_a: For men, watching violent TV is not independent of abusiveness.
Step 2: Chi-square test of independence: one sample, the smallest expected count is $8.1 > 5$, sample not random, $\alpha = 0.05$. *Step 3:* Chi-square $= 5.02$, p-value $= 0.025$. *Step 4:* Reject H_0: High TV violence as a child is associated with abusiveness as an adult in men, but don't generalize to all males and don't conclude causality.

10.21 a. Independence: one sample with two variables. **b.** *Step 1:* H_0: Gender and higher studies are independent. H_a: Gender and higher studies are associated (not independent). *Step 2:* Chi-square test of independence: random sample, the smallest expected count is $69.1 > 5$, $\alpha = 0.05$. *Step 3:* Chi-square $= 2.41$, p-value $= 0.300$. *Step 4*: Reject H_0. Gender and higher studies have been shown to be associated. **c.** The level of higher studies has been found to be significantly different for males and females.

10.23 a. HS Grad rate for no preschool: 29/64, or 45.3%. The preschool kids had a higher graduation rate. **b.** *Step 1:* H_0: Graduation and preschool are independent, H_a: Graduation and preschool are not independent (they are associated). *Step 2:* Chi-square test of homogeneity: random assignment, not a random sample, the smallest expected count is $25.91 > 5$, $\alpha = 0.05$. *Step 3:* $X^2 = 4.67$, p-value $= 0.031$. *Step 4:* Reject H_0. Graduation and preschool are associated; causality, yes; generalization, no.

10.25 a. For preschool, 50% graduated, and for no preschool, $21/39 = 53.8\%$ graduated. It is surprising to see that the boys who did not go to preschool had a bit higher graduation rate. **b.** *Step 1:* H_0: For the boys, graduation and preschool are independent, H_a: For the boys, graduation and preschool are associated. *Step 2:* Chi-square test for homogeneity: random assignment, not a random sample, the smallest expected count is $15.32 > 5$, $\alpha = 0.05$. *Step 3:* $X^2 = 0.10$, p-value $= 0.747$. *Step 4:* Do not reject H_0. For the boys, there is no evidence that attending preschool is associated with graduating from high school. **c.** The results do not generalize to other groups of boys and girls, but what evidence we have suggests that although preschool might be effective for girls, it may not be for boys, at least with regard to graduation from high school.

10.27 a. 0 in the control group, $0.75(20) = 15$ in the gastric bypass group, and $0.95(20) = 19$ in the biliopancreatic diversion group were free from diabetes after two years.

b.

	Control	Gastric	Bilio
Free	0	15	19
Not Free	20	5	1

c. *Step 1*: H_0: Form of treatment and whether the patient becomes free from diabetes are independent, H_a: Form of treatment and whether the patient becomes free from diabetes are not independent. *Step 2*: Chi-square test for homogeneity: random assignment, not random sample, smallest expected count is $8.67 > 5$, $\alpha = 0.05$. *Step 3*: Chi-square $= 40.86$, p-value < 0.001. *Step 4*: Reject H_0. Treatment and result are not independent; they are associated. The treatment causes the result, but do not generalize.

10.29 a. With no confederate, $6/18$ (33.3%) followed the directions and took the stairs. With a compliant confederate, $16/18$ (88.9%) followed directions. With a noncompliant confederate, $5/18$ (27.8%) followed directions. Thus the subjects tended to do the same thing as the confederate. **b.** A p-value can never be larger than 1. The p-value is about 2.7 times 10 to the negative fourth power, or 0.00027, which is less than 0.001. **c.** *Step 1:* H_0: Treatment and compliance are independent, H_a: Treatment and compliance are not independent. *Step 2:* Chi-square test of homogeneity: random assignment, not a random sample, expected counts are all $9 > 5$, $\alpha = 0.05$. *Step 3:* $X^2 = 16.44$, p-value < 0.001. *Step 4:* Reject H_0. There is a significant effect; causality, yes; generalization, no. This shows an association between treatment and behavior at the elevator.

10.31 a. NA, $2/20 = 10\%$ frogs with improved jumping performance; TRS, $15/20 = 5\%$ frogs with improved jumping performance. Thus, there is an improvement in the jumping performance of frogs that were kept in temperature-regulated structures.

b.

	NA	TRS
Frogs with improved jumping performance	2	15
Frogs with no improvement in jumping performance	18	5

c. *Step 1:* H_0: The jumping performance is not associated with temperature conditions. H_a: The jumping performance is associated with temperature conditions. *Step 2:* Chi-square test for homogeneity: random assignment, the smallest expected count is $8.5 > 5$. $\alpha = 0.05$. *Step 3:* $X^2 = 17.29$, p-value $=$ 0.000. *Step 4:* Reject H_0. Temperature affects jumping performance in frogs.

Section 10.3

10.33 a. No, we cannot generalize, because this was not a random sample. **b.** Yes, we can infer causality because of random assignment.

10.35 a. You can generalize to other people admitted to this hospital who would have been assigned a double room because of the random sampling from that group. **b.** Yes, you can infer causality because of the random assignment.

10.37 a. The treatment variable is whether the patient received EL or the placebo. The response variable is whether the patient avoided the need for a platelet transfusion. **b.** It was a controlled experiment, as you can see from the random assignment. **c.** Yes, 72% of those on the drug avoided platelet transfusions, and that was better than the 19% on the placebo who avoided them. **d.** You can reject the hypothesis that the treatments and outcomes are independent. It shows that EL had a significant effect. **e.** Yes, you can infer causality (EL reduces the chance . . .) because the study was randomized and placebo-controlled, and there was a significant effect.

10.39 Because of the usage of the words "tend to exhibit," it is suspected that the conclusion is likely to be the result of an observational study, from which causality cannot be inferred.

10.41 No. The study was probably observational, and we cannot infer causality from one observational study.

10.43 Randomly assign about half the women to an iron supplement and half to a placebo. You could flip a coin for each woman: Heads she gets the iron, and tails she gets the placebo. Study and compare the death rates over one or two years.

10.45 a. The treatment variable is the drug: 5 mg of drug, 10 mg of drug, or placebo. The response variable is whether the patient experienced a 20% improvement of symptoms in three months. **b.** It was a controlled experiment, as you can see by the random assignment. **c.** Yes, the rate of 20% improvement was highest at the higher level of drug and lowest for those who took the placebo. **d.** The small p-values show that the percentage of patients who improved with tofacitinib is higher than the percentage who improved with the placebo, and the difference in percentages did not occur by chance. **e.** Yes, you can conclude that the use of tofacitinib increases the chances of a 20% improvement in symptoms, because this is a well-designed experiment with random assignment. The fact that the higher dose provided a larger chance of improvement adds to the evidence that the drug is effective.

10.47 **a.** Take a nonrandom sample of students and randomly assign some to the reception and some to attend a "control group" meeting where they do something else (such as learn the history of the college). **b.** Take a random sample of students and offer them the choice of attending the reception or attending a "control group" meeting where they do something else (such as learn the history of the college). **c.** Take a random sample of students. Then randomly assign some of the students in this sample to the reception and some to the "control group" meeting.

10.49 The answers follow the steps shown in the Guided Exercises.
1: Is the new drug better than a placebo with regard to worsening of asthma?
2: Yes, the new drug is significantly better.
3: This was a controlled experiment because of the random assignment.
4: Yes. Mentioning that dupilumab therapy, as compared with placebo, was associated with fewer asthma exacerbations is acceptable, but a stronger statement inferring causality could have been made. For example, "Dupilumab therapy caused fewer asthma exacerbations than the placebo."
5: Because there was no random sampling from the population, we cannot generalize widely, and the results apply only to these patients.
6: There was no mention of other articles.

10.51 Ten percent of the tests would be wrong (assuming none of the groups were different from the others). Because $0.10 \times 10 = 1$, you would expect 1 test out of 10 to appear significant just by chance.

Chapter Review Exercises

10.53 **a.** $31/65$, or 47.7%, of those in the control group were arrested, and $8/58$, or 13.8%, of those who attended preschool were arrested. Thus there was a lower rate of arrest for those who went to preschool.

b.

	Preschool	No Preschool
Arrest	8	31
No Arrest	50	34

Step 1: H_0: The treatment and arrest rate are independent, H_a: The treatment and arrest rate are associated. *Step 2:* Chi-square for homogeneity: random assignment, not a random sample, the smallest expected count is $18.39 > 5$, $\alpha = 0.05$. Step 3: Chi-square $= 16.27$, p-value $= 0.000055$ (or < 0.001). Step 4: We can reject the hypothesis of no association at the 0.05 level. Don't generalize. We conclude that preschool attendance affects the arrest rate. **c.** Two-proportion z-test. *Step 1*: H_0: $p_{pre} = p_{nopre}$, H_a: $p_{pre} < p_{nopre}$ (p is the rate of arrest). *Step 2:* Two-proportion z-test: the smallest expected count is $18.39 > 10$, $\alpha = 0.05$. *Step 3:* $z = 4.03$, p-value $= 0.000028$ (or < 0.001). *Step 4:* Reject the null hypothesis. Preschool lowers the rate of arrest, but we cannot generalize. **d.** The z-test enables us to test the alternative hypothesis that preschool attendance *lowers* the risk of later arrest. The Chi-square test allows for testing for some sort of association, but we can't specify whether it is a positive or a negative association. Note that the p-value for the one-sided hypothesis with the z-test is half the p-value for the two-sided hypothesis with the Chi-square test.

10.55 The data are percentages (not counts), and we cannot convert them to counts without knowing the total number of live births each year.

10.57 **a.** The death rate before the vaccine was 18.1 deaths per 100,000 children. After the vaccine, the death rate fell to 11.8 deaths per 100,000 children. The difference is 6.3 fewer deaths per 100,000 children after the vaccine was introduced. The small p-value (less than 0.001) means we can reject the null hypothesis that the death rate was unchanged and conclude that the death rate decreased. **b.** Although there are many indications that the vaccine is effective, this was not a randomized study. We cannot rule out the possibility that a confounding variable, not the vaccine, caused the decrease in death rates. (For example, because the comparison was done using different years, a difference in weather might have contributed to the difference in disease rates.)

10.59 **a.** $200/347 = 57.6\%$ of the surgery group died, and $247/348 = 71.0\%$ of the watchful waiting group died. Thus the surgery group did better (with regard to death) in the sample. **b.** $63/347 = 18.2\%$ of the surgery group died from prostate cancer, and $99/348 = 28.4\%$ of the watchful waiting group died from prostate cancer. So again the surgery group did better with regard to death from prostate cancer. **c.** It was a controlled experiment, as you can see from the random assignment.

This page intentionally left blank

Appendix D: Credits

PHOTO CREDITS

Cover Oksana Shufrych/Shutterstock
Page vi (top) Mansour Bethoney/Pearson Education, Inc., (middle) Colleen Ryan, (bottom) Rebecca Wong; **Page xviii** CarpathianPrince/Shutterstock.

Chapter 1

Page 26 Mark III Photonics/Shutterstock; **Page 27** Darren Baker/Fotolia; **Page 30** (top) Chad Bontrager/Shutterstock, (bottom) NASA; **Page 31** S. Kuelcue/Shutterstock; **Page 33** Tiler84/Fotolia; **Page 36** Laborant/Shutterstock; **Page 38** Art_zzz/Fotolia; **Page 42** Netsuthep/Fotolia; **Page 45** ImageryMajestic/Shutterstock; **Page 46** Coprid/Shutterstock; **Page 47** Kesu/Shutterstock; **Page 49** (background) Ase/Shutterstock, (binoculars) Evgeny Karandaev/Shutterstock, (right) Monkey Business/Shutterstock. (Note: Background and binoculars photos also appear on pages 85, 142, 201, 257, 307, 361, 411, 476, and 527.)

Chapter 2

Page 60 Denis Vrublevski/Shutterstock; **Page 61** Harvard College Library; **Page 66** Junial Enterprises/Shutterstock; **Page 70** Pashin Georgiy/Shutterstock; **Page 71** Filmfoto/Shutterstock; **Page 73** (top) George Doyle & Ciaran Griffin/Stockbyte/Getty Images, (bottom) Iofoto/Shutterstock; **Page 74** Luis Louro/Shutterstock; **Page 75** RedGreen/Shutterstock; **Page 79** Stephen VanHorn/Shutterstock; **Page 81** Cla78/Shutterstock; **Page 85** Alamy.

Chapter 3

Page 106 Jiri Hera/Shutterstock; **Page 107** Andy Dean/Fotolia; **Page 110** Saurabh13/Shutterstock; **Page 111** Rafa Irusta/Shutterstock; **Page 116** El Greco/Shutterstock; **Page 120** Shutterstock; **Page 122** OLJ Studio/Shutterstock; **Page 126** BW Folsom/Shutterstock; **Page 129** Jon Delorey; **Page 131** Jessmine/Shutterstock; **Page 133** George Doyle & Ciaran Griffin/Stockbyte/Getty Images; **Page 134** Arka38/Shutterstock; **Page 137** Ilona Ignatova/Shutterstock; **Page 142** Pearson Education, Inc.

Chapter 4

Page 166 Steve Rosset/Shutterstock; **Page 167** Hempuli/Shutterstock; **Page 172** Birute Vijeikiene/Shutterstock; **Page 183** Joingate/Shutterstock; **Page 188** Creations/Shutterstock; **Page 190** R. Gino Santa Maria/Shutterstock; **Page 191** Vixit/Shutterstock; **Page 201** PCN Photography/Alamy.

Chapter 5

Page 228 Brian Jackson/Fotolia; **Page 231** Bilder/Shutterstock; **Page 235** Pearson Education, Inc.; **Page 237** Pearson Education, Inc.; **Page 239** Pearson Education, Inc.; **Page 241** (left) Koncz/Shutterstock, (right) HomeStudio/Shutterstock; **Page 244** Rannev/ Shutterstock; **Page 250** Shutterstock; **Page 251** Guzel Studio/Shutterstock; **Page 254** Gjermund/Shutterstock; **Page 257** (top left) Evgeny Karandaev/Shutterstock, (top right) Siri Stafford/Digital Vision/Getty Images, (center) Eric Isselee/Shutterstock, (bottom) Shutterstock.

Chapter 6

Page 272 Joyfull/Shutterstock; **Page 273** Jupiterimages/Photos.com/Getty Images; **Page 275** Studiotouch/Shutterstock; **Page 277** Shutterstock; **Page 279** Alaettin Yildirim/Shutterstock; **Page 284** Frank Greenaway/Dorling Kindersley Limited; **Page 289** Uros Jonic/Shutterstock; **Page 293** Pearson Education, Inc.; **Page 296** Micha Rosenwirth/Shutterstock; **Page 298** Nixx Photography/Shutterstock; **Page 304** Alekss/Fotolia; **Page 305** PaulPaladin/Shutterstock; **Page 307** Karin Hildebrand Lau/Shutterstock.

Chapter 7

Page 324 Chad McDermott/Shutterstock; **Page 325** Elena Yakusheva/Shutterstock; **Page 327** Rido/Fotolia; **Page 332** Orla/Shutterstock; **Page 333** (all) Pearson Education, Inc.; **Page 338** Pearson Education, Inc.; **Page 340** Pashin Georgiy/Shutterstock; **Page 345** Paulo Williams/Shutterstock; **Page 347** Seregam/Shutterstock; **Page 353** Imagewell/Shutterstock; **Page 357** Whatafoto/Shutterstock; **Page 360** Xuejun Ii/Fotolia; **Page 361** Justasc/Shutterstock.

Chapter 8

Page 378 Koi88/Fotolia; **Page 379** Robert Daly/OJO Images/Getty Images; **Page 380** Yellowj/Fotolia; **Page 382** Mipan/Fotolia; **Page 383** Helder Almeida/Shutterstock; **Page 386** Raywoo/Fotolia; **Page 387** Mircea Maties/Shutterstock; **Page 401** Julian Rovagnati/Shutterstock; **Page 399** Shotgun/Shutterstock; **Page 402** Ajt/Shutterstock; **Page 406** Jan Kaliciak/Shutterstock; **Page 408** Gabriele Maltinti/Fotolia; **Page 411** Nathan B Dappen/Shutterstock.

Chapter 9

Page 428 Todd Klassy/Shutterstock; **Page 429** Shutterstock; **Page 433** Mmaxer/Shutterstock; **Page 435** Rihardzz/Shutterstock; **Page 438** Sashkin/Shutterstock; **Page 444** Elnur/Shutterstock; **Page 447** John Baran/Alamy; **Page 448** Africa Studio/Shutterstock; **Page 451** Martin Allinger/Shutterstock; **Page 462** Africa Studio/Shutterstock; **Page 467** 4matic/Fotolia; **Page 469** Bajinda/Fotolia; **Page 474** Valentyn Volkov/Shutterstock; **Page 476** AVAVA/Shutterstock.

Chapter 10

Page 500 Ljansempoi/Shutterstock; **Page 501** Chris Curtis/Shutterstock; **Page 505** Robertopalace/Shutterstock; **Page 507** Pakhnyushcha/Shutterstock; **Page 510** Oleksiy Mark/Fotolia; **Page 512** Shutterstock; **Page 514** Science Magazine; **Page 519** Jim Barber/Shutterstock; **Page 522** Sunabesyou/Shutterstock; **Page 527** Jennifer Nickert/Shutterstock.

TEXT CREDITS

Texas Instruments graphing calculator screenshots courtesy of Texas Instruments. Data and screenshots from StatCrunch used by permission of StatCrunch. Screenshots from Minitab courtesy of Minitab Corporation. Screenshot from SOCR used by the permission of the Statistics Online Computational Resource, UCLA. XLSTAT screenshots courtesy of Addinsoft, Inc. Used with permission. All Rights Reserved. XLSTAT is a registered trademark of Addinsoft SARL.

Chapter 1

Pages 54–55 Excerpt from U.S. HOSPITALIZATIONS FOR PNEUMONIA AFTER A DECADE OF PNEUMOCOCCAL VACCINATION by M.R. Griffin et al. *The New England Journal of Medicine,* July 11, 2013.

Chapter 2

Page 83 Screenshot of President Obama's 2013 State of the Union speech. Copyright ©Brad Borevitz. Used with permission.

Chapter 7

Page 328 Excerpt from "1 in 5 New Survey Express Some Doubt About the Holocaust," *New York Times*, April 20, 1993; **Page 361** Based on an activity from Workshop Statistics, © 2004, Dr. Allan Rossman and Dr. Beth Chance, California Polytechnic State University.

Chapter 8

Page 381 Definition of *hypothesis*: Used by permission of Merriam Webster.

Chapter 10

Page 520 Excerpt from BOCEPREVIR FOR UNTREATED CHRONIC HCV GENOTYPE 1 INFECTION by Poordad et al. Copyright © 2011. Used by permission of The New England Journal of Medicine; **Page 521** Excerpt from TEN-YEAR EFFECTS OF THE ADVANCED COGNITIVE TRAINING FOR INDEPENDENT AND VITAL ELDERLY COGNITIVE TRAINING TRIAL ON COGNITION AND EVERYDAY FUNCTIONING IN OLDER ADULTS by Rebok et al. *Journal of American Geriatics Society*. Copyright © 2014. Used by the permission of Journal of American Geriatrics Society; **Page 535** Excerpt from ELTROMBOPAG BEFORE PROCEDURES IN PATIENTS WITH CIRRHOSIS AND THROMBOCYTOPENIA by Nezam Afdhal et al. *The New England Journal of Medicine*. Copyright © 2012. Published by The New England Journal of Medicine; **Page 535** Excerpt from EFFECT OF INHALED GLUCOCORTICOIDS IN CHILDHOOD ON ADULT HEIGHT by L. H. William Kelly et al. *The New England Journal of Medicine.* Copyright © 2012. Published by The New England Journal of Medicine; **Page 538** Prostate Cancer Treatment from RADICAL PROSTATECTOMY OR WATCHFUL WAITING IN EARLY PROSTATE CANCER by Anna Bill Axelson. Published by The New England Journal of Medicine; **Page 538** Exercise 10.60: LENALIDOMIDE AFTER STEM-CELL TRANSPLANTATION FOR MULTIPLE MYELOMA, by P. L. McCarthy et al. *The New England Journal of Medicine.* Copyright © 2012. Published by The New England Journal of Medicine.

This page intentionally left blank

Index

S

T

This page intentionally left blank

This page intentionally left blank

GLOBAL EDITION

For these Global Editions, the editorial team at Pearson has collaborated with educators across the world to address a wide range of subjects and requirements, equipping students with the best possible learning tools. This Global Edition preserves the cutting-edge approach and pedagogy of the original, but also features alterations, customization, and adaptation from the North American version.

Pearson Global Edition

ISBN-10 1-292-16122-1
ISBN-13 978-1-292-16122-8

90000>

9 781292 161228