Introductory STATISTICS

10TH EDITION

MyLab Revision with Tech Updates

Pearson

Introductory
STATISTICS

10TH EDITION

MyLab Revision with Tech Updates

Neil A. Weiss, Ph.D.
School of Mathematical and Statistical Sciences
Arizona State University

with Contributions by
Toni Garcia
George Mason University

Biographies by **Carol A. Weiss**

Pearson

On the cover: Pelicans are large, social water birds found primarily along coastlines. They are best known for their aerial dives and elastic throat pouch with which they catch fish. According to the National Geographic Society, pelicans live for about 10 to 25 years in the wild. On average, they are 5.8 ft long, have a wingspan of 10 ft, and weigh 30 lb. Although, in North America, the brown pelican is endangered, populations have somewhat recovered after decades of population decline from the deleterious effects of pesticides, such as DDT, on egg survival.

Cover photograph: Tirc83/Getty Images

Director, Portfolio Management: Deirdre Lynch
Portfolio Manager: Suzanna Bainbridge
Courseware Portfolio Management Assistant: Morgan Danna
Managing Producer: Karen Wernholm
Content Producer: Ron Hampton
Media Producer: Shana Siegmund
Manager, Content Development: Bob Carroll
Product Marketing Manager: Emily Ockay
Field Marketing Manager: Andrew Noble
Marketing Assistant: Shannon McCormack
Senior Author Support/Technology Specialist: Joe Vetere
Manager, Rights and Permissions: Gina Cheselka
Senior Procurement Specialist: Carol Melville
Cover Designer: Jenny Willingham; Pearson CSC
Text Design: Rokusek Design, Inc.
Production Coordination, Composition, and Illustrations: Pearson CSC

Acknowledgments of third-party content appear on page C-1, which constitutes an extension of this copyright page.

Library of Congress Cataloging-in-Publication Data

Weiss, N. A. (Neil A.)
Introductory statistics MyLab revision with tech updates / Neil A. Weiss, Arizona State University; Contributions by Toni Garcia biographies by Carol A. Weiss. – 10th edition.
pages cm
Includes index.
ISBN 978-0-13-516305-4
1. Statistics–Textbooks. I. Weiss, Carol A. II. Title.
QA276.12.W45 2016
519.5–dc23

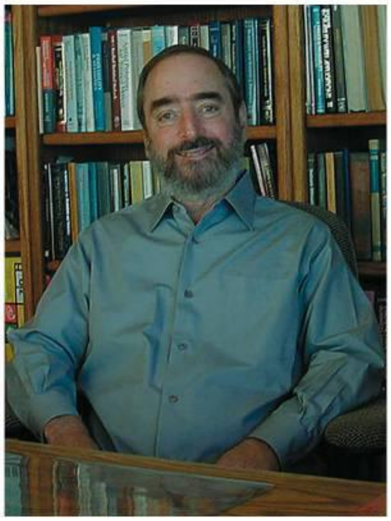
2014019105

Pearson www.pearson.com

ISBN 13: 9780135163054
ISBN 10: 0135163056

1 19

About the Author



The late Neil A. Weiss received his Ph.D. from UCLA and subsequently accepted an assistant professor position at Arizona State University (ASU), where he was ultimately promoted to the rank of full professor. Dr. Weiss taught statistics, probability, and mathematics—from the freshman level to the advanced graduate level—for more than 30 years.

In recognition of his excellence in teaching, Dr. Weiss received the *Dean's Quality Teaching Award* from the ASU College of Liberal Arts and Sciences. He also was runner-up twice for the *Charles Wexler Teaching Award* in the ASU School of Mathematical and Statistical Sciences. Dr. Weiss's comprehensive knowledge and experience ensures that his texts are mathematically and statistically accurate, as well as pedagogically sound.

In addition to his numerous research publications, Dr. Weiss was the author of *A Course in Probability* (Addison-Wesley, 2006). He also authored or coauthored books in finite mathematics, statistics, and real analysis. His texts—well known for their precision, readability, and pedagogical excellence—are used worldwide.

Dr. Weiss was a pioneer of the integration of statistical software into textbooks and the classroom, first providing such integration in the book *Introductory Statistics* (Addison-Wesley, 1982). Since this first publication, Neil Weiss's work continues to inspire future statisticians and current students of statistics, alike.

Dedicated to Aaron and Greg

Contents

*Indicates optional material.

PART III Probability, Random Variables, and Sampling Distributions

PART IV Inferential Statistics

*Indicates optional material.

*Indicates optional material.

**Indicates optional material on the WeissStats site.

PART V Regression, Correlation, and ANOVA

PART VI Multiple Regression and Model Building; Experimental Design and ANOVA**

*Indicates optional material.

**Indicates optional material on the WeissStats site.

Appendixes

WeissStats Resource Site (brief contents)

Note: **Visit the WeissStats Resource Site at www.pearsonhighered.com/weiss-stats for detailed contents.**

*indicates optional material

Preface

Using and understanding statistics and statistical procedures have become required skills in virtually every profession and academic discipline. The purpose of this book is to help students master basic statistical concepts and techniques and to provide real-life opportunities for applying them.

Audience

Introductory Statistics is intended for one- or two-semester courses or for quarter-system courses. Instructors can easily fit the text to the pace and depth they prefer. Introductory high school algebra is a sufficient prerequisite.

Although mathematically and statistically sound (the author has also written books at the senior and graduate levels), the approach does not require students to examine complex concepts. Rather, the material is presented in a natural and intuitive way. Simply stated, students will find this book's presentation of introductory statistics easy to understand.

About This Book

Introductory Statistics presents the fundamentals of statistics, featuring data production and data analysis. Data exploration is emphasized as an integral prelude to statistical inference.

This edition of *Introductory Statistics* continues the book's tradition of being on the cutting edge of statistical pedagogy, technology, and data analysis. It includes hundreds of new and updated exercises with real data from journals, magazines, newspapers, and websites.

The following Guidelines for Assessment and Instruction in Statistics Education (GAISE), funded and endorsed by the American Statistical Association, are supported and adhered to in *Introductory Statistics:*

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

Technology and Other Updates to the Tenth Edition

The book's technology coverage includes the use of Minitab, Excel, and the TI-83/84 Plus. Instructors can concentrate on one technology or cover and compare two or more technologies.

In addition to the MyLab updates that can be found on pp. xvi–xvii, all of the Technology Center features in this edition have been updated to reflect the latest software releases. **The Technology Center** is an in-text feature that includes step-by-step instructions for the implementation of each of these three applications.

This edition contains more than **3000 high-quality exercises**, which far exceeds what is found in typical introductory statistics books.

The **WeissStats Resource Site** (aka WeissStats site) provides an extensive array of resources for both instructors and students, including additional topics, applets, all data sets from the book in multiple formats, a procedures booklet, and technology appendixes. In addition to several new items, the site offers universal access to those items formerly included on the WeissStats CD. Refer to the table of contents for a brief list of the contents of the WeissStats site or visit the site at www.pearsonhighered.com/weiss-stats. *Note:* Resources for instructors only are available on the Instructor Resource Center at www.pearsonhighered.com/irc.

Chebyshev's Rule and the Empirical Rule. Section 3 of Chapter 3 is dedicated to an examination of Chebyshev's rule and the empirical rule. The empirical rule is further examined in Chapter 6 when the normal distribution is discussed.

Quartiles. The method for calculating quartiles has been modified to make it more easily accessible to students. Furthermore, a dedicated procedure that provides a step-by-step method for finding the quartiles of a data set has been included.

Distribution Shapes. The material on distribution shapes in Section 2.4 has been significantly modifiedand clarified.

Students will find this revised approach easier to understand and apply.

Regression Analysis. Major improvements have been made to the chapter on Descriptive Methods in Regression and Correlation. These improvements include a comprehensive discussion of scatterplots, a simpler introduction to the least-squares criterion, and easier introductory examples for the regression equation, the sums of squares and coefficient of determination, and the linear correlation coefficient.

Warm-up Exercises. Hundreds of "warm-up" exercises provide context-free problems that allow students to concentrate solely on the relevant concepts before moving on to applied exercises.

Density Curves. The discussion of density curves has been significantly expanded and now includes several examples and many more exercises.

Type II Error Probabilities and Power. Section 9.7, which covers Type II error probabilities and power, has undergone major revision, including increased visuals and the addition of procedures for calculating Type II error probabilities and for constructing power curves.

Note: See the Technology section of this preface for a discussion of technology additions, revisions, and improvements.

Hallmark Features and Approach

Chapter-Opening Features. Each chapter begins with a general description of the chapter, an explanation of how the chapter relates to the text as a whole, and a chapter outline. A classic or contemporary case study highlights the real-world relevance of the material.

End-of-Chapter Features. Each chapter ends with features that are useful for review, summary, and further practice.

- *Chapter Reviews.* Each chapter review includes *chapter objectives*, a list of *key terms* with page references, and *review problems* to help students review and study the chapter. Items related to optional materials are marked with asterisks, unless the entire chapter is optional.
- *Focusing on Data Analysis.* This feature lets students work with large data sets, practice technology use, and discover the many methods of exploring and analyzing data. For details, see the introductory Focusing on Data Analysis section on page 34 of Chapter 1.
- *Case Study Discussion.* At the end of each chapter, the chapter-opening case study is reviewed and discussed in light of the chapter's major points, and then problems are presented for students to solve.
- *Biographical Sketches.* Each chapter ends with a brief biography of a famous statistician. Besides being of general interest, these biographies teach students about the development of the science of statistics.

Formula/Table Card. The book's detachable formula/table card (FTC) contains all the formulas and many of the tables that appear in the text. The FTC is helpful for quick-reference purposes; many instructors also find it convenient for use with examinations.

Procedure Boxes, Index, and Booklet. To help students learn how to perform statistical analyses, easy-to-follow, step-by-step procedures have been provided. Each step is highlighted and presented again within the illustrating example. This approach shows how the procedure is applied and helps students master its steps. Additionally:

- A *Procedure Index* (located near the front of the book) provides a quick and easy way to find the right procedure for performing any statistical analysis.
- A *Procedures Booklet* (available in the Procedures Booklet section of the WeissStats Resource Site) provides a convenient way to access any required procedure.

ASA/MAA–Guidelines Compliant. *Introductory Statistics* follows American Statistical Association (ASA) and Mathematical Association of America (MAA) guidelines, which stress the interpretation of statistical results, the contemporary applications of statistics, and the importance of critical thinking.

Populations, Variables, and Data. Through the book's consistent and proper use of the terms *population, variable*, and *data*, statistical concepts are made clearer and more unified. This strategy is essential for the proper understanding of statistics.

Data Analysis and Exploration. Data analysis is emphasized, both for exploratory purposes and to check assumptions required for inference. Recognizing that not all readers have access to technology, the book provides ample opportunity to analyze and explore data without the use of a computer or statistical calculator.

Parallel Critical-Value/*P*-Value Approaches. Through a parallel presentation, the book offers complete flexibility in the coverage of the critical-value and P-value approaches to hypothesis testing. Instructors can concentrate on either approach, or they can cover and compare both approaches. The dual procedures, which provide both the critical-value and P-value approaches to a hypothesis-testing method, are combined in a side-by-side, easy-to-use format.

Interpretations. This feature presents the meaning and significance of statistical results in everyday language and highlights the importance of interpreting answers and results.

You Try It! This feature, which follows most examples, allows students to immediately check their understanding by working a similar exercise.

What Does It Mean? This margin feature states in "plain English" the meanings of definitions, formulas, key facts, and some discussions—thus facilitating students' understanding of the formal language of statistics.

Examples and Exercises

Real-World Examples. Every concept discussed in the text is illustrated by at least one detailed example. Based on real-life situations, these examples are interesting as well as illustrative.

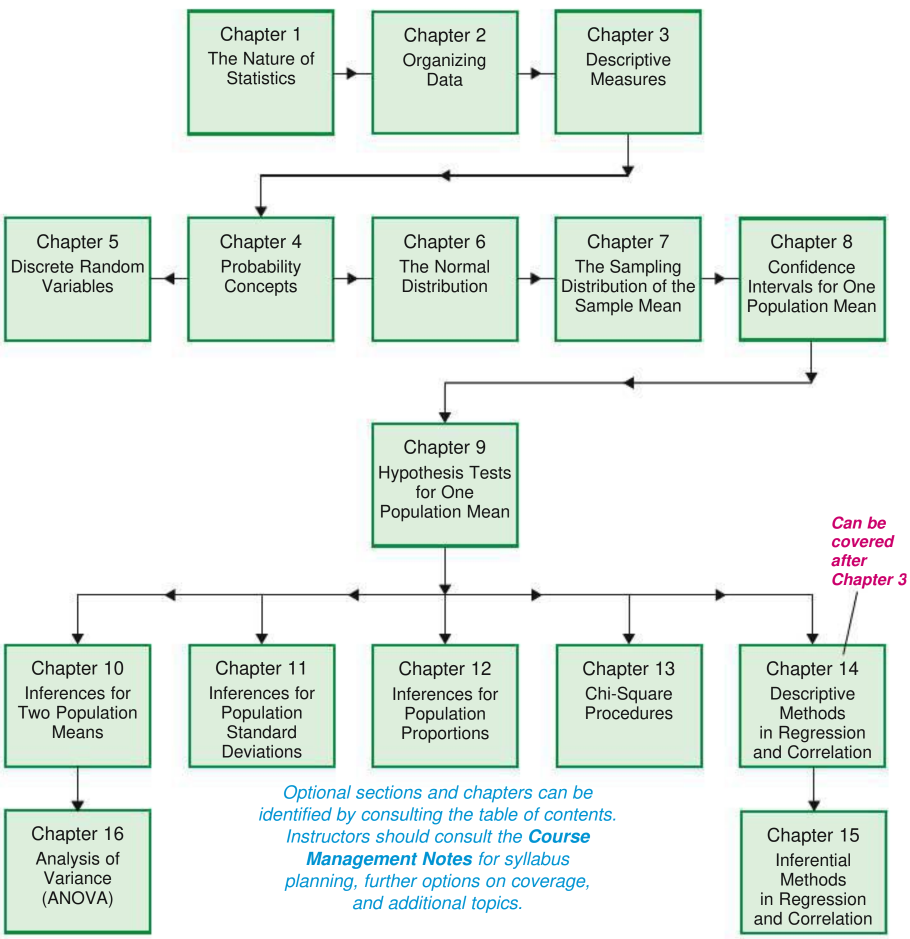
Real-World Exercises. Constructed from an extensive variety of articles in newspapers, magazines, statistical abstracts, journals, and websites, the exercises provide current, real-world applications whose sources are explicitly cited.
New to this edition, a fourth category of exercises has been added, namely, Applying the Concepts and Skills. As a consequence, the exercise sets are now divided into the following four categories:

- *Understanding the Concepts and Skills* exercises help students master the basic concepts and skills explicitly discussed in the section. These exercises consist of two types: (1) Non-computational problems that test student understanding of definitions, formulas, and key facts; (2) "warm-up" exercises, which require only simple computations and provide context-free problems that allow students to concentrate solely on the relevant concepts before moving on to applied exercises. For pedagogical reasons, it is recommended that warm-up exercises be done without the use of a statistical technology.
- *Applying the Concepts and Skills* exercises provide students with an extensive variety of applied problems that hone student skills with real-life data. These exercises can be done with or without the use of a statistical technology, at the instructor's discretion.
- *Working with Large Data Sets* exercises are intended to be done with a statistical technology and let students apply and interpret the computing and statistical capabilities of Minitab®, Excel®, the TI-83/84 Plus®, or any other statistical technology.
- *Extending the Concepts and Skills* exercises invite students to extend their skills by examining material not necessarily covered in the text. These exercises include many critical-thinking problems.

Notes: An exercise number set in cyan indicates that the exercise belongs to a group of exercises with common instructions. Also, exercises related to optional materials are marked with asterisks, unless the entire section is optional.

Organization

Introductory Statistics offers considerable flexibility in choosing material to cover. The following flowchart indicates different options by showing the interdependence among chapters; the prerequisites for a given chapter consist of all chapters that have a path that leads to that chapter.



Technology Resources

Technology Appendixes. The appendixes for Excel, Minitab, and the TI-83/84 Plus introduce the three statistical technologies, explain how to input data, and discuss how to perform other basic tasks. They are entitled *Getting Started with …* and are located in the Technology Basics section of the WeissStats Resource Site, www.pearsonhighered.com/weiss-stats.

Built-in Technology Manuals. The Technology Center features (in the book) and the technology appendixes (on the WeissStats site) make it unnecessary for students to purchase technology manuals. Students who will be using Minitab, Excel, or the TI-83/84 Plus to solve exercises should study the appropriate technology appendix(es) before commencing with the Technology Center sections.

TI Programs. The TI-83/84 Plus does not have built-in applications for a number of the statistical analyses discussed in the book. So that users of the TI-83/84 Plus can do such analyses with their calculators, the author has made available TI programs. Those programs are obtainable from the TI Programs section of the WeissStats Resource Site.

Acknowledgments

For this and the previous few editions of the book, it is our pleasure to thank the following reviewers, whose comments and suggestions resulted in significant improvements:

Olcay Akman, *Illinois State University*
James Albert, *Bowling Green State University*
John F. Beyers, II, *University of Maryland, University College*
David K. Britz, *Raritan Valley Community College*
Josef Brown, *New Mexico Tech*
Yvonne Brown, *Pima Community College*
Beth Chance, *California Polytechnic State University*
Brant Deppa, *Winona State University*
Carol DeVille, *Louisiana Tech University*
Jacqueline Fesq, *Raritan Valley Community College*
Robert Forsythe, *Frostburgh State University*
Richard Gilman, *Holy Cross College*
Donna Gorton, *Butler Community College*
David Groggel, *Miami University*
Joel Haack, *University of Northern Iowa*
Bernard Hall, *Newbury College*
Jessica Hartnett, *Gannon College*
Jane Harvill, *Baylor University*
Lance Hemlow, *Raritan Valley Community College*
Susan Herring, *Sonoma State University*
David Holmes, *The College of New Jersey*
Lorraine Hughes, *Mississippi State University*
Michael Hughes, *Miami University*
Satish Iyengar, *University of Pittsburgh*
Yvette Janecek, *Blinn College*
Jann-Huei Jinn, *Grand Valley State University*
Jeffrey Jones, *County College of Morris*
Thomas Kline, *University of Northern Iowa*
Lynn Kowski, *Raritan Valley Community College*
Christopher Lacke, *Rowan University*
Sheila Lawrence, *Rutgers University*
Tze-San Lee, *Western Illinois University*
Ennis Donice McCune, *Stephen F. Austin State University*
Jackie Miller, *The Ohio State University*
Luis F. Moreno, *Broome Community College*
Bernard J. Morzuch, *University of Massachusetts, Amherst*
Dennis M. O'Brien, *University of Wisconsin, La Crosse*
Dwight M. Olson, *John Carroll University*
Bonnie Oppenheimer, *Mississippi University for Women*
JoAnn Paderi, *Lourdes College*
Melissa Pedone, *Valencia Community College*
Alan Polansky, *Northern Illinois University*
Cathy D. Poliak, *Northern Illinois University*
Kimberley A. Polly, *Indiana University*
Geetha Ramachandran, *California State University*
B. Madhu Rao, *Bowling Green State University*
Gina F. Reed, *Gainesville College*
Steven E. Rigdon, *Southern Illinois University, Edwardsville*
Kevin M. Riordan, *South Suburban College*
Sharon Ross, *Georgia Perimeter College*
Edward Rothman, *University of Michigan*
Rina Santos, *College of Alameda*
George W. Schultz, *St. Petersburg College*
Arvind Shah, *University of South Alabama*
Sean Simpson, *Westchester Community College, SUNY*
Cid Srinivasan, *University of Kentucky, Lexington*
W. Ed Stephens, *McNeese State University*
Kathy Taylor, *Clackamas Community College*
Alane Tentoni, *Northwest Mississippi Community College*
Bill Vaughters, *Valencia Community College*
Roumen Vesselinov, *University of South Carolina*
Brani Vidakovic, *Georgia Institute of Technology*
Jackie Vogel, *Austin Peay State University*
Donald Waldman, *University of Colorado, Boulder*
Daniel Weiner, *Boston University*
Dawn White, *California State University, Bakersfield*
Marlene Will, *Spalding University*
Latrica Williams, *St. Petersburg College*
Matthew Wood, *University of Missouri, Columbia*
Nicholas A. Zaino Jr., *University of Rochester*

Our thanks are also extended to Joe Fred Gonzalez, Jr., for his many suggestions over the years for improving the book; and to Daniel Collins, Fuchun Huang, Charles Kaufman, Sharon Lohr, Richard Marchand, Shahrokh Parvini, Kathy Prewitt, Walter Reid, and Bill Steed, with whom we have had several illuminating consultations. Thanks also go to Matthew Hassett and Ronald Jacobowitz for their many helpful comments and suggestions.

Several other people provided useful input and resources. They include Thomas A. Ryan, Jr., Webster West, William Feldman, Frank Crosswhite, Lawrence W. Harding, Jr., George McManus, Greg Weiss, Jeanne Sholl, R. B. Campbell, Linda Holderman, Mia Stephens, Howard Blaut, Rick Hanna, Alison Stern-Dunyak, Dale Phibrick, Christine Sarris, and Maureen Quinn. Our sincere thanks go to all of them for their help in making this a better book.

Thanks to Larry Griffey for his formula/table card. Our gratitude also goes to Toni Garcia for writing the *Instructor's Solutions Manual* and the *Student's Solutions Manual*.

We express our appreciation to Dennis Young for his linear models modules and for his collaboration on numerous statistical and pedagogical issues. For checking the accuracy of the entire text and answers to the exercises, we extend our gratitude to Todd Hendricks and Susan Herring.

We are also grateful to David Lund and Patricia Lee for obtaining the database for the Focusing on Data Analysis sections. Our thanks are extended to the following people for their research in finding myriad interesting statistical studies and data for the examples, exercises, and case studies: Toni Garcia, Traci Gust, David Lund, Jelena Milovanovic, and Greg Weiss.

Many thanks go to Christine Stavrou and Stephanie Green for directing the development of the WeissStats Resource Site and to Cindy Scott, Carol Weiss, and Dennis Young for constructing the data files. Our appreciation also goes to our software editor, Bob Carroll.

We are grateful to Ron Hampton and Joe Vetere of Pearson Education, who coordinated the production of the book. We also thank our copyeditor, Bret Workman, and our proofreaders, Carol Weiss, Greg Weiss, Danielle Kortan, and Cindy Scott.

To Barbara Atkinson (Pearson Education) and Rokusek Design, Inc., we express our thanks for awesome interior and cover designs. Our sincere thanks also go to all the people at Pearson CSC for a terrific job of composition and illustration. We thank Aptara Corporation for photo research.

Without the help of many people at Pearson Education, this book and its numerous ancillaries would not have been possible; to all of them go our heartfelt thanks. In addition to the Pearson Education people mentioned above, we give special thanks to Deirdre Lynch, and to the following other people at Pearson Education: Suzanna Bainbridge, Ruth Berry, Emily Ockay, Erin Kelly, and Carol Melville.

Finally, we convey our appreciation to Carol A. Weiss. Apart from writing the text, she was involved in every aspect of development and production. Moreover, Carol did a superb job of researching and writing the biographies.

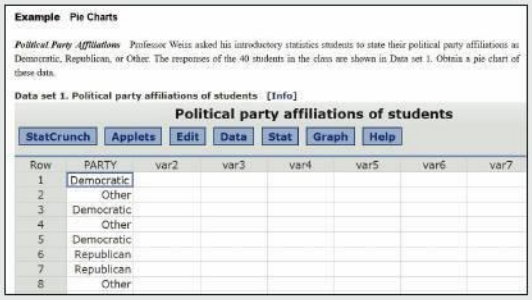
MyLab Statistics Revision for *Introductory Statistics,* 10e

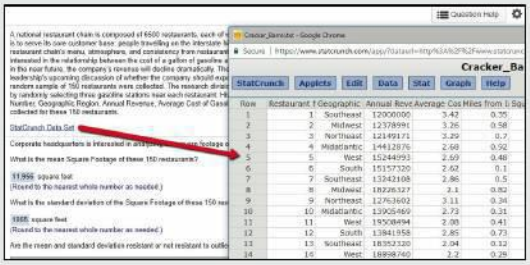
(access code required)

MyLab™ Statistics 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 MyLab Statistics course, making learning the material as seamless as possible. This updated version of the tenth edition includes a MyLab revision with the following enhancements:

StatCrunch Reports

StatCrunch Reports get students hands-on with statistical procedures by guiding them through real data analysis in StatCrunch. When results are generated with just a few clicks, students can spend more time interpreting and communicating results. StatCrunch Reports are integrated into the text and now include assignable questions in MyLab Statistics.



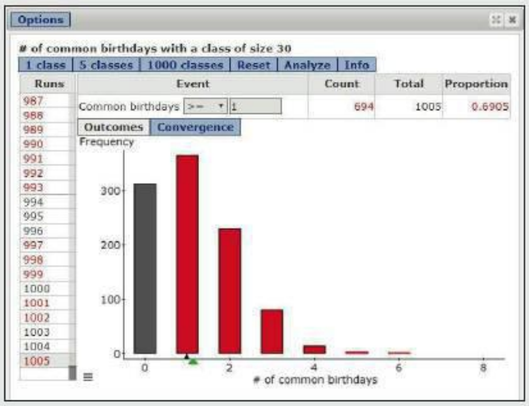


NEW! StatCrunch Projects

StatCrunch Projects provide opportunities for students to explore data beyond the classroom. In each project, students analyze a data set in StatCrunch® and answer assignable MyLab questions for immediate feedback. StatCrunch Projects span the entire curriculum or focus on certain key concepts. Questions from each project can also be assigned individually.

UPDATED! Interactive Applets

Interactive applets are a powerful tool for exploring statistical concepts and enhancing understanding. All text-specific applets for *Introductory Statistics* were updated and converted to HTML5 making them more user friendly. These applets join the existing suite of interactive applets that are available in StatCrunch. Now with assignable exercises.



pearson.com/mylab/statistics

Resources for Success

Pearson MyLab

Instructor Resources

Instructor's Edition

The Instructor's Edition contains answers to all text exercises.
ISBN-13: 978-0-13-518933-7; ISBN-10: 0-13-518933-0

Instructor's Solutions Manual

Written by Toni Garcia (George Mason University), the Instructor's Solutions Manual contains worked-out solutions to all text exercises. It can be downloaded from MyLab Statistics or from www.pearson.com.

PowerPoint Lecture Slides

Slides feature key graphics, concepts, and definitions from this text. These files can be downloaded from within MyLab Statistics or from www.pearson.com.

TestGen

TestGen® (www.pearson.com/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. TestGen is algorithmically based, allowing instructors to create multiple but equivalent versions of the same question or test with the click of a button. Instructors can also modify test bank questions or add new questions. The software and test bank are available for download from pearson.com. The questions are also assignable in MyLab Statistics.

Learning Catalytics

Now included in all MyLab Statistics courses, this student response tool uses students' smartphones, tablets, or laptops to engage them in more interactive tasks and thinking during lecture. Learning Catalytics™ fosters student engagement and peer-to-peer learning with real-time analytics. Access prebuilt exercises created specifically for statistics.

Question Libraries

In addition to StatCrunch Projects, MyLab Statistics also includes a Getting Ready for Statistics library that contains more than 450 exercises on prerequisite topics and a Conceptual Question Library with 1,000 questions that assess conceptual understanding.

Minitab and Minitab Express™

Bundling Minitab software with educational materials ensures students have access to the software they need in the classroom, around campus, and at home. And having 12-month access to Minitab and Minitab Express ensures students can use the software for the duration of their course.
ISBN 13: 978-0-13-445640-9 ISBN 10: 0-13-445640-8

JMP Student Edition

An easy-to-use, streamlined version of JMP desktop statistical discovery software from SAS Institute, Inc. is available for bundling with the text.
ISBN-13: 978-0-13-467979-2 ISBN-10: 0-13-467979-2

XLSTAT™

An Excel add-in that enhances the analytical capabilities of Excel. XLSTAT is used by leading businesses and universities around the world. It is available to bundle with this text. For more information go to www.pearsonhighered.com/xlstat/.
ISBN-13: 978-0-321-75932-0; ISBN-10: 0-321-75932-X

Student Resources

StatCrunch

StatCrunch® is powerful web-based statistical software that allows users to collect, crunch, and communicate with data. The vibrant online community offers tens of thousands of shared data sets for students and instructors to analyze, in addition to all of the data sets in the text or online homework. StatCrunch is integrated directly into MyLab Statistics or it can be purchased separately. Learn more at www.statcrunch.com.

Student's Solutions Manual

Written by Toni Garcia (George Mason University), this manual provides detailed, worked-out solutions to all odd-numbered section exercises and all Review Problems. It is available in print and can be downloaded from MyLab Statistics.
(ISBN-13: 978-0-32-198928-4; ISBN-10: 0-32-198928-7)

Data Sets

All data sets from the textbook are available in MyLab Statistics and at www.pearsonhighered.com/weiss-stats. They can be analyzed in StatCrunch or downloaded for use in other statistical software programs.

Statistical Software Support

Instructors and students can copy data sets from the text and MyLab Statistics exercises directly into software such as StatCrunch or Excel®. Students can also access instructional support tools including tutorial videos, Study Cards, and manuals for a variety of statistical software programs including, StatCrunch, Excel, Minitab®, JMP®, R, SPSS, and TI-83/84 calculators.

pearson.com/mylab/statistics

Data Sources

1stock1
A Handbook of Small Data Sets
A. C. Nielsen Company
AAA Foundation for Traffic Safety
AAMC Faculty Roster
AAUP Annual Report on the Economic Status of the Profession
ABC Global Kids Study
ABCNEWS Poll
ABCNews.com
About.com Pediatrics
Accident Facts
ACT High School Profile Report
ACT, Inc.
Acta Opthalmologica
AFI's 100 Years... 100 Movies — 10th Anniversary Edition
Agricultural Marketing Service
Agricultural Research Service
AHA Hospital Statistics
Air Travel Consumer Report
Alcohol Consumption and Related Problems: Alcohol and Health Monograph 1
All About Diabetes
Alliance for Cervical Cancer Protection
Alzheimer's Care Quarterly
American Association of University Professors
American Community Survey
American Council of Life Insurers
American Demographics
American Diabetes Association
American Express Retail Index
American Film Institute
American Hospital Association
American Hospital Association Annual Survey
American Housing Survey for the United States
American Industrial Hygiene Association Journal
American Journal of Applied Sciences
American Journal of Clinical Nutrition
American Journal of Obstetrics and Gynecology
American Journal of Physical Anthropology
American Journal of Political Science
American Laboratory
American Scientist
American Statistical Association
American Statistician
American Veterinary Medical Association
American Wedding Study
America's Families and Living Arrangements
America's Network Telecom Investor Supplement
Amstat News
Amusement Business
Analytical Chemistry
Analytical Services Division Transport Statistics
Animal Action Report
Animal Behaviour
Annals of Epidemiology
Anthropometric Reference Data for Children and Adults
Appetite
Applied Psychology in Criminal Justice
Aquaculture
Aquatic Biology
Arbitron
Archives of Physical Medicine and Rehabilitation
Arizona Chapter of the American Lung Association
Arizona Department of Revenue
Arizona Republic
Arizona Residential Property Valuation System
Arizona State University
Arizona State University Enrollment Summary
Arthritis Today
Asian Import
Associated Newspapers Ltd
Associated Press
Association of American Medical Colleges
Association of American Universities
Atlantic Oceanographic & Meteorological Laboratory
Atlantic Hurricane Database
Auckland University of Technology
Augusta National Golf Club
Australian Journal of Rural Health
Australian Journal of Zoology
Auto Trader
Avis Rent-A-Car
Baltimore Ravens
BARRON'S
Baseball Almanac
BBC News Magazine
Beachbody, LLC
Beer Institute Annual Report
Behavior Research Center
Behavioral Ecology and Sociobiology
Behavioral Risk Factor Surveillance System Summary Prevalence Report
Behavioural Pharmacology
Bell Systems Technical Journal
Biofuel Transportation Database
Biological Conservation
Biology of Sex Differences
Biometrics
Biometrika
BioScience
Board of Governors of the Federal Reserve System
Boston Athletic Association
Boston Globe
Box Office Mojo
Boyce Thompson Southwestern Arboretum
Brewer's Almanac
Brides
British Bankers' Association
British Journal of Educational Psychology
British Journal of Haematology
British Journal of Visual Impairment
British Medical Journal
Brokerage Report
Bureau of Crime Statistics and Research of Australia
Bureau of Economic Analysis
Bureau of Educational and Cultural Affairs
Bureau of Justice Statistics
Bureau of Justice Statistics Special Report
Bureau of Labor Statistics
Business Times
Buyers of New Cars
California Agriculture
California Wild: Natural Sciences for Thinking Animals
Car Shopping Trends Report
CareerBuilder
CBS News
Celebrity Net Worth
Cellular Telecommunications & Internet Association
Census of Agriculture
Centers for Disease Control and Prevention
Central Election Commission of the Russian Federation

Central Intelligence Agency
Chance
Characteristics of New Housing
Chemical & Pharmaceutical Bulletin
Chesapeake Biological Laboratory
Climates of the World
Climatography of the United States
Clinical Journal of Sports Medicine
Clinical Linguistics and Phonetics
CNBC
CNN/USA TODAY
Coleman & Associates, Inc.
College Board
College Entrance Examination Board
College-Bound Seniors Total Group Profile Report
Communications Industry Forecast & Report
Compendium of Federal Justice Statistics
Conde Nast Bridal Group
Congressional Directory
Consumer Expenditure Survey
Consumer Reports
Contributions to Boyce Thompson Institute
Controlling Road Rage: A Literature Review and Pilot Study
Crime in the United States
Criminal Justice and Behavior
CTIA–The Wireless Association
Current Housing Reports
Current Population Reports
Current Population Survey
Daily Mail
Daily Racing Form
Dallas Mavericks Roster
Data from the National Health Interview Survey
DataGenetics
Dave Leip's Atlas of U.S. Presidential Elections
Deep Sea Research Part I: Oceanographic Research Papers
Demographic Profiles
Demography
Department of Information Resources and Communications
Desert Samaritan Hospital
Dietary Guidelines for Americans
Dietary Reference Intakes
Digest of Education Statistics
Discover
Early Medieval Europe
Eastern Mediterranean Health Journal
Ecology
Economic Development Corporation Report
Edinburgh Medical and Surgical Journal
Edison Research
Edmunds.com
Educational Attainment in the United States
Educational Research
Educational Testing Service
eMarketer
Employment and Earnings
Energy Information Administration
Environmental Geology
ESPN
ESPN MLB Scoreboard
Estimates of School Statistics Database
Everyday Health Network
Experimental Agriculture
Experimental Brain Research
Family & Intimate Partner Violence Quarterly
Family Planning Perspectives
Federal Bureau of Investigation
Federal Highway Administration
Federal Highway Administration Annual Highway Statistics
Federal Reserve System
Federal Bureau of Prisons
Financial Planning
Fixed-Site Amusement Ride Injury Survey
FlightStats On-time Performance Report Summary
Florida Department of Environmental Protection
Florida State Center for Health Statistics
Food Consumption, Prices, and Expenditures
Footwear News
Forbes
Forest Mensuration
Fortune
Friends of the Earth
Fuel Economy Guide
Gallup
Gallup Poll
Gallup, Inc.
Geography
Georgia State University
Global Attractions Attendance Report
Global Index of Religiosity and Atheism
Golf Laboratories, Inc.
Golf.com
Governors' Political Affiliations & Terms of Office
Graduating Student and Alumni Survey
GRE Guide to the Use of Scores
Hanna Properties
Harris Interactive
Harris Poll
Harvard University
Heredity
Higher Education Research Institute
Highway Construction Safety and the Aging Driver
Highway Statistics
Hilton Hotels Corporation
Hirslanden Clinic
Historical Income Tables
HIV Surveillance Report
Hollywood Demographics
Homestyle Pizza
Hospital Statistics
HuffPost
HuffPost Style
Human Biology
Hydrobiologia
Income, Individual Income Tax Returns
Income, Poverty and Health Insurance Coverage in the United States
Indiewire
Industry Research
Infoplease
Information Please Almanac
Injury Facts
Inside MS
Institute of Medicine of the National Academy of Sciences
Insurance Institute for Highway Safety
Internal Revenue Service
International Association of Amusement Parks and Attractions.
International Classification of Diseases
International Communications Research
International Data Base
International Journal of Obesity
International Journal of Public Health
Iowa Agriculture Experiment Station
Iowa State University
Japan Automobile Manufacturer's Association
Japan Statistics Bureau
JAVMA News
Joint Committee on Printing
Journal of Abnormal Psychology
Journal of Advertising Research
Journal of American College Health
Journal of Anaesthesiology Clinical Pharmacology
Journal of Anatomy
Journal of Applied Behavioral Analysis
Journal of Applied Ecology
Journal of Applied Ichthyology
Journal of Applied Psychology
Journal of Applied Research in Higher Education
Journal of Applied Social Psychology
Journal of Applied Sport Psychology
Journal of Bone and Joint Surgery
Journal of Chemical Ecology
Journal of Child Nutrition and Management
Journal of Chronic Diseases
Journal of Clinical Endocrinology & Metabolism
Journal of Clinical Oncology
Journal of Early Adolescence
Journal of Environmental Psychology
Journal of Environmental Science and Health
Journal of Experimental Biology
Journal of Experimental Social Psychology
Journal of Family Violence
Journal of Forensic Identification
Journal of Gerontology Series A: Biological Sciences and Medical Sciences
Journal of Health, Population and Nutrition
Journal of Herpetology
Journal of Human Evolution
Journal of Mammalogy

Journal of Nursing and Healthcare of Chronic Illness
Journal of Nutrition
Journal of Organizational Behavior
Journal of Paleontology
Journal of Pediatrics
Journal of Pharmaceutical Sciences
Journal of Poverty & Social Justice
Journal of Pregnancy
Journal of Prosthetic Dentistry
Journal of Real Estate and Economics
Journal of Statistics Education
Journal of Sustainable Tourism
Journal of the American Academy of Child and Adolescent Psychiatry
Journal of the American Geriatrics Society
Journal of the American Medical Association
Journal of the American Public Health Association
Journal of the Royal Statistical Society
Journal of Tropical Ecology
Journal of Water Resources Planning and Management
Journal of Wildlife Management
Journal of Zoology, London
Journalism & Mass Communication Quarterly
Kelley Blue Book
Kelley Blue Book Company
Kennedy: The Classical Biography
Labor Force Statistics
Land Economics
Lawlink
Leonard Maltin Movie Guide
Life Expectancy at Birth
Life Insurers Fact Book
Limnology and Oceanography
Literary Digest
Los Angeles Dodgers
Los Angeles Times
Main Economic Indicators
Mammalia
Management
Manufactured Housing Statistics
Marine Ecology Progress Series
Marine Mammal Science
Market Survey of Long-Term Care Costs
Mayo Clinical Proceedings
Median Sales Price of Existing Single-Family Homes for Metropolitan Areas
Medical Biology and Etruscan Origins
Medical College of Wisconsin Eye Institute
Medical Principles and Practice
Medicine and Science in Sports & Exercise
Mega Millions
Mellman Group
Merck Manual
MetLife Mature Market Institute
Minitab Inc.
MLB.com
Money Stock Measures
Monitoring the Future
Monthly Labor Review
Monthly Tornado Statistics
Morningstar
Morrison Planetarium
Motor Vehicle Statistics of Japan
Motorcycle USA
National Aeronautics and Space Administration
National Agricultural Statistics Service
National Anti-Vivisection Society
National Association of Colleges and Employers
National Association of Realtors
National Association of State Racing Commissioners
National Basketball Association
National Cancer Institute
National Center for Education Statistics
National Center for Health Statistics
National Center on Addiction and Substance Abuse at Columbia University
National Collegiate Athletic Association
National Constitution Center
National Corrections Reporting Program
National Education Association
National Football League
National Geographic
National Geographic Society
National Golf Foundation
National Governors Association
National Health and Nutrition Examination Survey
National Health Interview Study
National Health Interview Survey
National Highway Traffic Safety Administration
National Household Travel Survey, Summary of Travel Trends
National Hurricane Center
National Institute of Aging
National Institute of Child Health and Human Development Neonatal Research Network
National Institute of Hygiene
National Institute of Mental Health
National Institute on Drug Abuse
National Interagency Fire Center
National Longitudinal Survey of Youth
National Low Income Housing Coalition
National Mortgage News
National Oceanic & Atmospheric Administration
National Safety Council
National Science Foundation
National Survey on Drug Use and Health
National Vital Statistics Reports
Nature
NCAA
NCAA.com
New England Journal of Medicine
New England Patriots Roster
New Scientist
New York Times
Newsweek
Newsweek, Inc
NewYork Times
Nielsen Media Research
Nielsen Report on Television
Nigerian Medical Journal
Nutrition
Obstetrics & Gynecology
Occupant Restraint Use
OECD in Figures
Office of Aviation Enforcement and Proceedings
Office of Immigration Statistics
Office of Justice Programs
Opinion Dynamics Poll
Opinion Research Corporation
Organisation for Economic Co-operation and Development
Origin of Species
Osteoporosis International
Out of Reach
Parade Magazine
Payless ShoeSource
Peacecorps.org
Pediatric Research
Pediatrics
Penn Schoen Berland
Pew Forum on Religion and Public Life
Pew Internet & American Life Project
Pew Research Center
PGA TOUR
Philosophical Magazine
Phoenix Gazette
Physician Specialty Data Book
PIN analysis
Player Roster
PLOS Biology
PLOS ONE
Pollstar
Popular Mechanics
Population-at-Risk Rates and Selected Crime Indicators
Preventative Medicine
pricewatch.com
Primetime Broadcast Programs
Prison Statistics
Proceedings of the 6th Berkeley Symposium on Mathematics and Statistics, VI
Proceedings of the American Zoo and Aquarium Association Nutrition Advisory Group
Proceedings of the National Academy of Science USA
Proceedings of the Royal Society of London
Professional Golfers' Association of America
Profile of Jail Inmates
Psychology of Addictive Behaviors
Pulse Opinion Research, LLC
Quality Engineering
Quinnipiac University
R. L. Polk & Co.

R. R. Bowker Company
Ranking of the States and Estimates of School Statistics
Rasmussen Reports
Recording Industry Association of America
Religious Landscape Survey
Reports
Research Quarterly for Exercise and Sport
Research Resources, Inc.
Residential Energy Consumption Survey: Consumption and Expenditures
Richard's Heating and Cooling
Robson Communities, Inc.
Roche
Roper Starch Worldwide, Inc.
Rubber Age
Runner's World
Salary Survey
Scarborough Research
Science
Science and Engineering Degrees
Science and Engineering Doctorate Awards
Science and Engineering Indicators
Science News
Scientific American
Scottish Executive
Selected Manpower Statistics
Semi-annual Wireless Survey
Sexually Transmitted Disease Surveillance
Significance Magazine
Smartphone Ownership
Sneak Previews
Snell, Perry and Associates
Soccer & Society
Social Forces
Social Indicators Research
South Carolina Budget and Control Board
South Carolina Statistical Abstract
Southwest Airlines
Sports Illustrated
Sports Illustrated Sites
SportsCenturyRetrospective
Stanford Revision of the Binet–Simon Intelligence Scale
Statistical Abstract of the United States
Statistical Report
Statistical Summary of Students and Staff
Statistical Yearbook
Statistics Norway
Statistics of Income, Individual Income Tax Returns
STATS
Status of the Profession
Stock Performance Guide
Stockholm Transit District
Storm Prediction Center
Summary of Travel Trends
Surveillance Epidemiology and End Results Fact Sheet
Survey of Consumer Finances
Survey of Current Business
Survey of Graduate Science Engineering Students and Postdoctorates
TalkBack Live
Teaching Issues and Experiments in Ecology
Technometrics
Television Bureau of Advertising
Tempe Daily News
Tesla Motors
Texas Comptroller of Public Accounts
The AMATYC Review
The American Freshman
The American Statistician
The Bowker Annual Library and Book Trade Almanac
The Business Journal
The Cross-Platform Report
The Design and Analysis of Factorial Experiments
The Earth: Structure, Composition and Evolution
The Geyser Observation and Study Association
The History of Statistics
The Infinite Dial
The Journal of Arachnology
The Lancet
The Lobster Almanac
The Marathon: Physiological, Medical, Epidemiological, and Psychological Studies
The Methods of Statistics
The Nielsen Company
The Open University
The Plant Cell
The Street
The Washington Post
The World Bank
Themed Entertainment Association
Thoroughbred Times
Time
Time Spent Viewing
Times Higher Education
TIMS
TNS Intersearch
Trade & Environment Database (TED) Case Studies
Trademark Reporter
Travel + Leisure Golf
Trends in Television
Tropical Biodiversity
Tropical Cyclone Report
TV Basics
TVbytheNumbers
U.S. Agency for International Development
U.S. Agricultural Trade Update
U.S. Bureau of Citizenship and Immigration Services
U.S. Bureau of Economic Analysis
U.S. Census Bureau
U.S. Coast Guard
U.S. Congress, Joint Committee on Printing
U.S. Department of Agriculture
U.S. Department of Commerce
U.S. Department of Defense
U.S. Department of Education
U.S. Department of Energy
U.S. Department of Health and Human Services
U.S. Department of Housing and Urban Development
U.S. Department of Justice
U.S. Energy Information Administration
U.S. Environmental Protection Agency
U.S. Federal Highway Administration
U.S. Geological Survey
U.S. National Center for Health Statistics
U.S. National Oceanic and Atmospheric Administration
U.S. Naturalizations
U.S. News and World Report
U.S. Overseas Loans and Grants
U.S. Postal Service
U.S. Substance Abuse and Mental Health Services Administration
U.S. Women's Open
Ultrasound in Medicine and Biology
Uniform Crime Reports
United States Pharmacopeia
University of Delaware
University of Helsinki
University of Malaysia
University of Maryland
University of Nevada, Las Vegas
Urban Studies
USA Today
Usability News
Utah Behavioral Risk Factor Surveillance System (BRFSS) Local Health District Report
Vegetarian Journal
Vegetarian Resource Group
VentureOne Corporation
Veronis Suhler Stevenson
Vital and Health Statistics
Vital Statistics of the United States
Wall Street Journal
Washington Post
Weatherwise
Weekly Retail Gasoline and Diesel Prices
Wichita Eagle
Wikipedia
WIN-Gallup International
Women and Cardiovascular Disease Hospitalizations
Women's Health Initiative
WONDER database
World Almanac
World FactBook
World Meteorological Association
World Radiation Center
World Series History
www.house.gov
Yahoo News
Yahoo! Contributor Network
Year-End Industry Shipment and Revenue Statistics
YouGov
Zillow.com
Zogby International

CHAPTER
1

The Nature of Statistics

CHAPTER OUTLINE

CHAPTER OBJECTIVES

What does the word *statistics* bring to mind? To most people, it suggests numerical facts or data, such as unemployment figures, farm prices, or the number of marriages and divorces. Two common definitions of the word *statistics* are as follows:

1. [used with a plural verb] facts or data, either numerical or nonnumerical, organized and summarized so as to provide useful and accessible information about a particular subject.

2. [used with a singular verb] the science of organizing and summarizing numerical or nonnumerical information.

Statisticians also analyze data for the purpose of making generalizations and decisions. For example, a political analyst can use data from a portion of the voting population to predict the political preferences of the entire voting population, or a city council can decide where to build a new airport runway based on environmental impact statements and demographic reports that include a variety of statistical data.

In this chapter, we introduce some basic terminology so that the various meanings of the word *statistics* will become clear to you. We also examine two primary ways of producing data, namely, through sampling and experimentation. We discuss sampling designs in Sections 1.2 and 1.3 and experimental designs in Section 1.4.

CASE STUDY

Top Films of All Time

Honoring the 10th anniversary of its award-winning series, the American Film Institute (AFI) again conducted a poll of 1500 film artists, critics, and historians, asking them to pick their 100 favorite films from a list of 400. The films on the list were made between 1915 and 2005.

After tallying the responses, AFI compiled a list representing the top 100 films. *Citizen Kane*, made in 1941, again finished in first place, followed by *The Godfather*, which was made in 1972. The following table shows the top 40 finishers in the poll. [SOURCE: Data from *AFI's 100 Years... 100 Movies — 10th Anniversary Edition*. Published by the American Film Institute.]

Rank	Film	Year	Rank	Film	Year
1	Citizen Kane	1941	21	Chinatown	1974
2	The Godfather	1972	22	Some Like It Hot	1959
3	Casablanca	1942	23	The Grapes of Wrath	1940
4	Raging Bull	1980	24	E.T. The Extra-Terrestrial	1982
5	Singin' in the Rain	1952	25	To Kill a Mockingbird	1962
6	Gone with the Wind	1939	26	Mr. Smith Goes to Washington	1939
7	Lawrence of Arabia	1962	27	High Noon	1952
8	Schindler's List	1993	28	All About Eve	1950
9	Vertigo	1958	29	Double Indemnity	1944
10	The Wizard of Oz	1939	30	Apocalypse Now	1979
11	City Lights	1931	31	The Maltese Falcon	1941
12	The Searchers	1956	32	The Godfather Part II	1974
13	Star Wars	1977	33	One Flew Over the Cuckoo's Nest	1975
14	Psycho	1960	34	Snow White and the Seven Dwarfs	1937
15	2001: A Space Odyssey	1968	35	Annie Hall	1977
16	Sunset Blvd.	1950	36	The Bridge on the River Kwai	1957
17	The Graduate	1967	37	The Best Years of Our Lives	1946
18	The General	1927	38	The Treasure of the Sierra Madre	1948
19	On the Waterfront	1954	39	Dr. Strangelove	1964
20	It's a Wonderful Life	1946	40	The Sound of Music	1965

Armed with the knowledge that you gain in this chapter, you will be asked to further analyze this AFI poll at the end of the chapter.

1.1 Statistics Basics

You probably already know something about statistics. If you read newspapers, surf the Web, watch the news on television, or follow sports, you see and hear the word *statistics* frequently. In this section, we use familiar examples such as baseball statistics and voter polls to introduce the two major types of statistics: **descriptive statistics** and **inferential statistics.** We also introduce terminology that helps differentiate among various types of statistical studies.

Descriptive Statistics

Each spring in the late 1940s, President Harry Truman officially opened the major league baseball season by throwing out the "first ball" at the opening game of the Washington Senators. We use the 1948 baseball season to illustrate the first major type of statistics, descriptive statistics.

EXAMPLE 1.1 Descriptive Statistics

The 1948 Baseball Season In 1948, the Washington Senators (Nationals) played 153 games, winning 56 and losing 97. They finished seventh in the American League and were led in hitting by Bud Stewart, whose batting average was .279. Baseball statisticians compiled these and many other statistics by organizing the complete records for each game of the season.

Although fans take baseball statistics for granted, much time and effort is required to gather and organize them. Moreover, without such statistics, baseball would be much harder to follow. For instance, imagine trying to select the best hitter in the American League given only the official score sheets for each game. (More than 600 games were played in 1948; the best hitter was Ted Williams, who led the league with a batting average of .369.)

The work of baseball statisticians is an illustration of *descriptive statistics*.

DEFINITION 1.1

Descriptive Statistics

Descriptive statistics consists of methods for organizing and summarizing information.

Descriptive statistics includes the construction of graphs, charts, and tables and the calculation of various descriptive measures such as averages, measures of variation, and percentiles. We discuss descriptive statistics in detail in Chapters 2 and 3.

Inferential Statistics

We use the 1948 presidential election to introduce the other major type of statistics, inferential statistics.

EXAMPLE 1.2 Inferential Statistics

The 1948 Presidential Election In the fall of 1948, President Truman was concerned about statistics. The Gallup Poll taken just prior to the election predicted that he would win only 44.5% of the vote and be defeated by the Republican nominee, Thomas E. Dewey. But the statisticians had predicted incorrectly. Truman won more than 49% of the vote and, with it, the presidency. The Gallup Organization modified some of its procedures and has correctly predicted the winner ever since.

Political polling provides an example of inferential statistics. Interviewing everyone of voting age in the United States on their voting preferences would be expensive and unrealistic. Statisticians who want to gauge the sentiment of the entire **population** of U.S. voters can afford to interview only a carefully chosen group of a few thousand voters. This group is called a **sample** of the population. Statisticians analyze the information obtained from a sample of the voting population to make inferences (draw conclusions) about the preferences of the entire voting population. Inferential statistics provides methods for drawing such conclusions.

The terminology just introduced in the context of political polling is used in general in statistics.

DEFINITION 1.2

Population and Sample

Population: The collection of all individuals or items under consideration in a statistical study.

Sample: That part of the population from which information is obtained.

Figure 1.1 on the following page depicts the relationship between a population and a sample from the population.

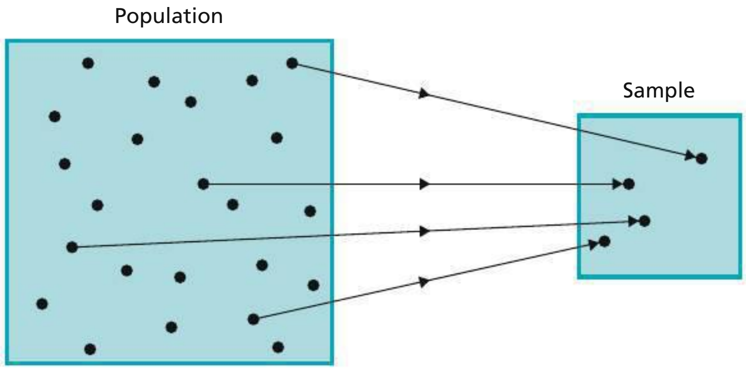
Now that we have discussed the terms *population* and *sample,* we can define *inferential statistics.*

DEFINITION 1.3

Inferential Statistics

Inferential statistics consists of methods for drawing and measuring the reliability of conclusions about a population based on information obtained from a sample of the population.

FIGURE 1.1
Relationship between population and sample



Descriptive statistics and inferential statistics are interrelated. You must almost always use techniques of descriptive statistics to organize and summarize the information obtained from a sample before carrying out an inferential analysis. Furthermore, as you will see, the preliminary descriptive analysis of a sample often reveals features that lead you to the choice of (or to a reconsideration of the choice of) the appropriate inferential method.

Classifying Statistical Studies

As you proceed through this book, you will obtain a thorough understanding of the principles of descriptive and inferential statistics. In this section, you will classify statistical studies as either descriptive or inferential. In doing so, you should consider the purpose of the statistical study.

If the purpose of the study is to examine and explore information for its own intrinsic interest only, the study is descriptive. However, if the information is obtained from a sample of a population and the purpose of the study is to use that information to draw conclusions about the population, the study is inferential.

Thus, a descriptive study may be performed either on a sample or on a population. Only when an inference is made about the population, based on information obtained from the sample, does the study become inferential.

Examples 1.3 and 1.4 further illustrate the distinction between descriptive and inferential studies. In each example, we present the result of a statistical study and classify the study as either descriptive or inferential. Classify each study yourself before reading our explanation.

EXAMPLE 1.3 Classifying Statistical Studies

The 1948 Presidential Election Table 1.1 displays the voting results for the 1948 presidential election.

TABLE 1.1
Final results of the 1948 presidential election

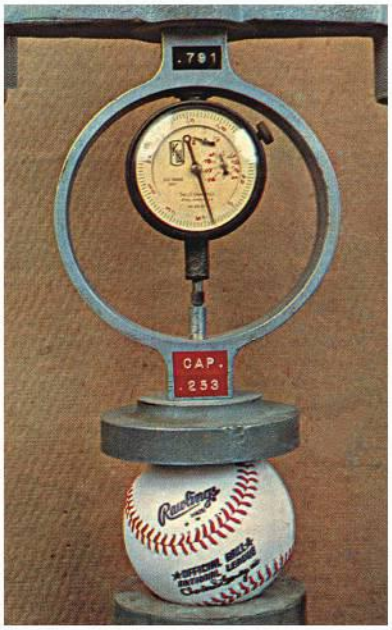
Ticket	Votes	Percentage
Truman–Barkley (Democratic)	24,179,345	49.7
Dewey–Warren (Republican)	21,991,291	45.2
Thurmond–Wright (States Rights)	1,176,125	2.4
Wallace–Taylor (Progressive)	1,157,326	2.4
Thomas–Smith (Socialist)	139,572	0.3

You try it!

Exercise 1.7 on page 7

Classification This study is descriptive. It is a summary of the votes cast by U.S. voters in the 1948 presidential election. No inferences are made.

EXAMPLE 1.4 Classifying Statistical Studies



Testing Baseballs For the 101 years preceding 1977, the major leagues purchased baseballs from the Spalding Company. In 1977, that company stopped manufacturing major league baseballs, and the major leagues then bought their baseballs from the Rawlings Company.

Early in the 1977 season, pitchers began to complain that the Rawlings ball was "livelier" than the Spalding ball. They claimed it was harder, bounced farther and faster, and gave hitters an unfair advantage. Indeed, in the first 616 games of 1977, 1033 home runs were hit, compared to only 762 home runs hit in the first 616 games of 1976.

Sports Illustrated magazine sponsored a study of the liveliness question and published the results in the article "They're Knocking the Stuffing Out of It" (*Sports Illustrated*, June 13, 1977, pp. 23–27) by L. Keith. In this study, an independent testing company randomly selected 85 baseballs from the current (1977) supplies of various major league teams. It measured the bounce, weight, and hardness of the chosen baseballs and compared these measurements with measurements obtained from similar tests on baseballs used in 1952, 1953, 1961, 1963, 1970, and 1973.

The conclusion was that "...the 1977 Rawlings ball is livelier than the 1976 Spalding, but not as lively as it could be under big league rules, or as the ball has been in the past."

Classification This study is inferential. The independent testing company used a sample of 85 baseballs from the 1977 supplies of major league teams to make an inference about the population of all such baseballs. (An estimated 360,000 baseballs were used by the major leagues in 1977.)

You try it!

Exercise 1.9 on page 7

The *Sports Illustrated* study also shows that it is often not feasible to obtain information for the entire population. Indeed, after the bounce and hardness tests, all of the baseballs sampled were taken to a butcher in Plainfield, New Jersey, to be sliced in half so that researchers could look inside them. Clearly, testing every baseball in this way would not have been practical.

The Development of Statistics

Historically, descriptive statistics appeared before inferential statistics. Censuses were taken as long ago as Roman times. Over the centuries, records of such things as births, deaths, marriages, and taxes led naturally to the development of descriptive statistics.

Inferential statistics is a newer arrival. Major developments began to occur with the research of Karl Pearson (1857–1936) and Ronald Fisher (1890–1962), who published their findings in the early years of the twentieth century. Since the work of Pearson and Fisher, inferential statistics has evolved rapidly and is now applied in a myriad of fields.

Familiarity with statistics will help you make sense of many things you read in newspapers and magazines and on the Internet. For instance, could the *Sports Illustrated* baseball test (Example 1.4), which used a sample of only 85 baseballs, legitimately draw a conclusion about 360,000 baseballs? After working through Chapter 9, you will understand why such inferences are reasonable.

What Does It Mean?

An understanding of statistical reasoning and of the basic concepts of descriptive and inferential statistics has become mandatory for virtually everyone, in both their private and professional lives.

Observational Studies and Designed Experiments

Besides classifying statistical studies as either descriptive or inferential, we often need to classify them as either *observational studies* or *designed experiments*. In an **observational study,** researchers simply observe characteristics and take measurements, as in a sample survey. In a **designed experiment,** researchers impose treatments and controls (discussed in Section 1.4) and then observe characteristics and take measurements. Observational studies can reveal only *association,* whereas designed experiments can help establish *causation.*

Note that, in an observational study, someone is observing data that already exist (i.e., the data were there and would be there whether someone was interested in them or not). In a designed experiment, however, the data do not exist until someone does something (the experiment) that produces the data. Examples 1.5 and 1.6 illustrate some major differences between observational studies and designed experiments.

EXAMPLE 1.5 An Observational Study

Vasectomies and Prostate Cancer Approximately 450,000 vasectomies are performed each year in the United States. In this surgical procedure for contraception, the tube carrying sperm from the testicles is cut and tied.

Several studies have been conducted to analyze the relationship between vasectomies and prostate cancer. The results of one such study by E. Giovannucci et al. appeared in the paper "A Retrospective Cohort Study of Vasectomy and Prostate Cancer in U.S. Men" (*Journal of the American Medical Association*, Vol. 269(7), pp. 878–882).

Dr. Giovannucci, study leader and epidemiologist at Harvard-affiliated Brigham and Women's Hospital, said that "…we found 113 cases of prostate cancer among 22,000 men who had a vasectomy. This compares to a rate of 70 cases per 22,000 among men who didn't have a vasectomy."

The study shows about a 60% elevated risk of prostate cancer for men who have had a vasectomy, thereby revealing an association between vasectomy and prostate cancer. But does it establish causation: that having a vasectomy causes an increased risk of prostate cancer?

You try it!
Exercise 1.19 on page 8

The answer is no, because the study was observational. The researchers simply observed two groups of men, one with vasectomies and the other without. Thus, although an association was established between vasectomy and prostate cancer, the association might be due to other factors (e.g., temperament) that make some men more likely to have vasectomies and also put them at greater risk of prostate cancer.

EXAMPLE 1.6 A Designed Experiment

Folic Acid and Birth Defects For several years, evidence had been mounting that folic acid reduces major birth defects. Drs. A. E. Czeizel and I. Dudas of the National Institute of Hygiene in Budapest directed a study that provided the strongest evidence to date. Their results were published in the paper "Prevention of the First Occurrence of Neural-Tube Defects by Periconceptional Vitamin Supplementation" (*New England Journal of Medicine*, Vol. 327(26), p. 1832).

For the study, the doctors enrolled 4753 women prior to conception and divided them randomly into two groups. One group took daily multivitamins containing 0.8 mg of folic acid, whereas the other group received only trace elements (minute amounts of copper, manganese, zinc, and vitamin C). A drastic reduction in the rate of major birth defects occurred among the women who took folic acid: 13 per 1000, as compared to 23 per 1000 for those women who did not take folic acid.

You try it!
Exercise 1.21 on page 8

In contrast to the observational study considered in Example 1.5, this is a designed experiment and does help establish causation. The researchers did not simply observe two groups of women but, instead, randomly assigned one group to take daily doses of folic acid and the other group to take only trace elements.

Exercises 1.1

Understanding the Concepts and Skills

1.1 Define the following terms:
a. Population
b. Sample

1.2 What are the two major types of statistics? Describe them in detail.

1.3 Identify some methods used in descriptive statistics.

1.4 Explain two ways in which descriptive statistics and inferential statistics are interrelated.

1.5 Define the following terms:
a. Observational study
b. Designed experiment

1.6 Fill in the following blank: Observational studies can reveal only association, whereas designed experiments can help establish ______.

Applying the Concepts and Skills

*In Exercises **1.7–1.12**, classify each of the studies as either descriptive or inferential. Explain your answers.*

1.7 TV Viewing Times. Data from a sample of Americans yielded the following estimates of average TV viewing time per month for all Americans 2 years old and older. The times are in hours and minutes; Q1 stands for first quarter. [SOURCE: *The Cross-Platform Report*, Quarter 1, 2011. Published by The Nielsen Company, © 2011.]

Viewing method	Q1 2011	Q1 2010	Change (%)
Watching TV in the home	158:47	158:25	0.2
Watching timeshifted TV	10:46	9:36	12.2
DVR playback	26:14	25:48	1.7
Using the Internet on a computer	25:33	25:54	−1.4
Watching video on the Internet	4:33	3:23	34.5
Mobile subscribers watching video on a mobile phone	4:20	3:37	20.0

1.8 Professional Athlete Salaries. From the *Statistical Abstract of the United States* and the article "Average Salaries in the NBA, NFL, MLB and NHL" by J. Dorish, published on the Yahoo! Contributor Network, we obtained the following data on average professional athletes' salaries for the years 2005 and 2011.

Sport	Average salary ($millions)	
	2005	2011
Baseball (MLB)	2.48	3.31
Basketball (NBA)	4.04	5.15
Football (NFL)	1.40	1.90

1.9 Home Sales. Zillow.com is an online database that provides real estate information for U.S. homes that are for rent or sale. It also presents statistics on recently sold homes. The following table gives various information on all homes sold in several different cities across the United States for the month of September 2012.

City	Price per square foot	Sale to list price ratio	% foreclosure re-sales
Scottsdale, AZ	$167	0.973	12.43%
Washington, DC	$436	0.990	2.88%
San Francisco, CA	$636	1.026	6.55%
Las Vegas, NV	$ 74	1.000	19.45%
Nashville, TN	$106	0.973	18.09%

1.10 Drug Use. The U.S. Substance Abuse and Mental Health Services Administration collects and publishes data on nonmedical drug use, by type of drug and age group, in *National Survey on Drug Use and Health*. The following table provides data for the years 2003 and 2008. The percentages shown are estimates for the entire nation based on information obtained from a sample (NA, not available).

Type of drug	Percentage, 18–25 years old			
	Ever used		Current user	
	2003	2008	2003	2008
Any illicit drug	60.5	56.6	20.3	19.6
Marijuana and hashish	53.9	50.4	17.0	16.5
Cocaine	15.0	14.4	2.2	1.5
Hallucinogens	23.3	17.7	1.7	1.7
Inhalants	14.9	10.4	0.4	0.3
Any psychotherapeutic	29.0	29.2	6.0	5.9
Alcohol	87.1	85.6	61.4	61.2
"Binge" alcohol use	NA	NA	41.6	41.8
Cigarettes	70.2	64.2	40.2	35.7
Smokeless tobacco	22.0	20.3	4.7	5.4
Cigars	45.2	41.4	11.4	11.3

1.11 Dow Jones Industrial Averages. From the *Stock Performance Guide*, published online by 1stock1 on the website 1Stock1.com, we found the closing values of the Dow Jones Industrial Averages as of the end of December for the years 2004 through 2013.

Year	Closing
2004	10,783.01
2005	10,717.50
2006	12,463.15
2007	13,264.82
2008	8,776.39
2009	10,428.05
2010	11,577.51
2011	12,217.56
2012	13,104.14
2013	16,576.66

1.12 In-Demand College Majors. In a June 2013 article, published online by The Street, B. O'Connell discussed the results of a survey on opportunities for graduating college students. In one aspect of the survey, the following percentage estimates were reported on which college majors were in demand among U.S. firms. [SOURCE: "The Most In-Demand College Majors This Year." Published by Career-Builder, LLC, © 2013.]

Major	Percentage of U.S. firms
Business studies	31%
Computer sciences	24%
Engineering	17%
Health care sciences	10%
Engineering technologies	9%
Math and statistics	9%
Communications	7%
Education	7%
Science technology	6%
Liberal arts	6%

1.13 Thoughts on Evolution. In an article titled "Who has designs on your student's minds?" (*Nature*, Vol. 434, pp. 1062–1065), author G. Brumfiel postulated that support for Darwinism increases with level of education. The following table provides percentages of U.S. adults, by educational level, who believe that evolution is a scientific theory well supported by evidence.

Education	Percentage
Postgraduate education	65%
College graduate	52%
Some college education	32%
High school or less	20%

a. Do you think that this study is descriptive or inferential? Explain your answer.
b. If, in fact, the study is inferential, identify the sample and population.

1.14 Big-Banks Break-up. A nationwide survey of 1000 U.S. adults, conducted in March 2013 by Rasmussen Reports (field work by Pulse Opinion Research, LLC), found that 50% of respondents favored a plan to break up the 12 megabanks, which then controlled about 69% of the banking industry.
a. Identify the population and sample for this study.
b. Is the percentage provided a descriptive statistic or an inferential statistic? Explain your answer.

1.15 Genocide. The document "American Attitudes about Genocide" provided highlights of a nationwide poll with 1000 participants. The survey, conducted by Penn Schoen Berland between June 30 and July 10, 2012, revealed that "66% of respondents believe that genocide is preventable."
a. Is the statement in quotes an inferential or a descriptive statement? Explain your answer.
b. Based on the same information, what if the statement had been "66% of Americans believe that genocide is preventable"?

1.16 Vasectomies and Prostate Cancer. Refer to the vasectomy/prostate cancer study discussed in Example 1.5 on page 6.
a. How could the study be modified to make it a designed experiment?
b. Comment on the feasibility of the designed experiment that you described in part (a).

*In Exercises **1.17–1.22**, state whether the investigation in question is an observational study or a designed experiment. Justify your answer in each case.*

1.17 The Salk Vaccine. In the 1940s and early 1950s, the public was greatly concerned about polio. In an attempt to prevent this disease, Jonas Salk of the University of Pittsburgh developed a polio vaccine. In a test of the vaccine's efficacy, involving nearly 2 million grade-school children, half of the children received the Salk vaccine; the other half received a placebo, in this case an injection of salt dissolved in water. Neither the children nor the doctors performing the diagnoses knew which children belonged to which group, but an evaluation center did. The center found that the incidence of polio was far less among the children inoculated with the Salk vaccine. From that information, the researchers concluded that the vaccine would be effective in preventing polio for all U.S. school children; consequently, it was made available for general use.

1.18 Do Left-Handers Die Earlier? According to a study published in the *Journal of the American Public Health Association*, left-handed people do not die at an earlier age than right-handed people, contrary to the conclusion of a highly publicized report done 2 years earlier. The investigation involved a 6-year study of 3800 people in East Boston older than age 65. Researchers at Harvard University and the National Institute of Aging found that the "lefties" and "righties" died at exactly the same rate. "There was no difference, period," said Dr. J. Guralnik, an epidemiologist at the institute and one of the coauthors of the report.

1.19 Sex, Sleep, and PTSD. In the article, "One's Sex, Sleep, and Posttraumatic Stress Disorder" (*Biology of Sex Differences*, Vol. 3, No. 29, pp. 1–7), I. Kobayashi et al. study the relationship between one's sex, sleep patterns, and posttraumatic stress disorder (PTSD) after trauma exposure. The authors report that women have a higher lifetime prevalence of PTSD as well as a greater risk of developing PTSD following trauma exposure. Relationships between sleep and physical health have been documented in a number of studies, and the authors explore the possibility that disruptive sleep habits are common among people with PTSD and also a possible risk factor for the development of PTSD. A questionnaire of men and women with and without PTSD produced data on their sleep habits.

1.20 Aspirin and Cardiovascular Disease. In the article by P. Ridker et al. titled "A Randomized Trial of Low-dose Aspirin in the Primary Prevention of Cardiovascular Disease in Women" (*New England Journal of Medicine*, Vol. 352, pp. 1293–1304), the researchers noted that "We randomly assigned 39,876 initially healthy women 45 years of age or older to receive 100 mg of aspirin or placebo on alternate days and then monitored them for 10 years for a first major cardiovascular event (i.e., nonfatal myocardial infarction, nonfatal stroke, or death from cardiovascular causes)."

1.21 Heart Failure. In the paper "Cardiac-Resynchronization Therapy with or without an Implantable Defibrillator in Advanced Chronic Heart Failure" (*New England Journal of Medicine*, Vol. 350, pp. 2140–2150), M. Bristow et al. reported the results of a study of methods for treating patients who had advanced heart failure due to ischemic or nonischemic cardiomyopathies. A total of 1520 patients were randomly assigned in a 1:2:2 ratio to receive optimal pharmacologic therapy alone or in combination with either a pacemaker or a pacemaker–defibrillator combination. The patients were then observed until they died or were hospitalized for any cause.

1.22 Starting Salaries. The National Association of Colleges and Employers (NACE) compiles information on salary offers to new college graduates and publishes the results in *Salary Survey*.

Extending the Concepts and Skills

1.23 Ballistic Fingerprinting. In an on-line press release, ABCNews.com reported that "...73 percent of Americans...favor a law that would require every gun sold in the United States to be test-fired first, so law enforcement would have its fingerprint in case it were ever used in a crime."

a. Do you think that the statement in the press release is inferential or descriptive? Can you be sure?

b. Actually, ABCNews.com conducted a telephone survey of a random national sample of 1032 adults and determined that 73% of them favored a law that would require every gun sold in the United States to be test-fired first, so law enforcement would have its fingerprint in case it were ever used in a crime. How would you rephrase the statement in the press release to make clear that it is a descriptive statement? an inferential statement?

1.24 Causes of Death. The National Center for Health Statistics published the following data on the leading causes of death in 2010 in *National Vital Statistics Reports*. Deaths are classified according to the tenth revision of the *International Classification of Diseases*. Rates are per 100,000 population.

Rank	Cause of death	Rate
1	Diseases of the heart	193.6
2	Malignant neoplasms	186.2
3	Chronic lower respiratory diseases	44.7
4	Cerebrovascular diseases	41.9
5	Accidents (unintentional injuries)	39.1
6	Alzheimer's disease	27.0
7	Diabetes mellitus	22.4

Do you think that these rates are descriptive statistics or inferential statistics? Explain your answer.

1.25 Medical Testing on Animals. In its Summer 2013 *Animal Action Report*, the National Anti-Vivisection Society stated that "59% of Americans between the ages of 18 and 29 oppose medical testing on animals." The percentage of 59% was computed from sample data.

a. Identify the population under consideration.

b. Identify the sample under consideration.

c. Is the statement in quotes descriptive or inferential?

d. If you wanted to make it clear that the the percentage of 59% was computed from sample data, how would you rephrase the statement in quotes?

1.26 Lobbying Congress. In the special report, "Bitter Pill: Why Medical Bills Are Killing Us" (*TIME*, Vol. 181, No. 8, 2013), S. Brill presented an in-depth investigation of hospital billing practices that reveals why U.S. health care spending is out of control. One of the many statistics provided in the report is that, during the period from 1998 through 2012, the pharmaceutical and health-care-products industries and organizations representing doctors, hospitals, nursing homes, health services, and HMOs spent $5.36 billion lobbying Congress.

a. Under what conditions would the $5.36 billion lobbying-expenditure figure be a descriptive statistic? Explain your answer.

b. Under what conditions would the $5.36 billion lobbying-expenditure figure be an inferential statistic? Explain your answer.

1.2 Simple Random Sampling

Throughout this book, we present examples of organizations or people conducting studies: A consumer group wants information about the gas mileage of a particular make of car, so it performs mileage tests on a sample of such cars; a teacher wants to know about the comparative merits of two teaching methods, so she tests those methods on two groups of students. This approach reflects a healthy attitude: To obtain information about a subject of interest, plan and conduct a study.

What Does It Mean?

You can often avoid the effort and expense of a study if someone else has already done that study and published the results.

Suppose, however, that a study you are considering has already been done. Repeating it would be a waste of time, energy, and money. Therefore, before planning and conducting a study, do a literature search. You do not necessarily need to go through the entire library or make an extensive Internet search. Instead, you might use an information collection agency that specializes in finding studies on specific topics.

Census, Sampling, and Experimentation

If the information you need is not already available from a previous study, you might acquire it by conducting a **census**—that is, by obtaining information for the entire population of interest. However, conducting a census may be time consuming, costly, impractical, or even impossible.

Two methods other than a census for obtaining information are **sampling** and **experimentation.** In much of this book, we concentrate on sampling. However, we introduce experimentation in Section 1.4, discuss it sporadically throughout the text, and examine it in detail in the chapter *Design of Experiments and Analysis of Variance* (Module C) in the Regression-ANOVA Modules section on the WeissStats site.

If sampling is appropriate, you must decide how to select the sample; that is, you must choose the method for obtaining a sample from the population. Because the sample will be used to draw conclusions about the entire population, it should be a **representative sample**—that is, it should reflect as closely as possible the relevant characteristics of the population under consideration.

For instance, using the average weight of a sample of professional football players to make an inference about the average weight of all adult males would be unreasonable. Nor would it be reasonable to estimate the median income of California residents by sampling the incomes of Beverly Hills residents.

To see what can happen when a sample is not representative, consider the presidential election of 1936. Before the election, the *Literary Digest* magazine conducted an opinion poll of the voting population. Its survey team asked a sample of the voting population whether they would vote for Franklin D. Roosevelt, the Democratic candidate, or for Alfred Landon, the Republican candidate.

Based on the results of the survey, the magazine predicted an easy win for Landon. But when the actual election results were in, Roosevelt won by the greatest landslide in the history of presidential elections! What happened?

- The sample was obtained from among people who owned a car or had a telephone. In 1936, that group included only the more well-to-do people, and historically such people tend to vote Republican.
- The response rate was low (less than 25% of those polled responded), and there was a nonresponse bias (a disproportionate number of those who responded to the poll were Landon supporters).

The sample obtained by the *Literary Digest* was not representative.

Most modern sampling procedures involve the use of **probability sampling.** In probability sampling, a random device—such as tossing a coin, consulting a table of random numbers, or employing a random-number generator—is used to decide which members of the population will constitute the sample instead of leaving such decisions to human judgment.

The use of probability sampling may still yield a nonrepresentative sample. However, probability sampling helps eliminate unintentional selection bias and permits the researcher to control the chance of obtaining a nonrepresentative sample. Furthermore, the use of probability sampling guarantees that the techniques of inferential statistics can be applied. In this section and the next, we examine the most important probability-sampling methods.

Simple Random Sampling

The inferential techniques considered in this book are intended for use with only one particular sampling procedure: **simple random sampling.**

DEFINITION 1.4

Simple Random Sampling; Simple Random Sample

Simple random sampling: A sampling procedure for which each possible sample of a given size is equally likely to be the one obtained.

Simple random sample: A sample obtained by simple random sampling.

What Does It Mean?

Simple random sampling corresponds to our intuitive notion of random selection by lot.

There are two types of simple random sampling. One is **simple random sampling with replacement (SRSWR),** whereby a member of the population can be selected more than once; the other is **simple random sampling without replacement (SRS),** whereby a member of the population can be selected at most once. *Unless we specify otherwise, assume that simple random sampling is done without replacement.*

In Example 1.7, we chose a very small population—the five top Oklahoma state officials—to illustrate simple random sampling. In practice, we would not sample from such a small population but would instead take a census. Using a small population here makes understanding the concept of simple random sampling easier.

EXAMPLE 1.7 Simple Random Samples

TABLE 1.2
Five top Oklahoma state officials

Governor (G)
Lieutenant Governor (L)
Secretary of State (S)
Attorney General (A)
Treasurer (T)

TABLE 1.3
The 10 possible samples of two officials

G, L	G, S	G, A	G, T	L, S
L, A	L, T	S, A	S, T	A, T

TABLE 1.4
The five possible samples of four officials

G, L, S, A	G, L, S, T
G, L, A, T	G, S, A, T
L, S, A, T	

Sampling Oklahoma State Officials As reported by the *World Almanac*, the top five state officials of Oklahoma are as shown in Table 1.2. Consider these five officials a population of interest.

a. List the possible samples (without replacement) of two officials from this population of five officials.
b. Describe a method for obtaining a simple random sample of two officials from this population of five officials.
c. For the sampling method described in part (b), what are the chances that any particular sample of two officials will be the one selected?
d. Repeat parts (a)–(c) for samples of size 4.

Solution For convenience, we represent the officials in Table 1.2 by using the letters in parentheses.

a. Table 1.3 lists the 10 possible samples of two officials from this population of five officials.
b. To obtain a simple random sample of size 2, we could write the letters that correspond to the five officials (G, L, S, A, and T) on separate pieces of paper. After placing these five slips of paper in a box and shaking it, we could, while blindfolded, pick two slips of paper.
c. The procedure described in part (b) will provide a simple random sample. Consequently, each of the possible samples of two officials is equally likely to be the one selected. There are 10 possible samples, so the chances are $\frac{1}{10}$ (1 in 10) that any particular sample of two officials will be the one selected.
d. Table 1.4 lists the five possible samples of four officials from this population of five officials. A simple random sampling procedure, such as picking four slips of paper out of a box, gives each of these samples a 1 in 5 chance of being the one selected.

Exercise 1.43 on page 16

Random-Number Tables

Obtaining a simple random sample by picking slips of paper out of a box is usually impractical, especially when the population is large. Fortunately, we can use several practical procedures to get simple random samples. One common method involves a **table of random numbers**—a table of randomly chosen digits, as illustrated in Example 1.8.

EXAMPLE 1.8 Random-Number Tables

Sampling Student Opinions Student questionnaires, known as "teacher evaluations," gained widespread use in the late 1960s and early 1970s. Generally, professors hand out evaluation forms a week or so before the final.

That practice, however, poses several problems. On some days, less than 60% of students registered for a class may attend. Moreover, many of those who are present complete their evaluation forms in a hurry in order to prepare for other classes. A better method, therefore, might be to select a simple random sample of students from the class and interview them individually.

During one semester, Professor Hassett wanted to sample the attitudes of the students taking college algebra at his school. He decided to interview 15 of the 728 students enrolled in the course. Using a registration list on which the 728 students were numbered 1–728, he obtained a simple random sample of 15 students by randomly selecting 15 numbers between 1 and 728. To do so, he used the random-number table that appears in Appendix A as Table I and here as Table 1.5.

TABLE 1.5
Random numbers

Line number	Column number 00–09		10–19		20–29		30–39		40–49	
00	15544	80712	97742	21500	97081	42451	50623	56071	28882	28739
01	01011	21285	04729	39986	73150	31548	30168	76189	56996	19210
02	47435	53308	40718	29050	74858	64517	93573	51058	68501	42723
03	91312	75137	86274	59834	69844	19853	06917	17413	44474	86530
04	12775	08768	80791	16298	22934	09630	98862	39746	64623	32768
05	31466	43761	94872	92230	52367	13205	38634	55882	77518	36252
06	09300	43847	40881	51243	97810	18903	53914	31688	06220	40422
07	73582	13810	57784	72454	68997	72229	30340	08844	53924	89630
08	11092	81392	58189	22697	41063	09451	09789	00637	06450	85990
09	93322	98567	00116	35605	66790	52965	62877	21740	56476	49296
10	80134	12484	67089	08674	70753	90959	45842	59844	45214	36505
11	97888	31797	95037	84400	76041	96668	75920	68482	56855	97417
12	92612	27082	59459	69380	98654	20407	88151	56263	27126	63797
13	72744	45586	43279	44218	83638	05422	00995	70217	78925	39097
14	96256	70653	45285	26293	78305	80252	03625	40159	68760	84716
15	07851	47452	66742	83331	54701	06573	98169	37499	67756	68301
16	25594	41552	96475	56151	02089	33748	65289	89956	89559	33687
17	65358	15155	59374	80940	03411	94656	69440	47156	77115	99463
18	09402	31008	53424	21928	02198	61201	02457	87214	59750	51330
19	97424	90765	01634	37328	41243	33564	17884	94747	93650	77668

TABLE 1.6
Registration numbers of students interviewed

69	303	458	652	178
386	97	9	694	578
539	628	36	24	404

To select 15 random numbers between 1 and 728, we first pick a random starting point, say, by closing our eyes and placing a finger on Table 1.5. Then, beginning with the three digits under the finger, we go down the table and record the numbers as we go. Because we want numbers between 1 and 728 only, we discard the number 000 and numbers between 729 and 999. To avoid repetition, we also eliminate duplicate numbers. If we have not found enough numbers by the time we reach the bottom of the table, we move over to the next column of three-digit numbers and go up.

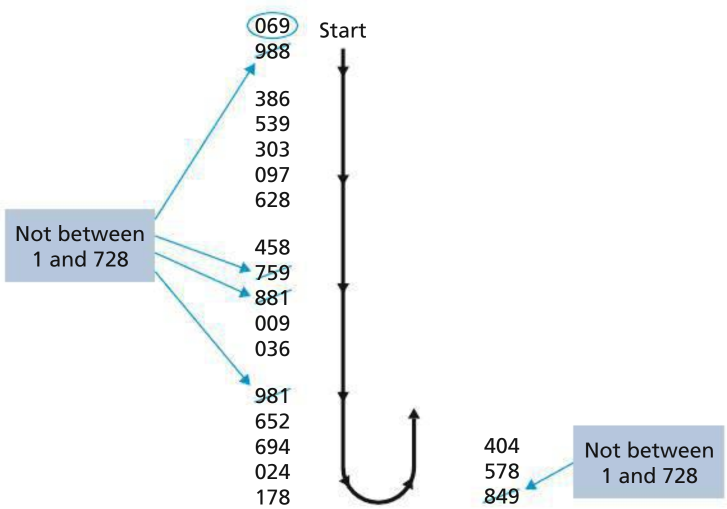
Using this procedure, Professor Hassett began with 069, circled in Table 1.5. Reading down from 069 to the bottom of Table 1.5 and then up the next column of three-digit numbers, he found the 15 random numbers displayed in Fig. 1.2 and in Table 1.6. Professor Hassett then interviewed the 15 students whose registration numbers are shown in Table 1.6.

StatCrunch Report 1.1

You try it!

Exercise 1.49(a) on page 16

FIGURE 1.2
Procedure used by Professor Hassett to obtain 15 random numbers between 1 and 728 from Table 1.5



Simple random sampling, the basic type of probability sampling, is also the foundation for the more complex types of probability sampling, which we explore in Section 1.3.

Random-Number Generators

Nowadays, statisticians prefer statistical software packages or graphing calculators, rather than random-number tables, to obtain simple random samples. The built-in programs for doing so are called **random-number generators.** When using random-number generators, be aware of whether they provide samples with replacement or samples without replacement. We discuss the use of random-number generators for obtaining simple random samples in the Technology Center at the end of this section.

Sources of Bias in Sample Surveys

Probability sampling eliminates bias in the selection of a sample from a complete and accurate list of the population. Nonetheless, many sources of bias often work their way into sample surveys, especially those involving large, human populations. We present a few of the many such sources in the exercises. For more details about this important topic, we recommend the book *Sampling: Design and Analysis, 2/e,* by Sharon L. Lohr (Boston: Brooks/Cole, 2010).

THE TECHNOLOGY CENTER

Today, programs for conducting statistical and data analyses are available in dedicated statistical software packages, general-use spreadsheet software, and graphing calculators. In this book, we discuss three of the most popular statistical technologies: Minitab, Excel, and the TI-83/84 Plus.

For Excel, we mostly use XLSTAT from Addinsoft, a statistics add-in that complements Excel's standard statistics capabilities. The version of XLSTAT that we use in this book is XLSTAT for Pearson, a special edition of XLSTAT that has been specifically built for Pearson Education.

"TI-83/84 Plus" is a shorthand for "TI-83 Plus and/or TI-84 Plus." The TI-calculator output that we show in this book is actually from the TI-84 Plus *C* Silver Edition. The style of the output from this calculator differs somewhat from that of the TI-83 Plus and earlier versions of the TI-84 Plus. Nonetheless, in most cases, keystrokes and output-content remain essentially the same for all three calculators.

At the end of most sections of this book, in subsections titled "The Technology Center," we present and interpret output from the three technologies that provides technology solutions to problems solved by hand earlier in the section. For this aspect of The Technology Center, you need neither a computer nor a graphing calculator, nor do you need working knowledge of any of the technologies.

Another aspect of The Technology Center provides step-by-step instructions for using the three technologies to obtain the output presented. When studying this material, you will get the best results by performing the steps described. *Note that, depending on the version of a statistical technology that you use, the appropriate instructions and the output you obtain may differ somewhat from that shown in the book.*

Successful use of technology requires knowing how to input data. We discuss that and other basic tasks for Minitab, Excel, and the TI-83/84 Plus in documents contained in the Technology Basics section of the WeissStats Resource Site (or, briefly, WeissStats site). The URL for the WeissStats site is www.pearsonhighered.com/weiss-stats. Note also that files for all appropriate data sets in the book can be found in multiple formats in the Data Sets section of the WeissStats site.

Using Technology to Obtain an SRS

In this Technology Center, we present output and step-by-step instructions for using technology to obtain a simple random sample without replacement.

Note to TI-83/84 Plus users:

- At the time of this writing, the TI-83/84 Plus does not have a built-in program for simple random sampling without replacement. However, a TI program called SRS, supplied in the TI Programs section of the WeissStats site, allows you to perform this procedure. Your instructor can show you how to download the program to your calculator.
- We recommend that you first *seed* the TI random-number generator, a task that only needs to be done once. Your instructor can also show you how to accomplish that task.
- After you apply the SRS program, the required SRS will be in List 1 (L1). *Warning:* Any data that you may have previously stored in List 1 will be erased during program execution, so copy those data to another list prior to program execution if you want to retain them.

EXAMPLE 1.9 Using Technology to Obtain an SRS

Sampling Student Opinions Recall that, during one semester, Professor Hassett wanted to sample the attitudes of the students taking college algebra at his school. He decided to interview 15 of the 728 students enrolled in the course. Use Minitab, Excel, or the TI-83/84 Plus to obtain a simple random sample without replacement (SRS) of 15 of the 728 students.

Solution Recall that Professor Hassett had a registration list on which the 728 students enrolled in the course were numbered 1–728. Thus, to obtain a simple random sample without replacement of 15 of the 728 students, we need only get an SRS of 15 numbers from the numbers 1–728.

To accomplish that, we applied the appropriate random-number programs. The results are shown in Output 1.1. Steps for generating that output are presented in Instructions 1.1.

OUTPUT 1.1 SRS of 15 numbers from 1–728

MINITAB

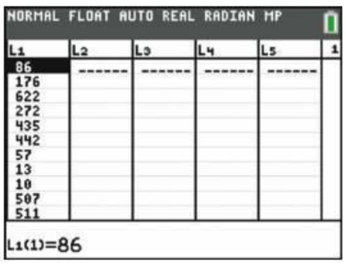
	REGNO	SAMPLE
1	1	321
2	2	430
3	3	241
4	4	11
5	5	545
6	6	185
7	7	16
8	8	534
9	9	428
10	10	596
11	11	460
12	12	641
13	13	331
14	14	46
15	15	203
16	16	
17	17	

EXCEL

Sampled data:

REGNO
541
325
278
387
596
109
563
69
300
142
163
399
416
370
157

TI-83/84 PLUS



Note: Only the first 11 numbers are visible in this output.

Of course, the 15 numbers that we obtained differ for the three technologies and will most likely differ from the 15 numbers that you would get by applying any of the programs.

INSTRUCTIONS 1.1 Steps for generating Output 1.1

MINITAB

1 Store the numbers 1–728 in a column named REGNO[†]
2 Choose **Calc ➤ Random Data ➤ Sample From Columns...**
3 Press the F3 key to reset the dialog box
4 Type 15 in the **Number of rows to sample** text box
5 Specify REGNO in the **From columns** text box
6 Type `SAMPLE` in the **Store samples in** text box
7 Click **OK**
8 The required sample is in the column named SAMPLE

EXCEL

1 Store the numbers 1–728 in a column named REGNO[‡]
2 Choose **XLSTAT ➤ Preparing data ➤ Data sampling**
3 Click the reset button ( ) in the lower left corner of the dialog box
4 Click in the **Data** selection box
5 Select the column of the worksheet that contains the REGNO data
6 Click the arrow button at the right of the **Sampling** drop-down list box and select **Random without replacement**
7 Type 15 in the **Sample size** text box
8 Click **OK**
9 Click the **Continue** button in the **XLSTAT – Selections** dialog box

TI-83/84 PLUS

1 Press **PRGM**
2 Arrow down to SRS and press **ENTER** twice
3 Type 1 for **MIN** (the smallest possible value) and press **ENTER**
4 Type 728 for **MAX** (the largest possible value) and press **ENTER**
5 Type 15 for **SAMPLE SIZE** and press **ENTER**
6 After the program completes, press **STAT** and then **ENTER**
7 The required sample is in **L1** (List 1)

† There are several ways to accomplish this task. Consult your instructor or search "Patterned data" in the Index tab of Minitab's Help facility.

‡ There are several ways to accomplish this task. Consult your instructor for his or her preferred method.

Exercises 1.2

Understanding the Concepts and Skills

1.27 Explain why a census is often not the best way to obtain information about a population.

1.28 Identify two statistical methods other than a census for obtaining information.

1.29 In sampling, explain why obtaining a representative sample is important.

1.30 Provide a scenario of your own in which a sample is not representative.

1.31 Regarding probability sampling:
a. What is it?
b. Does probability sampling always yield a representative sample? Explain your answer.
c. Identify some advantages of probability sampling.

1.32 Regarding simple random sampling:
a. What is simple random sampling?
b. What is a simple random sample?
c. Identify two forms of simple random sampling and explain the difference between the two.

1.33 The inferential procedures discussed in this book are intended for use with only one particular sampling procedure. What sampling procedure is that?

1.34 Identify two methods for obtaining a simple random sample.

1.35 What is the acronym used for simple random sampling without replacement?

1.36 The members of a population are numbered 1–5.
a. List the 10 possible samples (without replacement) of size 3 from this population.
b. If an SRS of size 3 is taken from the population, what are the chances of selecting 1, 3, and 5? Explain your answer.
c. Use Table I in Appendix A to obtain an SRS of size 3 from the population. Start at the single-digit number in line number 5 and column number 20, read down the column, up the next, and so on.

1.37 The members of a population are numbered 1–4.
a. List the 6 possible samples (without replacement) of size 2 from this population.
b. If an SRS of size 2 is taken from the population, what are the chances of selecting 2 and 3? Explain your answer.
c. Use Table I in Appendix A to obtain an SRS of size 2 from the population. Start at the single-digit number in line number 17 and column number 7, read down the column, up the next, and so on.

1.38 The members of a population are numbered 1–90.
a. Use Table I in Appendix A to obtain an SRS of size 5 from the population. Start at the two-digit number in line number 15 and column numbers 25–26, read down the column, up the next, and so on.
b. If you have access to a random-number generator, use it to solve part (a).

1.39 The members of a population are numbered 1–50.

a. Use Table I in Appendix A to obtain an SRS of size 6 from the population. Start at the two-digit number in line number 10 and column numbers 10−11, read down the column, up the next, and so on.

b. If you have access to a random-number generator, use it to solve part (a).

Applying the Concepts and Skills

1.40 Memorial Day Poll. In the year 2000, an on-line poll was conducted over Memorial Day weekend that asked people what they were doing to observe the holiday. The choices were: (1) stay home and relax, (2) vacation outdoors over the weekend, or (3) visit a military cemetery. More than 22,000 people participated in the poll, with 86% selecting option 1. Discuss this poll with regard to its suitability.

1.41 Estimating Median Income. Explain why a sample of 30 dentists from Seattle taken to estimate the median income of all Seattle residents is not representative.

1.42 Oklahoma State Officials. The five top Oklahoma state officials are displayed in Table 1.2 on page 11. Use that table to solve the following problems.

a. List the possible samples of size 1 that can be obtained from the population of five officials.

b. What is the difference between obtaining a simple random sample of size 1 and selecting one official at random?

c. List the possible samples (without replacement) of size 5 that can be obtained from the population of five officials.

d. What is the difference between obtaining a simple random sample of size 5 and taking a census of the five officials?

1.43 Oklahoma State Officials. The five top Oklahoma state officials are displayed in Table 1.2 on page 11. Use that table to solve the following problems.

a. List the 10 possible samples (without replacement) of size 3 that can be obtained from the population of five officials.

b. If a simple random sampling procedure is used to obtain a sample of three officials, what are the chances that it is the first sample on your list in part (a)? the second sample? the tenth sample?

1.44 Best-Selling Albums. The Recording Industry Association of America provides data on the best-selling albums of all time. As of May 28, 2013, the top six best-selling albums of all time (U.S. sales only), are by the artists the Eagles (E), Michael Jackson (M), Pink Floyd (P), Led Zeppelin (L), AC/DC (A), and Billy Joel (B).

a. List the 15 possible samples (without replacement) of two artists that can be selected from the six. For brevity, use the initial provided.

b. Describe a procedure for taking a simple random sample of two artists from the six.

c. If a simple random sampling procedure is used to obtain two artists, what are the chances of selecting P and A? M and E?

1.45 Best-Selling Albums. Refer to Exercise 1.44.

a. List the 15 possible samples (without replacement) of four artists that can be selected from the six.

b. Describe a procedure for taking a simple random sample of four artists from the six.

c. If a simple random sampling procedure is used to obtain four artists, what are the chances of selecting E, A, L, and B? P, B, M, and A?

1.46 Best-Selling Albums. Refer to Exercise 1.44.

a. List the 20 possible samples (without replacement) of three artists that can be selected from the six.

b. Describe a procedure for taking a simple random sample of three artists from the six.

c. If a simple random sampling procedure is used to obtain three artists, what are the chances of selecting M, A, and L? P, L, and E?

1.47 Social Networking Websites. From Wikipedia.com, we obtained the top seven major active social networking websites in the United States, excluding dating websites. Ranked according to registered users, as of April 2013, from most popular to least popular, they are Facebook (F), Twitter (T), Goggle+ (G), Habbo (H), LinkedIn (L), Bebo (B), and Tagged (A).

a. List the 21 possible samples (without replacement) of two social media websites that can be selected from the seven. For brevity, use the initial provided.

b. If a simple random sampling procedure is used to obtain two of these social media websites, what are the chances of selecting B and A? T and G?

1.48 Keno. In the game of keno, 20 balls are selected at random from 80 balls numbered 1–80.

a. Use Table I in Appendix A to simulate one game of keno by obtaining 20 random numbers between 1 and 80. Start at the two-digit number in line number 5 and column numbers 31–32, read down the column, up the next, and so on.

b. If you have access to a random-number generator, use it to solve part (a).

1.49 The International 500. Each year, *Fortune Magazine* publishes an article titled "The International 500" that provides a ranking by sales of the top 500 firms outside the United States. Suppose that you want to examine various characteristics of successful firms. Further suppose that, for your study, you decide to take a simple random sample of 10 firms from *Fortune Magazine*'s list of "The International 500."

a. Use Table I in Appendix A to obtain 10 random numbers that you can use to specify your sample. Start at the three-digit number in line number 14 and column numbers 10–12, read down the column, up the next, and so on.

b. If you have access to a random-number generator, use it to solve part (a).

1.50 Megacities Risk. In an issue of *Discover* (Vol. 26, No. 5, p. 14), A. Casselman looked at the natural-hazards risk index of megacities to evaluate potential loss from catastrophes such as earthquakes, storms, and volcanic eruptions. Urban areas have more to lose from natural perils, technological risks, and environmental hazards than rural areas. The top 10 megacities in the world are Tokyo, San Francisco, Los Angeles, Osaka, Miami, New York, Hong Kong, Manila, London, and Paris.

a. Suppose that you decide to take a simple random sample of five of these 10 megacities. Use Table I in Appendix A to obtain five random numbers that you can use to specify your sample.

b. If you have access to a random-number generator, use it to solve part (a).

1.51 Element Hunters. In the article "Element Hunters" (*National Geographic*, Vol. 233, No. 5, pp. 112−120), R. Dunn reports about the search for new undiscovered elements. Since 1940, scientists have been synthesizing elements one by one. The first was neptunium (Np), element number 93. There are, as of this writing, a total of 26 new synthetic elements. The following table provides their element numbers and symbols.

Number	Symbol	Number	Symbol
93	Np	106	Sg
94	Pu	107	Bh
95	Am	108	Hs
96	Cm	109	Mt
97	Bk	110	Ds
98	Cf	111	Rg
99	Es	112	Cn
100	Fm	113	Uut
101	Md	114	Fl
102	No	115	Uup
103	Lr	116	Lv
104	Rf	117	Uus
105	Db	118	Uuo

a. Suppose that you decide to take a simple random sample of eight of these new elements. Use Table I in Appendix A to obtain eight random numbers that you can use to specify your sample.
b. If you have access to a random-number generator, use it to solve part (a).

Extending the Concepts and Skills

Sources of Bias in Sample Surveys. As you know, probability sampling eliminates bias in choosing a sample from a (list of the entire) population. Nonetheless, many sources of bias often work their way into sample surveys of large human populations, such as those done by government agencies and in opinion polls. Exercises 1.52–1.54 present some aspects of sources of bias in sample surveys.

1.52 Undercoverage. Oftentimes, an accurate and complete list of the population is unavailable. In such cases, one or more groups will be omitted from the sampling process because they are not listed as part of the population. This type of bias is called *undercoverage.*
a. Explain why a sample survey of households will generally suffer from undercoverage.
b. Provide another example where bias due to undercoverage is likely to occur.

1.53 Nonresponse. When responses are not obtained from some of the individuals in the sample because either those individuals cannot be reached or refuse to participate, we have *nonresponse bias*.
a. Discuss some of the dangers of nonresponse.
b. Many sample surveys that are reported in the media have response rates as low as 10%. Explain the consequences of such low response rates in trying to generalize the results to the entire population.

1.54 Response bias. When the behavior of the interviewer or respondent results in inaccurate responses, we have *response bias.*
a. Explain why a survey question “Do you smoke marijuana” might result in response bias?
b. Provide some additional survey situations that might be conducive to response bias.
c. Provide some additional factors that might lead to response bias.

1.3 Other Sampling Designs*

Simple random sampling is the most natural and easily understood method of probability sampling—it corresponds to our intuitive notion of random selection by lot. However, simple random sampling does have drawbacks. For instance, it may fail to provide sufficient coverage when information about subpopulations is required and may be impractical when the members of the population are widely scattered geographically.

In this section, we examine some commonly used sampling procedures that are often more appropriate than simple random sampling. Remember, however, that the inferential procedures discussed in this book must be modified before they can be applied to data that are obtained by sampling procedures other than simple random sampling.

Systematic Random Sampling

One method that takes less effort to implement than simple random sampling is **systematic random sampling**. Procedure 1.1 presents a step-by-step method for implementing systematic random sampling.

PROCEDURE 1.1 Systematic Random Sampling

Step 1 Divide the population size by the sample size and round the result down to the nearest whole number, m.

Step 2 Use a random-number table or a similar device to obtain a number, k, between 1 and m.

Step 3 Select for the sample those members of the population that are numbered $k, k + m, k + 2m, \ldots$.

EXAMPLE 1.10 Systematic Random Sampling

Sampling Student Opinions Recall Example 1.8, in which Professor Hassett wanted a sample of 15 of the 728 students enrolled in college algebra at his school. Use systematic random sampling to obtain the sample.

Solution We apply Procedure 1.1.

Step 1 Divide the population size by the sample size and round the result down to the nearest whole number, *m*.

The population size is the number of students in the class, which is 728, and the sample size is 15. Dividing the population size by the sample size and rounding down to the nearest whole number, we get $728/15 = 48$ (rounded down). Thus, $m = 48$.

Step 2 Use a random-number table or a similar device to obtain a number, *k*, between 1 and *m*.

Referring to Step 1, we see that we need to randomly select a number between 1 and 48. Using a random-number table, we obtained the number 22 (but we could have conceivably gotten any number between 1 and 48, inclusive). Thus, $k = 22$.

TABLE 1.7
Numbers obtained by systematic random sampling

22	166	310	454	598
70	214	358	502	646
118	262	406	550	694

Step 3 Select for the sample those members of the population that are numbered $k, k + m, k + 2m, \ldots$.

From Steps 1 and 2, we see that $k = 22$ and $m = 48$. Hence, we need to list every 48th number, starting at 22, until we have 15 numbers. Doing so, we get the 15 numbers displayed in Table 1.7.

You try it!

Exercise 1.71 on page 24

Interpretation If Professor Hassett had used systematic random sampling and had begun with the number 22, he would have interviewed the 15 students whose registration numbers are shown in Table 1.7.

Systematic random sampling is easier to execute than simple random sampling and usually provides comparable results. The exception is the presence of some kind of cyclical pattern in the listing of the members of the population (e.g., male, female, male, female, . . .), a phenomenon that is relatively rare.

Cluster Sampling

Another sampling method is **cluster sampling,** which is particularly useful when the members of the population are widely scattered geographically. Procedure 1.2 provides a step-by-step method for implementing cluster sampling.

PROCEDURE 1.2 Cluster Sampling

Step 1 **Divide the population into groups (clusters).**

Step 2 **Obtain a simple random sample of the clusters.**

Step 3 **Use all the members of the clusters obtained in Step 2 as the sample.**

EXAMPLE 1.11 Cluster Sampling

The 300 members of a population have been divided into clusters of equal size 20. Use cluster sampling to obtain a sample of size 60 from the population.

Solution We apply Procedure 1.2.

Step 1 Divide the population into groups (clusters).

We know that the population size is 300 and that the population has been divided into clusters of equal size 20. Therefore, there are 15 (300/20) clusters. We number the 20 members of the population in cluster #1 from 1–20, those in cluster #2 from 21–40, and so forth.

Step 2 Obtain a simple random sample of the clusters.

Because each cluster contains 20 members of the population and we want a sample of size 60, we need a simple random sample of 3 (60/20) of the 15 clusters. Using a random-number generator, we obtained clusters #3, #4, and #10.

You try it!

Exercise 1.61 on page 23

Step 3 Use all the members of the clusters obtained in Step 2 as the sample.

From Step 2, we see that the 60 members of the population comprising the sample are those whose numbers are 41–60, 61–80, and 181–200. *Note:* Of course, the 60 members sampled would be different if one or more of the three clusters that were randomly selected in Step 2 were different.

Many years ago, citizens' groups pressured the city council of Tempe, Arizona, to install bike paths in the city. The council members wanted to be sure that they were supported by a majority of the taxpayers, so they decided to poll the city's homeowners.

Their first survey of public opinion was a questionnaire mailed out with the city's 18,000 homeowner water bills. Unfortunately, this method did not work very well. Only 19.4% of the questionnaires were returned, and a large number of those had written comments that indicated they came from avid bicyclists or from people who strongly resented bicyclists. The city council realized that the questionnaire generally had not been returned by the average homeowner.

An employee in the city's planning department had sample survey experience, so the council asked her to do a survey. She was given two assistants to help her interview 300 homeowners and 10 days to complete the project.

The planner first considered taking a simple random sample of 300 homes: 100 interviews for herself and for each of her two assistants. However, the city was so spread out that an interviewer of 100 randomly scattered homeowners would have to drive an average of 18 minutes from one interview to the next. Doing so would require approximately 30 hours of driving time for each interviewer and could delay completion of the report. The planner needed a different sampling design.

EXAMPLE 1.12 Cluster Sampling

Bike Paths Survey To save time, the planner decided to use cluster sampling. The residential portion of the city was divided into 947 blocks, each containing 20 homes, as shown in Fig. 1.3. Explain how the planner used cluster sampling to obtain a sample of 300 homes.

FIGURE 1.3
A typical block of homes

Solution We apply Procedure 1.2.

Step 1 Divide the population into groups (clusters).

The planner used the 947 blocks as the clusters, thus dividing the population (residential portion of the city) into 947 groups.

Step 2 Obtain a simple random sample of the clusters.

Because each block (cluster) contained 20 homes and the planner needed a sample of 300 homes, she required 15 (300/20) blocks. The planner numbered the blocks (clusters) from 1 to 947 and then used a table of random numbers to obtain a simple random sample of 15 of the 947 blocks.

Step 3 Use all the members of the clusters obtained in Step 2 as the sample.

The sample consisted of the 300 homes comprising the 15 sampled blocks:

$$15 \text{ blocks} \times 20 \text{ homes per block} = 300 \text{ homes.}$$

Exercise 1.73(a) on page 24

Interpretation The planner used cluster sampling to obtain a sample of 300 homes: 15 blocks of 20 homes per block. Each of the three interviewers was then assigned 5 of these 15 blocks. This method gave each interviewer 100 homes to visit (5 blocks of 20 homes per block) but saved much travel time because an interviewer could complete the interviews on an entire block before driving to another neighborhood. The report was finished on time.

Although cluster sampling can save time and money, it does have disadvantages. Ideally, each cluster should mirror the entire population. In practice, however, members of a cluster may be more homogeneous than the members of the entire population, which can cause problems.

For instance, consider a simplified small town, as depicted in Fig. 1.4. The town council wants to build a town swimming pool. A town planner needs to sample homeowner opinion about using public funds to build the pool. Many upper-income and middle-income homeowners may say "No" if they own or can access pools. Many low-income homeowners may say "Yes" if they do not have access to pools.

FIGURE 1.4
Clusters for a small town

If the planner uses cluster sampling and interviews the homeowners of, say, three randomly selected clusters, there is a good chance that no low-income homeowners will be interviewed.† And if no low-income homeowners are interviewed, the results of the survey will be misleading. If, for instance, the planner surveyed clusters #3, #5, and #8, then his survey would show that only about 30% of the homeowners want a pool. However, that is not true, because more than 40% of the homeowners actually want a pool. The clusters most strongly in favor of the pool would not have been included in the survey.

In this hypothetical example, the town is so small that common sense indicates that a cluster sample may not be representative. However, in situations with hundreds of clusters, such problems may be difficult to detect.

---

†There are 120 possible three-cluster samples, and 56 of those contain neither of the low-income clusters, #9 and #10. In other words, 46.7% of the possible three-cluster samples contain neither of the low-income clusters.

Stratified Sampling

Another sampling method, known as **stratified sampling,** is often more reliable than cluster sampling. In stratified sampling, the population is first divided into subpopulations, called **strata,** and then sampling is done from each stratum. Ideally, the members of each stratum should be homogeneous relative to the characteristic under consideration.

In stratified sampling, the strata are often sampled in proportion to their size, which is called **proportional allocation.** Procedure 1.3 presents a step-by-step method for implementing stratified (random) sampling with proportional allocation.

PROCEDURE 1.3 Stratified Random Sampling with Proportional Allocation

Step 1 **Divide the population into subpopulations (strata).**

Step 2 **From each stratum, obtain a simple random sample of size proportional to the size of the stratum; that is, the sample size for a stratum equals the total sample size times the stratum size divided by the population size.**

Step 3 **Use all the members obtained in Step 2 as the sample.**

EXAMPLE 1.13 Stratified Sampling with Proportional Allocation

The 2000 members of a population have been divided into four strata of sizes 400, 600, 800, and 200. Use stratified sampling with proportional allocation to obtain a sample of size 10 from the population.

Solution We apply Procedure 1.3.

Step 1 **Divide the population into subpopulations (strata).**

We know that the population size is 2000 and that the population has been divided into four strata of sizes 400, 600, 800, and 200. We number the 400 members of the population in strata #1 from 1–400, those in strata #2 from 401–1000, and so forth.

Step 2 **From each stratum, obtain a simple random sample of size proportional to the size of the stratum; that is, the sample size for a stratum equals the total sample size times the stratum size divided by the population size.**

The sample size for stratum #1 is

$$\text{Total sample size} \times \frac{\text{Size of stratum \#1}}{\text{Population size}} = 10 \times \frac{400}{2000} = 2.$$

Similarly, we find that the sample sizes for strata #2, #3, and #4 are 3, 4, and 1, respectively.

We used a random-number generator to get a simple random sample of size 2 from the 400 members of stratum #1 (numbered 1–400) and obtained 166 and 264. Proceeding similarly, we obtained the required simple random samples shown in the final column of the following table.

Stratum	Size	Numbered	Sample size	Sample
#1	400	1–400	2	166, 264
#2	600	401–1000	3	454, 511, 620
#3	800	1001–1800	4	1246, 1420, 1759, 1793
#4	200	1801–2000	1	1938

Step 3 Use all the members obtained in Step 2 as the sample.

From Step 2, we see that the 10 members of the population comprising the sample are those whose numbers are 166, 264, 454, 511, 620, 1246, 1420, 1759, 1793, and 1938. *Note:* Of course, the 10 members sampled would be different if one or more of the four samples in the last column of the preceding table were different.

You try it!

Exercise 1.63 on page 23

EXAMPLE 1.14 Stratified Sampling with Proportional Allocation

Town Swimming Pool Consider again the town swimming pool situation discussed on page 20. The town has 250 homeowners of which 25, 175, and 50 are upper income, middle income, and low income, respectively. Explain how we can obtain a sample of 20 homeowners, using stratified sampling with proportional allocation, stratifying by income group.

Solution We apply Procedure 1.3.

Step 1 Divide the population into subpopulations (strata).

We divide the homeowners in the town into three strata according to income group: upper income, middle income, and low income.

Step 2 From each stratum, obtain a simple random sample of size proportional to the size of the stratum; that is, the sample size for a stratum equals the total sample size times the stratum size divided by the population size.

Of the 250 homeowners, 25 are upper income, 175 are middle income, and 50 are lower income. The sample size for the upper-income homeowners is, therefore,

$$\text{Total sample size} \times \frac{\text{Number of upper-income homeowners}}{\text{Total number of homeowners}} = 20 \cdot \frac{25}{250} = 2.$$

Similarly, we find that the sample sizes for the middle-income and lower-income homeowners are 14 and 4, respectively. Thus, we take a simple random sample of size 2 from the 25 upper-income homeowners, of size 14 from the 175 middle-income homeowners, and of size 4 from the 50 lower-income homeowners.

Step 3 Use all the members obtained in Step 2 as the sample.

The sample consists of the 20 homeowners selected in Step 2, namely, the 2 upper-income, 14 middle-income, and 4 lower-income homeowners.

Interpretation This stratified sampling procedure ensures that no income group is missed. It also improves the precision of the statistical estimates (because the homeowners within each income group tend to be homogeneous) and makes it possible to estimate the separate opinions of each of the three strata (income groups).

You try it!

Exercise 1.73(c) on page 24

In both Examples 1.13 and 1.14, the calculated sample sizes for all the strata turned out to be whole numbers. Sometimes, however, things don't work out that easily.

For instance, suppose that the population size is 100, the total sample size is 10, and the strata sizes are 29, 41, and 30. Applying the formula

$$\text{Stratum sample size} = \text{Total sample size} \times \frac{\text{Size of stratum}}{\text{Population size}},$$

we find the strata "sample sizes" to be 2.9, 4.1, and 3. Of course, sample sizes must be whole numbers. To remedy the situation here, we can simply round to get the strata sample sizes 3, 4, and 3. Note that, in this case, rounding preserves the condition that the total sample size (10) equals the sum of the strata sample sizes $(3 + 4 + 3)$.

Simple rounding, however, doesn't always remedy the situation. For instance, suppose that the population size is 100, the total sample size is 10, and the strata sizes are 26, 47, and 27. Applying the preceding formula, we find the strata "sample sizes" to be 2.6, 4.7, and 2.7. Rounding these "sample sizes" yields the whole numbers 3, 5, and 3, which do not sum to the total sample size of 10. One way to remedy the situation here would be to round up the two strata "sample sizes" that are closest to their rounded-up values (i.e., round 4.7 and 2.7 up to 5 and 3, respectively) and round down the other strata "sample size" (i.e., round 2.6 down to 2).

Generally, we often need to get creative in deciding how to deal with fractional strata "sample sizes." You will be asked to do so in some of the exercises at the end of this section.

Multistage Sampling

Most large-scale surveys combine one or more of simple random sampling, systematic random sampling, cluster sampling, and stratified sampling. Such **multistage sampling** is used frequently by pollsters and government agencies.

For instance, the U.S. National Center for Health Statistics conducts surveys of the civilian noninstitutional U.S. population to obtain information on illnesses, injuries, and other health issues. Data collection is by a multistage probability sample of approximately 42,000 households. Information obtained from the surveys is published in the *National Health Interview Survey*.

Exercises 1.3

Understanding the Concepts and Skills

*In each of Exercises **1.55–1.58**, fill in the blank(s).*

1.55 Systematic random sampling is easier to execute than simple random sampling and usually provides comparable results. The exception is the presence of some kind of ______ ______ in the listing of the members of the population.

1.56 Ideally, in cluster sampling, each cluster should ______ the entire population.

1.57 Ideally, in stratified sampling, the members of each stratum should be ______ relative to the characteristic under consideration.

1.58 Surveys that combine one or more of simple random sampling, systematic random sampling, cluster sampling, and stratified sampling employ what is called ______ sampling.

1.59 The members of a population have been numbered 1–372. A sample of size 5 is to be taken from the population, using systematic random sampling.

a. Apply Procedure 1.1 on page 17 to determine the sample (i.e., the numbers corresponding to the members of the population that are included in the sample).

b. Suppose that, in Step 2 of Procedure 1.1, the random number chosen is 10 (i.e., $k = 10$). Determine the sample.

1.60 The members of a population have been numbered 1–500. A sample of size 9 is to be taken from the population, using systematic random sampling.

a. Apply Procedure 1.1 on page 17 to determine the sample (i.e., the numbers corresponding to the members of the population that are included in the sample).

b. Suppose that, in Step 2 of Procedure 1.1, the random number chosen is 48 (i.e., $k = 48$). Determine the sample.

1.61 The members of a population have been numbered 1–50. A sample of size 20 is to be taken from the population, using cluster sampling. The clusters are of equal size 10, where cluster #1 consists of the members of the population numbered 1–10, cluster #2 consists of the members of the population numbered 11–20, and so forth.

a. Apply Procedure 1.2 on page 18 to determine the sample (i.e., the numbers corresponding to the members of the population that are included in the sample).

b. Suppose that, in Step 2 of Procedure 1.2, clusters #1 and #3 are selected. Determine the sample.

1.62 The members of a population have been numbered 1–100. A sample of size 30 is to be taken from the population, using cluster sampling. The clusters are of equal size 10, where cluster #1 consists of the members of the population numbered 1–10, cluster #2 consists of the members of the population numbered 11–20, and so forth.

a. Apply Procedure 1.2 on page 18 to determine the sample (i.e., the numbers corresponding to the members of the population that are included in the sample).

b. Suppose that, in Step 2 of Procedure 1.2, clusters #2, #6, and #9 are selected. Determine the sample.

1.63 The members of a population have been numbered 1–1000. A sample of size 20 is to be taken from the population, using stratified random sampling with proportional allocation. The strata are of sizes 300, 200, 400, and 100, where stratum #1 consists of the members of the population numbered 1–300, stratum #2 consists of the members of the population numbered 301–500, and so forth.

a. Determine the sample sizes that will be taken from the strata.

b. Apply Procedure 1.3 on page 21 to determine the sample (i.e., the numbers corresponding to the members of the population that are included in the sample).

1.64 The members of a population have been numbered 1–500. A sample of size 10 is to be taken from the population, using stratified random sampling with proportional allocation. The strata are of sizes 200, 150, and 150, where stratum #1 consists of the members of the population numbered 1–200, stratum #2 consists of the members of the population numbered 201–350, and so forth.

a. Determine the sample sizes that will be taken from the strata.
b. Apply Procedure 1.3 on page 21 to determine the sample (i.e., the numbers corresponding to the members of the population that are included in the sample).

Applying the Concepts and Skills

1.65 Ghost of Speciation Past. In the article, "Ghost of Speciation Past" (*Nature*, Vol. 435, pp. 29–31), T. Kocher looked at the origins of a diverse flock of cichlid fishes in the lakes of southeast Africa. Suppose that you wanted to select a sample from the hundreds of species of cichlid fishes that live in the lakes of southeast Africa. If you took a simple random sample from the species of each lake and combined all the simple random samples into one sample, which type of sampling design would you have used? Explain your answer.

1.66 Number of Farms. The National Agricultural Statistics Service (NASS) conducts studies of the number of acres devoted to farms in each county of the United States. Suppose that we divide the United States into the four census regions (Northeast, North Central, South, and West), take a simple random sample of counties from each of the four regions, and combine all four simple random samples into one sample. What type of sampling design have we used? Explain your answer.

1.67 John F. Kennedy. In one of his books, Ted Sorenson, Special Counsel to President John F. Kennedy, presents an intimate biography of the extraordinary man. According to Sorenson, Kennedy "read every fiftieth letter of the thirty thousand coming weekly to the White House." What type of sampling design was Kennedy using in this case? Explain your answer. [SOURCE: From *Kennedy: The Classical Biography* by Ted Sorenson. Published by Harper Perennial, © 2009.]

1.68 Litigation Surveys. In the article, "Non-probability Sampling Designs for Litigation Surveys (*Trademark Reporter*, Vol. 81, pp. 169–179), J. Jacoby and H. Handlin discussed the controversy about whether nonprobability samples are acceptable as evidence in litigation. The authors randomly selected 26 journals from a list of 1285 scholarly journals in the social and behavioral sciences. They examined all articles published during one year in each of the 26 journals selected with regard to sampling methods. What type of sampling design was used by these two authors in their investigation? Explain your answer.

1.69 Immunization of Schoolchildren. In the article, "Reasons for Non-uptake of Measles, Mumps, and Rubella Catch Up Immunisation in a Measles Epidemic and Side Effects of the Vaccine" (*British Medical Journal*, Vol. 310, pp. 1629–1632), R. Roberts et al. discussed a follow-up survey to examine why almost 10,000 children, ages 11–15, whose records showed no previous immunization were not immunized. In the survey, 10 of the 46 schools participating in the immunization campaign were randomly chosen and then the parents of all the nonimmunized children at the 10 selected schools were sent a questionnaire. What type of sampling design was used by these authors in their survey? Explain your answer.

1.70 University Parking Facilities. During one year, a university wanted to gauge the sentiment of the people using the university's parking facilities. Each of the 8493 people that used the parking facilities had a sticker with a unique number between 1 and 8493. The university committee on parking decided to sample 30 users of the parking facilities and obtain their views on those facilities. The committee selected a number at random between 1 and 283 and got the number 10. The people interviewed were the ones whose stickers had numbers 10, 293, 576, ..., 8217. What type of sampling design was used by the university committee on parking? Explain your answer.

1.71 The International 500. In Exercise 1.49 on page 16, you used simple random sampling to obtain a sample of 10 firms from *Fortune Magazine*'s list of "The International 500."

a. Use systematic random sampling to accomplish that same task.
b. Which method is easier: simple random sampling or systematic random sampling?
c. Does it seem reasonable to use systematic random sampling to obtain a representative sample? Explain your answer.

1.72 Keno. In the game of keno, 20 balls are selected at random from 80 balls numbered 1–80. In Exercise 1.48 on page 16, you used simple random sampling to simulate one game of keno.

a. Use systematic random sampling to obtain a sample of 20 of the 80 balls.
b. Which method is easier: simple random sampling or systematic random sampling?
c. Does it seem reasonable to use systematic random sampling to simulate one game of keno? Explain your answer.

1.73 Sampling Dorm Residents. Students in the dormitories of a university in the state of New York live in clusters of four double rooms, called *suites*. There are 48 suites, with eight students per suite.

a. Describe a cluster sampling procedure for obtaining a sample of 24 dormitory residents.
b. Students typically choose friends from their classes as suitemates. With that in mind, do you think cluster sampling is a good procedure for obtaining a representative sample of dormitory residents? Explain your answer.
c. The university housing office has separate lists of dormitory residents by class level. The number of dormitory residents in each class level is as follows.

Class level	Number of dorm residents
Freshman	128
Sophomore	112
Junior	96
Senior	48

Use the table to design a procedure for obtaining a stratified sample (with proportional allocation) of 24 dormitory residents.

1.74 Best High Schools. In an issue of *Newsweek* (Vol. CXLV, No. 20, pp. 48–57), B. Kantrowitz listed "The 100 best high schools in America" according to a ranking devised by J. Mathews. Another characteristic measured from the high school is the percent free lunch, which is the percentage of student body that is eligible for free and reduced-price lunches, an indicator of socioeconomic status. A percentage of 40% or more generally signifies a high concentration of children in poverty. The top 100 schools, grouped according to their percent free lunch, is as follows.

Percent free lunch	Number of top 100 ranked high schools
0–under 10	50
10–under 20	18
20–under 30	11
30–under 40	8
40 or over	13

a. Use the table to design a procedure for obtaining a stratified sample (with proportional allocation) of 25 high schools from the list of the top 100 ranked high schools.
b. If stratified random sampling with proportional allocation is used to select the sample of 25 high schools, how many would be selected from the stratum with a percent-free-lunch value of 30–under 40?

1.75 U.S. House of Representatives. There are 435 representatives in the 113th session of the U.S. House of Representatives. On the website www.house.gov, you can find an alphabetized list of the 435 congresspersons. In 2013, the first representative listed is Robert Aderholt, a Republican from Alabama, and the last representative listed is Todd Young, a Republican from Indiana. Suppose that the alphabetized list is indexed 1 through 435.
a. Use systematic random sampling to obtain a sample of 15 of the 435 representatives.
b. Suppose that, in Step 2 of Procedure 1.1, the random number chosen is 12 (i.e., $k = 12$). Determine the sample.

1.76 Peace Corps Volunteers. The Peace Corps is an independent U.S. government agency that provides trained volunteers for countries requesting assistance. According to Peacecorps.org, as of September 2012, volunteers currently serve in about 76 different host countries. The average age of a volunteer is about 28 years old. The following table reports the percentage of total volunteers serving by geographic region.

Region	Percent of volunteers
Africa	43%
Latin America	21%
Eastern Europe/Central Asia	15%
Asia	10%
Caribbean	4%
North Africa/Middle East	4%
Pacific Islands	3%

a. Use the table to design a procedure for obtaining a stratified sample (with proportional allocation) of 50 Peace Corps volunteers.
b. If stratified random sampling with proportional allocation is used to select the sample of 50 Peace Corps volunteers, how many would be selected from the Caribbean?

Extending the Concepts and Skills

1.77 The Terri Schiavo Case. In the early part of 2005, the Terri Schiavo case received national attention as her husband sought to have life support removed, and her parents sought to maintain that life support. The courts allowed the life support to be removed, and her death ensued. A Harris Poll of 1010 U.S. adults was taken by telephone on April 21, 2005, to determine how common it is for life support systems to be removed. Those questioned in the sample were asked: (1) Has one of your parents, a close friend, or a family member died in the last 10 years? (2) Before (this death/these deaths) happened, was this person/were any of these people, kept alive by any support system? (3) Did this person die while on a life support system, or had it been withdrawn? Respondents were also asked questions about age, sex, race, education, region, and household income to ensure that results represented a cross section of U.S. adults.
a. What kind of sampling design was used in this survey? Explain your answer.
b. If 78% of the respondents answered the first question in the affirmative, what was the approximate sample size for the second question?
c. If 28% of those responding to the second question answered "yes," what was the approximate sample size for the third question?

1.78 In simple random sampling, all samples of a given size are equally likely. Is that true in systematic random sampling? Explain your answer.

1.79 In simple random sampling, it is also true that each member of the population is equally likely to be selected, the chance for each member being equal to the sample size divided by the population size.
a. Under what circumstances is that fact also true for systematic random sampling? Explain your answer.
b. Provide an example in which that fact is not true for systematic random sampling.

1.80 In simple random sampling, it is also true that each member of the population is equally likely to be selected, the chance for each member being equal to the sample size divided by the population size. Show that this fact is also true for stratified random sampling with proportional allocation.

1.4 Experimental Designs*

As we mentioned earlier, two methods for obtaining information, other than a census, are sampling and experimentation. In Sections 1.2 and 1.3, we discussed some of the basic principles and techniques of sampling. Now, we do the same for experimentation.

Principles of Experimental Design

The study presented in Example 1.6 on page 6 illustrates three basic principles of experimental design: **control, randomization,** and **replication.**

- *Control:* The doctors compared the rate of major birth defects for the women who took folic acid to that for the women who took only trace elements.
- *Randomization:* The women were divided randomly into two groups to avoid unintentional selection bias.
- *Replication:* A large number of women were recruited for the study to make it likely that the two groups created by randomization would be similar and also to increase the chances of detecting any effect due to the folic acid.

In the language of experimental design, each woman in the folic acid study is an **experimental unit,** or a **subject.** More generally, we have the following definition.

DEFINITION 1.5

Experimental Units; Subjects

In a designed experiment, the individuals or items on which the experiment is performed are called **experimental units.** When the experimental units are humans, the term **subject** is often used in place of experimental unit.

In the folic acid study, both doses of folic acid (0.8 mg and none) are called *treatments* in the context of experimental design. Generally, each experimental condition is called a **treatment,** of which there may be several.

Now that we have introduced the terms *experimental unit* and *treatment*, we can present the three basic principles of experimental design in a general setting.

KEY FACT 1.1

Principles of Experimental Design

The following principles of experimental design enable a researcher to conclude that differences in the results of an experiment not reasonably attributable to chance are likely caused by the treatments.

- **Control:** Two or more treatments should be compared.
- **Randomization:** The experimental units should be randomly divided into groups to avoid unintentional selection bias in constituting the groups.
- **Replication:** A sufficient number of experimental units should be used to ensure that randomization creates groups that resemble each other closely and to increase the chances of detecting any differences among the treatments.

One of the most common experimental situations involves a specified treatment and *placebo,* an inert or innocuous medical substance. Technically, both the specified treatment and placebo are treatments. The group receiving the specified treatment is called the **treatment group,** and the group receiving placebo is called the **control group.** In the folic acid study, the women who took folic acid constituted the treatment group, and those women who took only trace elements constituted the control group.

Terminology of Experimental Design

In the folic acid study, the researchers were interested in the effect of folic acid on major birth defects. Birth-defect classification (whether major or not) is the **response variable** for this study. The daily dose of folic acid is called the **factor.** In this case, the factor has two **levels,** namely, 0.8 mg and none.

When there is only one factor, as in the folic acid study, the treatments are the same as the levels of the factor. If a study has more than one factor, however, each treatment is a combination of levels of the various factors.

DEFINITION 1.6

Response Variable, Factors, Levels, and Treatments

Response variable: The characteristic of the experimental outcome that is to be measured or observed.

Factor: A variable whose effect on the response variable is of interest in the experiment.

Levels: The possible values of a factor.

Treatment: Each experimental condition. For one-factor experiments, the treatments are the levels of the single factor. For multifactor experiments, each treatment is a combination of levels of the factors.

EXAMPLE 1.15 Experimental Design

Weight Gain of Golden Torch Cacti The golden torch cactus (*Trichocereus spachianus*), a cactus native to Argentina, has excellent landscape potential. W. Feldman and F. Crosswhite, two researchers at the Boyce Thompson Southwestern Arboretum, investigated the optimal method for producing these cacti.

The researchers examined, among other things, the effects of a hydrophilic polymer and irrigation regime on weight gain. Hydrophilic polymers are used as soil additives to keep moisture in the root zone. For this study, the researchers chose Broadleaf P-4 polyacrylamide, abbreviated P4. The hydrophilic polymer was either used or not used, and five irrigation regimes were employed: none, light, medium, heavy, and very heavy. Identify the

a. experimental units. **b.** response variable. **c.** factors.
d. levels of each factor. **e.** treatments.

Solution

a. The experimental units are the cacti used in the study.
b. The response variable is weight gain.
c. The factors are hydrophilic polymer and irrigation regime.
d. Hydrophilic polymer has two levels: with and without. Irrigation regime has five levels: none, light, medium, heavy, and very heavy.
e. Each treatment is a combination of a level of hydrophilic polymer and a level of irrigation regime. Table 1.8 depicts the 10 treatments for this experiment. In the table, we abbreviated "very heavy" as "Xheavy."

You try it!

Exercise 1.93 on page 30

TABLE 1.8
Schematic for the 10 treatments in the cactus study

		Irrigation regime				
		None	Light	Medium	Heavy	Xheavy
Polymer	No P4	No water No P4 (Treatment 1)	Light water No P4 (Treatment 2)	Medium water No P4 (Treatment 3)	Heavy water No P4 (Treatment 4)	Xheavy water No P4 (Treatment 5)
	With P4	No water With P4 (Treatment 6)	Light water With P4 (Treatment 7)	Medium water With P4 (Treatment 8)	Heavy water With P4 (Treatment 9)	Xheavy water With P4 (Treatment 10)

Statistical Designs

Once we have chosen the treatments, we must decide how the experimental units are to be assigned to the treatments (or vice versa). The women in the folic acid study were randomly divided into two groups; one group received folic acid and the other only trace elements. In the cactus study, 40 cacti were divided randomly into 10 groups of 4 cacti each, and then each group was assigned a different treatment from among the 10 depicted in Table 1.8. Both of these experiments used a **completely randomized design.**

DEFINITION 1.7

Completely Randomized Design

In a **completely randomized design,** all the experimental units are assigned randomly among all the treatments.

Although the completely randomized design is commonly used and simple, it is not always the best design. Several alternatives to that design exist.

For instance, in a **randomized block design,** experimental units that are similar in ways that are expected to affect the response variable are grouped in **blocks.** Then the random assignment of experimental units to the treatments is made block by block.

DEFINITION 1.8

Randomized Block Design

In a **randomized block design,** the experimental units are assigned randomly among all the treatments separately within each block.

Example 1.16 contrasts completely randomized designs and randomized block designs.

EXAMPLE 1.16 Statistical Designs

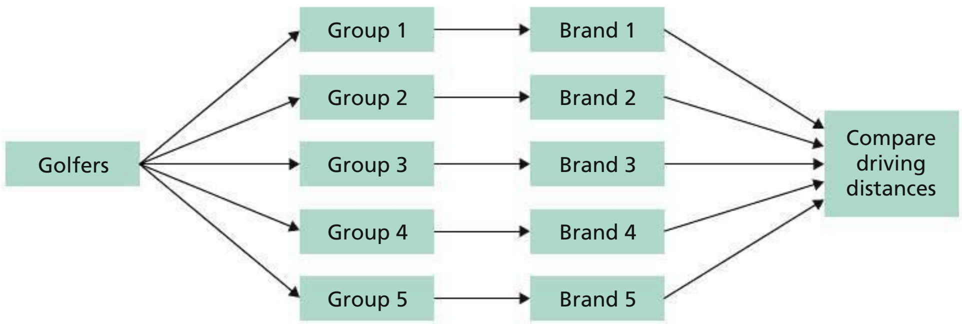
Golf Ball Driving Distances Suppose we want to compare the driving distances for five different brands of golf ball. For 40 golfers, discuss a method of comparison based on

a. a completely randomized design.
b. a randomized block design.

Solution Here the experimental units are the golfers, the response variable is driving distance, the factor is brand of golf ball, and the levels (and treatments) are the five brands.

a. For a completely randomized design, we would randomly divide the 40 golfers into five groups of 8 golfers each and then randomly assign each group to drive a different brand of ball, as illustrated in Fig. 1.5.

FIGURE 1.5 Completely randomized design for golf ball experiment



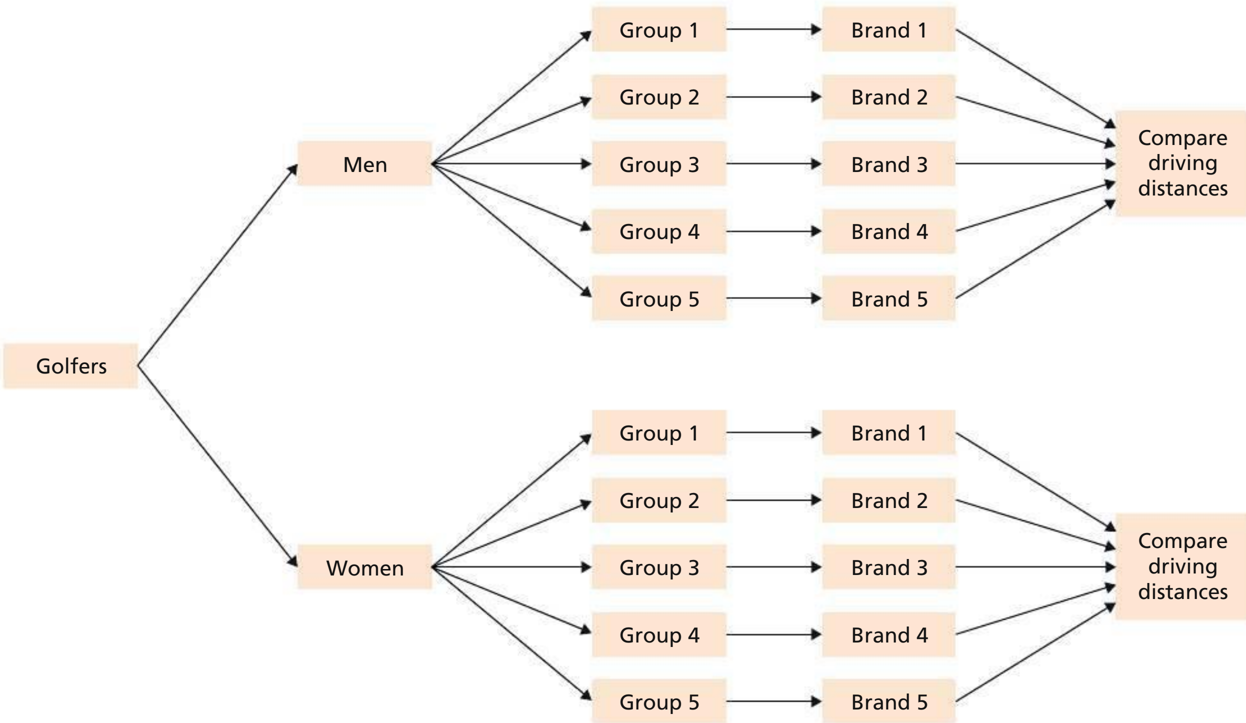
b. Because driving distance is affected by gender, using a randomized block design that blocks by gender is probably a better approach. We could do so by using 20 men golfers and 20 women golfers. We would randomly divide the 20 men into five groups of 4 men each and then randomly assign each group to drive a different brand of ball, as shown in Fig. 1.6. Likewise, we would randomly divide the 20 women into five groups of 4 women each and then randomly assign each group to drive a different brand of ball, as also shown in Fig. 1.6.

By blocking, we can isolate and remove the variation in driving distances between men and women and thereby make it easier to detect any differences in driving distances among the five brands of golf ball. Additionally, blocking permits us to analyze separately the differences in driving distances among the five brands for men and women.

Exercise 1.99 on page 30

As illustrated in Example 1.16, blocking can isolate and remove systematic differences among blocks, thereby making any differences among treatments easier to detect. Blocking also makes possible the separate analysis of treatment effects on each block.

FIGURE 1.6 Randomized block design for golf ball experiment



In this section, we introduced some of the basic terminology and principles of experimental design. However, we have just scratched the surface of this vast and important topic to which entire courses and books are devoted. Further discussion of experimental design is provided in the chapter *Design of Experiments and Analysis of Variance* (Module C) in the Regression-ANOVA Modules section on the WeissStats site.

Exercises 1.4

Understanding the Concepts and Skills

1.81 In a designed experiment,

a. what are the experimental units?

b. if the experimental units are humans, what term is often used in place of experimental unit?

1.82 State and explain the significance of the three basic principles of experimental design.

1.83 Define each of the following terms in the context of experimental design.

a. Response variable **b.** Factor

c. Levels **d.** Treatments

1.84 In this section, we discussed two types of statistical designs. Identify and explain the meaning of each one.

1.85 In a designed experiment, there is one factor with four levels. How many treatments are there?

1.86 In a designed experiment, there is one factor with five levels. How many treatments are there?

1.87 In a designed experiment, there are two factors, say, Factor A and Factor B. Factor A has three levels, say, a_1, a_2, and a_3; Factor B has four levels, say, b_1, b_2, b_3, and b_4.

a. Construct a schematic for the treatments similar to Table 1.8 on page 27.

b. Use part (a) to determine the number of treatments.

c. Could you have determined the answer for part (b) without constructing the schematic in part (a)? Explain your answer.

1.88 In a designed experiment, there are two factors, say, Factor A and Factor B. Factor A has four levels, say, a_1, a_2, a_3, and a_4; Factor B has two levels, say, b_1 and b_2.

a. Construct a schematic for the treatments similar to Table 1.8 on page 27.

b. Use part (a) to determine the number of treatments.

c. Could you have determined the answer for part (b) without constructing the schematic in part (a)? Explain your answer.

1.89 In a designed experiment, there are two factors. One factor has m levels and the other factor has n levels. Determine the number of treatments. *Hint:* Refer to Exercise 1.87 or Exercise 1.88.

Applying the Concepts and Skills

1.90 Adverse Effects of Prozac. Prozac (fluoxetine hydrochloride), a product of Eli Lilly and Company, is used for the treatment of depression, obsessive–compulsive disorder (OCD), and bulimia

nervosa. An issue of the magazine *Arthritis Today* contained an advertisement reporting on the "…treatment-emergent adverse events that occurred in 2% or more patients treated with Prozac and with incidence greater than placebo in the treatment of depression, OCD, or bulimia." In the study, 2444 patients took Prozac and 1331 patients were given placebo. Identify the
a. treatment group. **b.** control group. **c.** treatments.

1.91 Treating Heart Failure. In the journal article "Cardiac-Resynchronization Therapy with or without an Implantable Defibrillator in Advanced Chronic Heart Failure" (*New England Journal of Medicine*, Vol. 350, pp. 2140–2150), M. Bristow et al. reported the results of a study of methods for treating patients who had advanced heart failure due to ischemic or nonischemic cardiomyopathies. A total of 1520 patients were randomly assigned in a 1:2:2 ratio to receive optimal pharmacologic therapy alone or in combination with either a pacemaker or a pacemaker–defibrillator combination. The patients were then observed until they died or were hospitalized for any cause.
a. How many treatments were there?
b. Which group would be considered the control group?
c. How many treatment groups were there? Which treatments did they receive?
d. How many patients were in each of the three groups studied?
e. Explain how a table of random numbers or a random-number generator could be used to divide the patients into the three groups.

*In Exercises **1.92–1.97**, we present descriptions of designed experiments. In each case, identify the*
a. *experimental units.*
b. *response variable.*
c. *factor(s).*
d. *levels of each factor.*
e. *treatments.*

1.92 Increasing Unit Sales. Supermarkets are interested in strategies to increase temporarily the unit sales of a product. In one study, researchers compared the effect of display type and price on unit sales for a particular product. The following display types and pricing schemes were employed.

- Display types: normal display space interior to an aisle, normal display space at the end of an aisle, and enlarged display space.
- Pricing schemes: regular price, reduced price, and cost.

1.93 Highway Signs. A driver's ability to detect highway signs is an important consideration in highway safety. In his dissertation, *Highway Construction Safety and the Aging Driver*, S. Younes investigated the distance at which drivers can first detect highway caution signs. This distance is called the *detection distance.* Younes analyzed the effect that sign size and sign material have on detection distance. Drivers were randomly assigned to one combination of sign size (small, medium, and large), and sign material (1, 2, and 3). Each driver covered the same stretch of highway at a constant speed during the same time of day, and the detection distance, in feet, was determined for the driver's assigned caution sign.

1.94 Oat Yield and Manure. In a classic study, described by F. Yates in *The Design and Analysis of Factorial Experiments* (Commonwealth Bureau of Soils, Technical Communication No. 35), the effect on oat yield was compared for three different varieties of oats and four different concentrations of manure (0, 0.2, 0.4, and 0.6 cwt per acre).

1.95 The Lion's Mane. In a study by P. M. West titled "The Lion's Mane" (*American Scientist*, Vol. 93, No. 3, pp. 226–236), the effects of the mane of a male lion as a signal of quality to mates and rivals was explored. Four life-sized dummies of male lions provided a tool for testing female response to the unfamiliar lions whose manes varied by length (long or short) and color (blonde or dark). The female lions were observed to see whether they approached each of the four life-sized dummies.

1.96 Sexual Signals. In a study by A. Elliot et al., titled "Women's Use of Red Clothing as a Sexual Signal in Intersexual Interaction" (*Journal of Experimental Social Psychology*, Vol. 49, Issue 3, pp. 599–602), women were studied to determine the effect of apparel-color choice based on perceived attractiveness and gender of a new acquaintance. The following experiment is based on the researchers' investigation. Women are randomly assigned to be told that they would be conversing with an attractive male, an unattractive male, an attractive female, or an unattractive female. The women must wear a standardized shirt before the conversation, but can choose between red, green, and blue.

1.97 Dexamethasone and IQ. In the paper "Outcomes at School Age After Postnatal Dexamethasone Therapy for Lung Disease of Prematurity" (*New England Journal of Medicine*, Vol. 350, No. 13, pp. 1304–1313), T. Yeh et al. studied the outcomes at school age in children who had participated in a double-blind, placebo-controlled trial of early postnatal dexamethasone therapy for the prevention of chronic lung disease of prematurity. One result reported in the study was that, at school age, the control group of 74 children had an average IQ score of 84.4, whereas the dexamethasone group of 72 children had an average IQ score of 78.

1.98 Lifetimes of Flashlight Batteries. Two different options are under consideration for comparing the lifetimes of four brands of flashlight battery, using 20 flashlights.
a. One option is to randomly divide 20 flashlights into four groups of 5 flashlights each and then randomly assign each group to use a different brand of battery. Would this statistical design be a completely randomized design or a randomized block design? Explain your answer.
b. Another option is to use 20 flashlights—five different brands of 4 flashlights each—and randomly assign the 4 flashlights of each brand to use a different brand of battery. Would this statistical design be a completely randomized design or a randomized block design? Explain your answer.

1.99 Dental Hygiene: Which Toothbrush? In an experiment reported by J. Singer and D. Andrade in the article "Regression Models for the Analysis of Pretest/Posttest Data" (*Biometrics*, Vol. 53, pp. 729–735), the effect of using either a conventional or experimental (hugger) toothbrush was investigated. Twelve female and 12 male preschoolers were selected. Within each gender group, six were randomly assigned to the conventional toothbrush and the remaining six to the experimental toothbrush. After each subject brushed with the assigned toothbrush, a dental plaque index was measured. The higher the dental plaque index, the greater was the amount of plaque on an individual's teeth.
a. Is the statistical design described here a completely randomized design or a randomized block design? Explain your answer.
b. If the statistical design is a randomized block design, what are the blocks?

Extending the Concepts and Skills

1.100 The Salk Vaccine. In Exercise 1.17 on page 8, we discussed the Salk vaccine experiment. The experiment utilized a technique called *double-blinding* because neither the children nor the doctors involved knew which children had been given the vaccine and which had been given placebo. Explain the advantages of using double-blinding in the Salk vaccine experiment.

1.101 In sampling from a population, state which type of sampling design corresponds to each of the following experimental designs:
a. Completely randomized design
b. Randomized block design

CHAPTER IN REVIEW

You Should Be Able to

1. classify a statistical study as either descriptive or inferential.
2. identify the population and the sample in an inferential study.
3. explain the difference between an observational study and a designed experiment.
4. classify a statistical study as either an observational study or a designed experiment.
5. explain what is meant by a representative sample.
6. describe simple random sampling.
7. use a table of random numbers to obtain a simple random sample.

*8. describe systematic random sampling, cluster sampling, and stratified sampling.

*9. state the three basic principles of experimental design.

*10. identify the treatment group and control group in a study.

*11. identify the experimental units, response variable, factor(s), levels of each factor, and treatments in a designed experiment.

*12. distinguish between a completely randomized design and a randomized block design.

Key Terms

blocks,* *28*
census, *9*
cluster sampling,* *18*
completely randomized design,* *27*
control,* *26*
control group,* *26*
descriptive statistics, *3*
designed experiment, *5*
experimental unit,* *26*
experimentation, *9*
factor,* *26*
inferential statistics, *3*
levels,* *26*
multistage sampling,* *23*
observational study, *5*
population, *3*
probability sampling, *10*
proportional allocation,* *21*
randomization,* *26*
randomized block design,* *28*
random-number generator, *13*
replication,* *26*
representative sample, *10*
response variable,* *26*
sample, *3*
sampling, *9*
simple random sample, *10*
simple random sampling, *10*
simple random sampling with replacement (SRSWR), *10*
simple random sampling without replacement (SRS), *10*
strata,* *21*
stratified random sampling with proportional allocation,* *21*
stratified sampling,* *21*
subject,* *26*
systematic random sampling,* *17*
table of random numbers, *11*
treatment,* *26*
treatment group,* *26*

REVIEW PROBLEMS

Understanding the Concepts and Skills

1. In a newspaper or magazine, or on the Internet, find an example of
a. a descriptive study. **b.** an inferential study.

2. Almost any inferential study involves aspects of descriptive statistics. Explain why.

3. Regarding observational studies and designed experiments:
a. Describe each type of statistical study.
b. With respect to possible conclusions, what important difference exists between these two types of statistical studies?

4. Before planning and conducting a study to obtain information, what should be done?

5. Explain the meaning of
a. a representative sample.
b. probability sampling.
c. simple random sampling.

6. Which of the following sampling procedures involve the use of probability sampling?
a. A college student is hired to interview a sample of voters in her town. She stays on campus and interviews 100 students in the cafeteria.
b. A pollster wants to interview 20 gas station managers in Baltimore. He posts a list of all such managers on his wall, closes his eyes, and tosses a dart at the list 20 times. He interviews the people whose names the dart hits.

***7.** Describe each of the following sampling methods and indicate conditions under which each is appropriate.
a. Systematic random sampling
b. Cluster sampling
c. Stratified random sampling with proportional allocation

***8.** Identify and explain the significance of the three basic principles of experimental design.

Applying the Concepts and Skills

9. Baseball Scores. From ESPN MLB Scoreboard, we obtained the following major league baseball scores for August 14, 2013. Is this study descriptive or inferential? Explain your answer.

Games in AL stadiums	Games in NL stadiums
Indians 9, Twins 8	Reds 5, Cubs 0
Tigers 6, White Sox 4	Padres 2, Rockies 4
Marlins 5, Royals 2	Orioles 4, Diamondbacks 5
Angels 3, Yankees 11	Giants 5, Nationals 6
Red Sox 3, Blue Jays 4	Phillies 3, Braves 6
Mariners 4, Rays 5	Pirates 5, Cardinals 1
Brewers 4, Rangers 5	Mets 4, Dodgers 5
Astros 2, Athletics 1	

10. Working Lottery Winners. In a national poll taken on August 7–11, 2013, by Gallup, Inc., 1039 adults who were employed full or part time were asked the following question: "If you won 10 million dollars in the lottery, would you continue to work, or would you stop working?" Sixty-eight percent of the respondents said that they would continue to work.
a. Is this study descriptive or inferential? Explain your answer.
b. The title of the article discussing the survey was "In U.S., Most Would Still Work Even if They Won Millions." Is the statement in quotes here descriptive or inferential? Explain your answer.

11. British Backpacker Tourists. Research by G. Visser and C. Barker in "A Geography of British Backpacker Tourists in South Africa" (*Geography*, Vol. 89, No. 3, pp. 226–239) reflected on the impact of British backpacker tourists visiting South Africa. A sample of British backpackers was interviewed. The information obtained from the sample was used to construct the following table for the age distribution of all British backpackers. Classify this study as descriptive or inferential, and explain your answer.

Age (yr)	Percentage
Less than 21	9
21–25	46
26–30	27
31–35	10
36–40	4
Over 40	4

12. Peanut Allergies. In the article "Food Allergy Advice May Be Peanuts" (*Science News*, Vol. 174, No. 12, pp. 8–9), N. Seppa reports that early exposure to peanuts seems to lessen the risk of nut allergy. Of 4000 Jewish children sampled in Britain, 1.85% had peanut allergies; and of 4600 Jewish children sampled in Israel, where early peanut consumption is more common, 0.17% had peanut allergies. The researcher chose Jewish children in both countries to limit genetic differences between groups.
a. Is this study descriptive or inferential?
b. Is this study observational or experimental?

13. Persistent Poverty and IQ. An article appearing in an issue of the *Arizona Republic* reported on a study conducted by G. Duncan of the University of Michigan. According to the report, "Persistent poverty during the first 5 years of life leaves children with IQs 9.1 points lower at age 5 than children who suffer no poverty during that period. . . ." Is this statistical study an observational study or is it a designed experiment? Explain your answer.

14. Wasp Hierarchical Status. In an issue of *Discover* (Vol. 26, No. 2, pp. 10–11), J. Netting described the research of E. Tibbetts of the University of Arizona in the article, "The Kind of Face Only a Wasp Could Trust." Tibbetts found that wasps signal their strength and status with the number of black splotches on their yellow faces, with more splotches denoting higher status. Tibbetts decided to see if she could cheat the system. She painted some of the insects' faces to make their status appear higher or lower than it really was. She then placed the painted wasps with a group of female wasps to see if painting the faces altered their hierarchical status. Was this investigation an observational study or a designed experiment? Justify your answer.

15. Incomes of College Students' Parents. A researcher wants to estimate the average income of parents of college students. To accomplish that, he surveys a sample of 250 students at Yale. Is this a representative sample? Explain your answer.

16. On-Time Airlines. From the *FlightStats On-time Performance Report Summary*, we found that, in July 2013, the top five North American airlines in terms of percentage of on-time arrivals were Hawaiian (H), Horizon (Z), Compass (C), Alaska (A), and Jazz (J).
a. List the 10 possible samples (without replacement) of size 3 that can be obtained from the population of five airlines. Use the parenthetical abbreviations in your list.
b. If a simple random sampling procedure is used to obtain a sample of three of these five airlines, what are the chances that it is the first sample on your list in part (a)? the second sample? the tenth sample?
c. Describe three methods for obtaining a simple random sample of three of these five airlines.
d. Use one of the methods that you described in part (c) to obtain a simple random sample of three of these five airlines.

17. Top North American Athletes. As part of ESPN's SportsCenturyRetrospective, a panel chosen by ESPN ranked the top 100 North American athletes of the twentieth century. For a class project, you are to obtain a simple random sample of 15 of these 100 athletes and briefly describe their athletic feats.
a. Explain how you can use Table I in Appendix A to obtain the simple random sample.
b. Starting at the three-digit number in line number 10 and column numbers 7–9 of Table I, read down the column, up the next, and so on, to find 15 numbers that you can use to identify the athletes to be considered.
c. If you have access to a random-number generator, use it to obtain the required simple random sample.

18. QuickVote. TalkBack Live, a production of CNN, conducted on-line surveys on various issues. One survey, called a *QuickVote*, of 680 people asked "Would you vote for a third-party candidate?" Of the 680 people surveyed, 608 (89.4%) responded "yes" and 72 (10.6%) responded "no." Beneath the vote tally, the following statement regarding the sampling procedure was found:

> This QuickVote is not scientific and reflects the opinions of only those Internet users who have chosen to

participate. The results cannot be assumed to represent the opinions of Internet users in general, nor the public as a whole. The QuickVote sponsor is not responsible for content, functionality or the opinions expressed therein.

Discuss the preceding statement in light of what you have learned in this chapter.

19. Leisure Activities and Dementia. An article appearing in the *Los Angeles Times* discussed the study "Leisure Activities and the Risk of Dementia in the Elderly" (*New England Journal of Medicine*, Vol. 348, pp. 2508–2516) by J. Verghese et al. The article in the *Times*, titled "Crosswords Reduce Risk of Dementia," contained the following statement: "Elderly people who frequently read, do crossword puzzles, practice a musical instrument or play board games cut their risk of Alzheimer's and other forms of dementia by nearly two-thirds compared with people who seldom do such activities." Comment on the statement in quotes, keeping in mind the type of study for which causation can be reasonably inferred.

20. Hepatitis B and Pancreatic Cancer. The article "Study Links Hepatitis B and Cancer of Pancreas" by D. Grady, appeared in the September 29, 2008 issue of the *NewYork Times*. It reported that, for the first time, a study showed that people with pancreatic cancer are more likely than those without the disease to have been infected with the hepatitis B virus. The study by M. Hassan et al., titled "Association Between Hepatitis B Virus and Pancreatic Cancer" (*Journal of Clinical Oncology*, Vol. 26, No. 28, pp. 4557–4562) compared 476 people who had pancreatic cancer with 879 healthy control subjects. All were tested to see whether they had ever been infected with the viruses that cause hepatitis B or hepatitis C. The results were that no connection was found to hepatitis C, but the cancer patients were twice as likely as the healthy subjects to have had hepatitis B. The researchers noted, however, that "... while the study showed an association, it did not prove cause and effect. More work is needed to determine whether the virus really can cause pancreatic cancer." Explain the validity of the statement in quotes.

***21. Top North American Athletes.** Refer to Problem 17.

a. Use systematic random sampling to obtain a sample of 15 athletes.

b. In this case, is systematic random sampling an appropriate alternative to simple random sampling? Explain your answer.

***22. Water Quality.** In the article "Randomized Stratified Sampling Methodology for Water Quality in Distribution Systems" (*Journal of Water Resources Planning and Management*, Vol. 130, Issue 4, pp. 330–338), V. Speight et al. proposed the method of stratified sampling to collect water samples for water-quality testing. The following table separates the Durham, North Carolina water distribution system into strata based on distance from the nearest treatment plant.

Distance from treatment center	Stratum size
Less than 1.5 miles	1310
1.5–less than 3.0 miles	3166
3.0–less than 4.5 miles	2825
4.5–less than 6.0 miles	1593
6.0–less than 7.5 miles	1350
7.5 miles or greater	1463

Use the table to design a procedure for obtaining a stratified sample (with proportional allocation) of 80 water samples from Durham. *Hint:* Refer to the remarks about strata sample size on pages 22–23 following Example 1.14.

***23. AVONEX and MS.** An issue of *Inside MS* contained an article describing AVONEX (interferon beta-1a), a drug used in the treatment of relapsing forms of multiple sclerosis (MS). Included in the article was a report on "... adverse events and selected laboratory abnormalities that occurred at an incidence of 2% or more among the 158 multiple sclerosis patients treated with 30 mcg of AVONEX once weekly by IM injection." In the study, 158 patients took AVONEX and 143 patients were given placebo.

a. Is this study observational or is it a designed experiment?

b. Identify the treatment group, control group, and treatments.

***24. Plant Density and Tomato Yield.** In "Effects of Plant Density on Tomato Yields in Western Nigeria" (*Experimental Agriculture*, Vol. 12(1), pp. 43–47), B. Adelana reported on the effect of tomato variety and planting density on yield. Four tomato varieties (Harvester, Pusa Early Dwarf, Ife No. 1, and Ibadan Local) were grown at four densities (10,000, 20,000, 30,000, and 40,000 plants/ha). Identify the

a. experimental units.
b. response variable.
c. factor(s).
d. levels of each factor.
e. treatments.

***25. Child-Proof Bottles.** Designing medication packaging that resists opening by children, but yields readily to adults, presents numerous challenges. In the article "Painful Design" (*American Scientist*, Vol. 93, No. 2, pp. 113–118), H. Petroski examined the packaging used for Aleve, a brand of pain reliever. Three new container designs were given to a panel of children aged 42 months to 51 months. For each design, the children were handed the bottle, shown how to open it, and then left alone with it. If more than 20% of the children succeeded in opening the bottle on their own within 10 minutes, even if by using their teeth, the bottle failed to qualify as child resistant. Identify the

a. experimental units.
b. response variable.
c. factor(s).
d. levels of each factor.
e. treatments.

***26. Doughnuts and Fat.** A classic study, conducted in 1935 by B. Lowe at the Iowa Agriculture Experiment Station, analyzed differences in the amount of fat absorbed by doughnuts in cooking with four different fats. For the experiment, 24 batches of doughnuts were randomly divided into four groups of 6 batches each. The four groups were then randomly assigned to the four fats. What type of statistical design was used for this study? Explain your answer.

***27. Comparing Gas Mileages.** An experiment is to be conducted to compare four different brands of gasoline for gas mileage.

a. Suppose that you randomly divide 24 cars into four groups of 6 cars each and then randomly assign the four groups to the four brands of gasoline, one group per brand. Is this experimental design a completely randomized design or a randomized block design? If it is the latter, what are the blocks?

b. Suppose, instead, that you use six different models of cars whose varying characteristics (e.g., weight and horsepower) affect gas mileage. Four cars of each model are randomly assigned to the four different brands of gasoline. Is this experimental design a completely randomized design or a randomized block design? If it is the latter, what are the blocks?

c. Which design is better, the one in part (a) or the one in part (b)? Explain your answer.

FOCUSING ON DATA ANALYSIS

UWEC UNDERGRADUATES

The file named Focus.txt in the Focus Database section of the WeissStats site contains information on the undergraduate students at the University of Wisconsin - Eau Claire (UWEC). Those students constitute the population of interest in the *Focusing on Data Analysis* sections that appear at the end of each chapter of the book.†

Thirteen variables are considered. Table 1.9 lists the variables and the names used for those variables in the data files. We call the database of information for those variables the **Focus database.**

Also provided in the Focus Database section is a file called FocusSample.txt that contains data on the same 13 variables for a simple random sample of 200 of the undergraduate students at UWEC. Those 200 students constitute a sample that can be used for making statistical inferences in the *Focusing on Data Analysis* sections. We call this sample data the **Focus sample.**

Large data sets are almost always analyzed by computer, and that is how you should handle both the Focus database and the Focus sample. We have supplied the Focus database and Focus sample in several file formats in the Focus Database section of the WeissStats site.

If you use a statistical software package for which we have not supplied a Focus database file, you should (1) input the file Focus.txt into that software, (2) ensure that the variables are named as indicated in Table 1.9, and (3) save the worksheet to a file named Focus in the format suitable to your software, that is, with the appropriate file extension. Then, any time that you want to analyze the Focus database, you can simply retrieve your Focus worksheet. These same remarks apply to the Focus sample, as well as to the Focus database.

TABLE 1.9
Variables and variable names for the Focus database

Variable	Variable name
Sex	SEX
High school percentile	HSP
Cumulative GPA	GPA
Age	AGE
Total earned credits	CREDITS
Classification	CLASS
School/college	COLLEGE
Primary major	MAJOR
Residency	RESIDENCY
Admission type	TYPE
ACT English score	ENGLISH
ACT math score	MATH
ACT composite score	COMP

†We have restricted attention to those undergraduate students at UWEC with complete records for all the variables under consideration.

CASE STUDY DISCUSSION

TOP FILMS OF ALL TIME

At the beginning of this chapter, we discussed the results of a survey by the American Film Institute (AFI). Now that you have learned some of the basic terminology of statistics, we want you to examine that survey in greater detail.

Answer each of the following questions pertaining to the survey. In doing so, you may want to reread the description of the survey given on pages 1–2.

a. Identify the population.
b. Identify the sample.
c. Is the sample representative of the population of all U.S. moviegoers? Explain your answer.
d. Consider the following statement: "Among the 1500 film artists, critics, and historians polled by AFI, the top-ranking film was *Citizen Kane.*" Is this statement descriptive or inferential? Explain your answer.
e. Suppose that the statement in part (d) is changed to the following statement: "Based on the AFI poll, *Citizen Kane* is the top-ranking film among all film artists, critics, and historians." Is this statement descriptive or inferential? Explain your answer.

BIOGRAPHY

FLORENCE NIGHTINGALE: LADY OF THE LAMP

Florence Nightingale (1820–1910), the founder of modern nursing, was born in Florence, Italy, into a wealthy English family. In 1849, over the objections of her parents, she entered the Institution of Protestant Deaconesses at Kaiserswerth, Germany, which "...trained country girls of good character to nurse the sick."

The Crimean War began in March 1854 when England and France declared war on Russia. After serving as superintendent of the Institution for the Care of Sick Gentlewomen in London, Nightingale was appointed by the English Secretary of State at War, Sidney Herbert, to be in charge of 38 nurses who were to be stationed at military hospitals in Turkey.