SEVENTH EDITION

Elementary STATISTICS

Picturing the World

Ron Larson Betsy Farber

Pearson

SEVENTH EDITION

Elementary Statistics

PICTURING THE WORLD

Ron Larson
The Pennsylvania State University
The Behrend College

Betsy Farber
Bucks County Community College

Pearson

Director, Portfolio Management: *Deirdre Lynch*
Senior Courseware Portfolio Manager: *Patrick Barbera*
Editorial Assistant: *Morgan Danna*
Content Producer: *Tamela Ambush*
Managing Producer: *Karen Wernholm*
Media Producer: *Audra Walsh*
Manager, Courseware QA: *Mary Durnwald*
Manager, Content Development: *Robert Carroll*
Product Marketing Manager: *Emily Ockay*
Field Marketing Manager: *Andrew Noble*
Marketing Assistants: *Shannon McCormack, Erin Rush*
Senior Author Support/Technology Specialist: *Joe Vetere*
Manager, Rights and Permissions: *Gina Cheselka*
Manufacturing Buyer: *Carol Melville, LSC Communications*
Text Design, Cover Design: *Cenveo Publisher Services*
Production Coordination, Composition, and Illustrations: *Larson Texts, Inc.*

Library of Congress Cataloging-in-Publication Data

Names: Larson, Ron, 1941- author. | Farber, Elizabeth, author.
Title: Elementary statistics : picturing the world / Ron Larson, The Pennsylvania State University, The Behrend College, Betsy Farber, Bucks County Community College.
Description: 7th edition. | Boston, MA : Pearson Education, Inc., 2017.
Identifiers: LCCN 2017040872| ISBN 9780134683416 | ISBN 9780134683584
Subjects: LCSH: Statistics--Textbooks. | Mathematical statistics--Textbooks.
Classification: LCC QA276.12 .L373 2017 | DDC 519.5--dc23
LC record available at *https://lccn.loc.gov/2017040872*

1—17

Pearson

ISBN 10: 0-13-468341-2
ISBN 13: 978-0-13-468341-6

CONTENTS

PART 2 PROBABILITY AND PROBABILITY DISTRIBUTIONS

5 Normal Probability Distributions 232

PART 3 STATISTICAL INFERENCE

6 Confidence Intervals 296

PART 4 MORE STATISTICAL INFERENCE

9 Correlation and Regression 468

10 Chi-Square Tests and the *F*-Distribution 524

11 Nonparametric Tests (Web Only)*

Where You've Been Where You're Going

* Available at *www.pearson.com/math-stats-resources* and in MyLab Statistics.

Appendices

PREFACE

Welcome to *Elementary Statistics: Picturing the World*, Seventh Edition. You will find that this textbook is written with a balance of rigor and simplicity. It combines step-by-step instruction, real-life examples and exercises, carefully developed features, and technology that makes statistics accessible to all.

I am grateful for the overwhelming acceptance of the first six editions. It is gratifying to know that my vision of combining theory, pedagogy, and design to exemplify how statistics is used to picture and describe the world has helped students learn about statistics and make informed decisions.

What's New in this Edition

The goal of the Seventh Edition was a thorough update of the key features, examples, and exercises:

Examples This edition has 213 examples, over 60% of which are new or revised. Also, several of the examples now show an alternate solution or a check using technology.

Technology Examples In addition to showing screen displays from Minitab®, Excel®, and the TI-84 Plus, this edition also shows screen displays from StatCrunch®.

Try It Yourself Over 40% of the 213 Try It Yourself exercises are new or revised.

Picturing the World Over 50% of these are new or revised.

Tech Tips New to this edition are technology tips that appear in most sections. These tips show how to use Minitab, Excel, the TI-84 Plus, or StatCrunch to solve a problem.

Exercises Over 40% of the more than 2300 exercises are new or revised.

Extensive Chapter Feature Updates Over 60% of the following key features are new or revised, making this edition fresh and relevant to today's students:

- Where You've Been and Where You're Going
- Uses and Abuses: Statistics in the Real World
- Real Statistics–Real Decisions: Putting it all together
- Chapter Technology Project

Revised Content Here is a summary of the content changes.

- **Section 1.1** now has more discussion about populations and samples, how to identify them, and their relationships to parameters and statistics. Also, the Venn Diagrams have been redrawn to use clearer labeling to help students distinguish between a population and a sample.
- **In Section 1.3,** the figure depicting systemic sampling has been redrawn to more clearly depict the sampling process.
- **Section 2.1** now has more discussion of class widths and open-ended classes. Also, a figure showing a histogram and its corresponding frequency polygon was added after Example 4.
- **In Section 2.4,** Example 9 was rewritten to explain the use of an open-ended class.
- **Section 2.5** now has a Study Tip discussing outliers and modified box-and-whisker plots. On pages 124 and 125, students are shown how to create modified box-and-whisker plots using technology.
- **In Section 3.1,** the solutions to the examples were rewritten to explain why a formula was chosen to find a probability.
- **In Chapter 5,** in addition to using a table, examples were revised and Tech Tips were added to show how to find areas or probabilities using technology.
- **In Chapter 6,** in addition to using a table, examples were revised and Tech Tips were added to show how to find critical values using technology. Also, the exercises in this chapter were revised to ask more conceptual questions.
- **Section 6.2** now has more explanation about why the t-distribution is needed when σ is unknown. Also, the flowchart on page 314 was revised to illustrate when it is not possible to use the normal distribution or the t-distribution to construct a confidence interval.
- **In Chapters 7–9,** in addition to using a table, examples were revised and Tech Tips were added to show how to find P-values and critical values using technology.
- **Section 8.2** now shows the formula for the number of degrees of freedom for the t-test often used by technology.
- **In Section 9.1,** the requirements to use a correlation coefficient r to make an inference about a population have been revised.

Features of the Seventh Edition

Guiding Student Learning

Where You've Been and Where You're Going Each chapter begins with a two-page visual description of a real-life problem. *Where You've Been* connects the chapter to topics learned in earlier chapters. *Where You're Going* gives students an overview of the chapter.

What You Should Learn Each section is organized by learning objectives, presented in everyday language in *What You Should Learn.* The same objectives are then used as subsection titles throughout the section.

Definitions and Formulas are clearly presented in easy-to-locate boxes. They are often followed by **Guidelines**, which explain *In Words* and *In Symbols* how to apply the formula or understand the definition.

Margin Features help reinforce understanding:

- **Study Tips** show how to read a table, interpret a result, help drive home an important interpretation, or connect different concepts.
- **Tech Tips** show how to use Minitab, Excel, the TI-84 Plus, or StatCrunch to solve a problem.
- **Picturing the World** is a "mini case study" in each section that illustrates the important concept or concepts of the section. Each Picturing the World concludes with a question and can be used for general class discussion or group work. The answers to these questions are included in the *Annotated Instructor's Edition.*

Examples and Exercises

Examples Every concept in the text is clearly illustrated with one or more step-by-step examples. Most examples have an interpretation step that shows the student how the solution may be interpreted within the real-life context of the example and promotes critical thinking and writing skills. Each example, which is numbered and titled for easy reference, is followed by a similar exercise called **Try It Yourself** so students can immediately practice the skill learned. The answers to these exercises are in the back of the book, and the worked-out solutions are in the *Student's Solutions Manual.*

Technology Examples Many sections contain an example that shows how technology can be used to calculate formulas, perform tests, or display data. Screen displays from Minitab, Excel, the TI-84 Plus, and StatCrunch are shown. Additional screen displays are presented at the ends of selected chapters, and detailed instructions are given in separate technology manuals available with the book.

Exercises The exercises give students practice in performing calculations, making decisions, providing explanations, and applying results to a real-life setting. The section exercises are divided into three parts:

- **Building Basic Skills and Vocabulary** are short answer, true or false, and vocabulary exercises carefully written to nurture student understanding.
- **Using and Interpreting Concepts** are skill or word problems that move from basic skill development to more challenging and interpretive problems.
- **Extending Concepts** go beyond the material presented in the section. They tend to be more challenging and are not required as prerequisites for subsequent sections.

Technology Answers Answers in the back of the book are found using calculations by hand and by tables. Answers found using technology (usually the TI-84 Plus) are also included when there are discrepancies due to rounding.

Review and Assessment

Chapter Summary Each chapter concludes with a Chapter Summary that answers the question *What did you learn?* The objectives listed are correlated to Examples in the section as well as to the Review Exercises.

Chapter Review Exercises A set of Review Exercises follows each Chapter Summary. The order of the exercises follows the chapter organization. Answers to all odd-numbered exercises are given in the back of the book.

Chapter Quizzes Each chapter has a Chapter Quiz. The answers to all quiz questions are provided in the back of the book. For additional help, see the step-by-step video solutions available in MyLab Statistics.

Chapter Tests Each chapter has a Chapter Test. The questions are in random order. The answers to all test questions are provided in the *Annotated Instructor's Edition.*

Cumulative Review There is a Cumulative Review after Chapters 2, 5, 8, and 10. Exercises in the Cumulative Review are in random order and may incorporate multiple ideas. Answers to all odd-numbered exercises are given in the back of the book.

Statistics in the Real World

Uses and Abuses: Statistics in the Real World Each chapter discusses how statistical techniques should be used, while cautioning students about common abuses. The discussion includes ethics, where appropriate. Exercises help students apply their knowledge.

Applet Activities Selected sections contain activities that encourage interactive investigation of concepts in the lesson with exercises that ask students to draw conclusions. The applets are available in MyLab Statistics and at **www.pearson.com/math-stats-resources.**

Chapter Case Study Each chapter has a full-page Case Study featuring actual data from a real-world context and questions that illustrate the important concepts of the chapter.

Real Statistics – Real Decisions: Putting it all together This feature encourages students to think critically and make informed decisions about real-world data. Exercises guide students from interpretation to drawing of conclusions.

Chapter Technology Project Each chapter has a Technology project using Minitab, Excel, and the TI-84 Plus that gives students insight into how technology is used to handle large data sets or real-life questions.

Continued Strong Pedagogy from the Sixth Edition

Versatile Course Coverage The table of contents was developed to give instructors many options. For instance, the *Extending Concepts* exercises, applet activities, Real Statistics—Real Decisions, and Uses and Abuses provide sufficient content for the text to be used in a two-semester course. More commonly, I expect the text to be used in a three-credit semester course or a four-credit semester course that includes a lab component. In such cases, instructors will have to pare down the text's 46 sections.

Graphical Approach As with most introductory statistics texts, this text begins the descriptive statistics chapter (Chapter 2) with a discussion of different ways to display data graphically. A difference between this text and many others is that **it continues to incorporate the graphical display of data throughout the text.** For example, see the use of stem-and-leaf plots to display data on page 387. This emphasis on graphical displays is beneficial to all students, especially those utilizing visual learning strategies.

Balanced Approach The text strikes a **balance among computation, decision making, and conceptual understanding.** I have provided many Examples, Exercises, and Try It Yourself exercises that go beyond mere computation.

Variety of Real-Life Applications I have chosen real-life applications that are representative of the majors of students taking introductory statistics courses. I want statistics to come alive and appear relevant to students so they understand the importance of and rationale for studying statistics. I wanted the applications to be **authentic**—but they also need to be **accessible.** See the Index of Applications on page xvi.

Data Sets and Source Lines The data sets in the book were chosen for interest, variety, and their ability to illustrate concepts. Most of the **250-plus data sets** contain real data with

source lines. The remaining data sets contain simulated data that are representative of real-life situations. All data sets containing 20 or more entries are available in a variety of formats in MyLab™ Statistics or at **www.pearson.com/math-stats-resources.** In the exercise sets, the data sets that are available electronically are indicated by the icon  .

Flexible Technology Although most formulas in the book are illustrated with "hand" calculations, I assume that most students have access to some form of technology, such as Minitab, Excel, StatCrunch, or the TI-84 Plus. Because technology varies widely, the text is flexible. **It can be used in courses with no more technology than a scientific calculator—or it can be used in courses that require sophisticated technology tools.** Whatever your use of technology, I am sure you agree with me that the goal of the course is not computation. Rather, it is to help students gain an understanding of the basic concepts and uses of statistics.

Prerequisites Algebraic manipulations are kept to a minimum—often I display informal versions of formulas using words in place of or in addition to variables.

Choice of Tables My experience has shown that students find a **cumulative distribution function** (CDF) table easier to use than a "0-to-z" table. Using the CDF table to find the area under the standard normal curve is a topic of Section 5.1 on pages 237–241. Because some teachers prefer to use the "0-to-z" table, an alternative presentation of this topic is provided in Appendix A.

Page Layout Statistics instruction is more accessible when it is carefully formatted on each page with a consistent open layout. This text is the first college-level statistics book to be written so that, when possible, its features are not split from one page to the next. Although this process requires extra planning, the result is a presentation that is clean and clear.

Meeting the Standards

MAA, AMATYC, NCTM Standards This text answers the call for a **student-friendly text that emphasizes the uses of statistics.** My goal is not to produce statisticians but to produce informed consumers of statistical reports. For this reason, I have included exercises that require students to interpret results, provide written explanations, find patterns, and make decisions.

GAISE Recommendations Funded by the American Statistical Association, the Guidelines for Assessment and Instruction in Statistics Education (GAISE) Project developed six recommendations for teaching introductory statistics in a college course. These recommendations are:

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

The examples, exercises, and features in this text embrace all of these recommendations.

Technology Resources

MyLab Statistics Online Course (access code required)

Used by nearly one million students a year, MyLab Statistics is the world's leading online program for teaching and learning statistics. MyLab Statistics delivers assessment, tutorials, and multimedia resources that provide engaging and personalized experiences for each student, so learning can happen in any environment.

Personalized Learning Not every student learns the same way or at the same rate. Personalized learning in the MyLab gives instructors the flexibility to incorporate the approach that best suits the needs of their course and students.

- Based on their performance on a quiz or test, **personalized homework** allows students to focus on just the topics they have not yet mastered.
- With **Companion Study Plan Assignments** you can assign the Study Plan as a prerequisite to a test or quiz, guiding students through the concepts they need to master.

Preparedness Preparedness is one of the biggest challenges in statistics courses. Pearson offers a variety of content and course options to support students with just-in-time remediation and key-concept review as needed.

- **Redesign-Ready Course Options** Many new course models have emerged in recent years, as institutions "redesign" to help improve retention and results. At Pearson, we're focused on tailoring solutions to support your plans and programs.
- **Getting Ready for Statistics Questions** This question library contains more than 450 exercises that cover the relevant developmental math topics for a given section. These can be made available to students for extra practice or assigned as a prerequisite to other assignments.

Conceptual Understanding Successful students have the ability to apply their statistical ideas and knowledge to new concepts and real-world situations. Providing frequent opportunities for data analysis and interpretation helps students develop the 21st century skills that they need in order to be successful in the classroom and workplace.

- **Conceptual Question Library** There are 1,000 questions in the Assignment Manager that require students to apply their statistical understanding.
- **Modern statistics is practiced with technology,** and MyLab Statistics makes learning and using software programs seamless and intuitive. Instructors 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.

Motivation Students are motivated to succeed when they are engaged in the learning experience and understand the relevance and power of statistics.

- **Exercises with Immediate Feedback** Homework and practice exercises in MyLab Statistics regenerate algorithmically to give students unlimited opportunity for

practice and mastery. Instructors can choose from the many exercises available for the author's approach—or even choose additional exercises from other MyLab Statistics courses. Most exercises include learning aids, such as guided solutions, sample problems, extra help at point-of-use, and immediate feedback when students enter incorrect answers.

- Instructors can create, import, and manage online homework assignments, quizzes, and tests—or start with sample assignments—all of which are automatically graded, allowing instructors to spend less time grading, and more time teaching.

Data & Analytics MyLab Statistics provides resources to help instructors assess and improve student results. A comprehensive gradebook with enhanced reporting functionality makes it easier for instructors to manage courses efficiently.

- **Reporting Dashboard** Instructors can view, analyze, and report learning outcomes, gaining the information they need to keep our students on track. Available via the Gradebook and fully mobile-ready, the Reporting Dashboard presents student performance data at the class, section, and program levels in an accessible, visual manner. Its finegrain reports allow instructors and administrators to compare performance across different courses, across individual sections and within each course.
- **Item Analysis** Instructors can track class-wide understanding of particular exercises in order to refine your class lectures or adjust the course/department syllabus. Just-in-time teaching has never been easier.

Accessibility Pearson works continuously to ensure our products are as accessible as possible to all students. We are working toward achieving WCAG 2.0 Level AA and Section 508 standards, as expressed in the Pearson Guidelines for Accessible Educational Web Media, **www.pearson.com/mylab/statistics/accessibility.**

StatCrunch

Integrated directly into MyLab Statistics, StatCrunch® is powerful web-based statistical software that allows users to perform complex analyses, share data sets, and generate compelling reports of their data.

- **Collect** Users can upload their own data to StatCrunch or search a large library of publicly shared data sets, spanning almost any topic of interest. A Featured Data page houses the best data sets, making it easy for instructors to use current data in their course. Data sets from the text and from online homework exercises can also be accessed and analyzed in StatCrunch. An online survey tool allows users to quickly collect data via web-based surveys.
- **Crunch** A full range of numerical and graphical methods allow users to analyze and gain insights from any data set. Interactive graphics help users understand statistical concepts, and are available for export to enrich reports with visual representations of data.
- **Communicate** Reporting options help users create a wide variety of visually-appealing representations of their data.

StatCrunch is integrated into MyLab Statistics, but it is also available by itself to qualified adopters. StatCrunch is also now available on your smartphone or tablet when you visit **www.statcrunch.com** from the device's browser. For more information, visit our website at **www.statcrunch.com,** or contact your Pearson representative.

MathXL Online Course (access code required)

Part of the world's leading collection of online homework, tutorial, and assessment products, MathXL® delivers assessment and tutorial resources that provide engaging and personalized experiences for each student. Each course is developed to accompany Pearson's best-selling content, authored by thought leaders across the math curriculum, and can be easily customized to fit any course format.

With MathXL, instructors can:

- Create, edit, and assign online homework and tests using algorithmically generated exercises correlated at the objective level to the textbook.
- Create and assign their own online exercises and import TestGen tests for added flexibility.
- Maintain records of all student work tracked in MathXL's online gradebook.

With MathXL, students can:

- Take chapter tests in MathXL and receive personalized study plans and/or personalized homework assignments based on their test results.
- Use the study plan and/or the homework to link directly to tutorial exercises for the objectives they need to study.
- Access supplemental animations and video clips directly from selected exercises.

MathXL is available to qualified adopters. For more information, visit our website at **www.pearson.com/mathxl,** or contact your Pearson representative.

Minitab and Minitab Express

Minitab and Minitab Express™ make learning statistics easy and provide students with a skill-set that is in demand in today's data driven workforce. 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 (access card only; not sold as stand alone)

JMP Student Edition

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

XLSTAT

XLSTAT™ is 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

Resources for Success

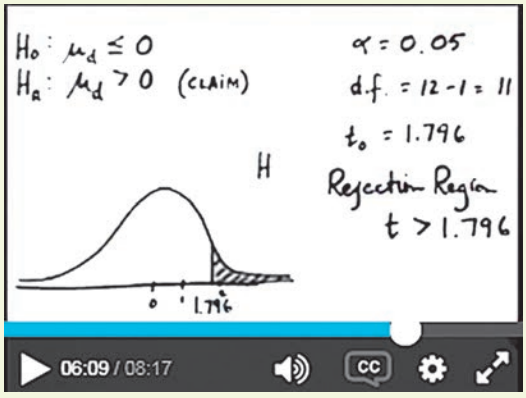
Pearson MyLab

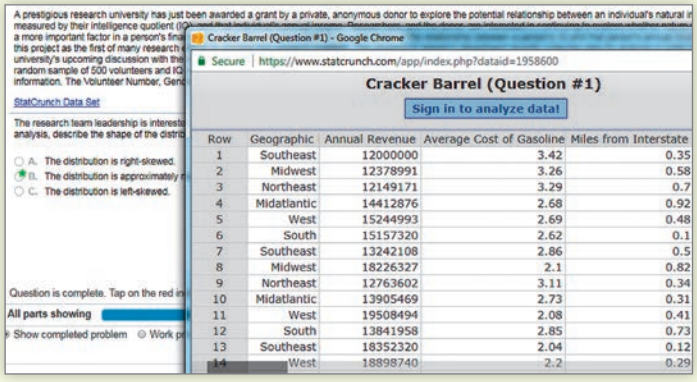
MyLab Statistics Online Course for *Elementary Statistics: Picturing the World*, 7e (access code required)

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. MyLab Statistics for *Elementary Statistics* includes the following new features, in addition to the resources listed on the previous page.

UPDATED! Video Program

Chapter Review Exercises come to life with new review videos that help students understand key chapter concepts. Section Lecture Videos work through examples and elaborate on key objectives.





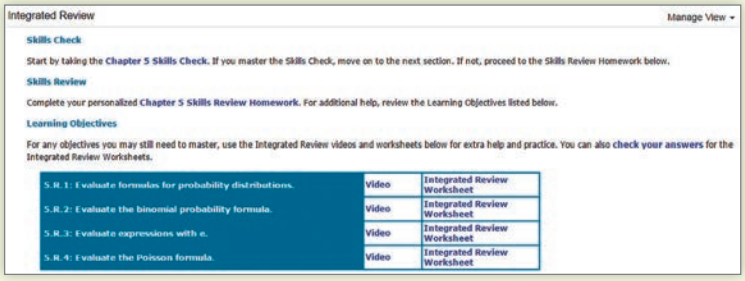
Row	Geographic	Annual Revenue	Average Cost of Gasoline	Miles from Interstate
1	Southeast	12000000	3.42	0.35
2	Midwest	12378991	3.26	0.58
3	Northeast	12149171	3.29	0.7
4	Midatlantic	14412876	2.68	0.92
5	West	15244993	2.69	0.48
6	South	15157320	2.62	0.1
7	Southeast	13242108	2.86	0.5
8	Midwest	18226327	2.1	0.82
9	Northeast	12763602	3.11	0.34
10	Midatlantic	13905469	2.73	0.31
11	West	19508494	2.08	0.41
12	South	13841958	2.85	0.73
13	Southeast	18352320	2.04	0.12
14	West	18898740	2.2	0.29

NEW! StatCrunch Question Library

This library of questions provides opportunities for students to analyze and interpret data sets in StatCrunch. Instructors can assign individual questions from the library by topic or they can assign questions from the same data set as a longer assignment that spans multiple learning objectives.

NEW! Integrated Review Course

Designed for just-in-time prerequisite review or for co-requisite courses, the Integrated Review version of the MyLab Statistics course provides pre-made, assignable skill-review quizzes and personalized homework assignments that are integrated throughout the regular statistics course content.



pearson.com/mylab/statistics

Resources for Success

Instructor Resources

Annotated Instructor's Edition

Includes suggested activities, additional ways to present material, common pitfalls, and other helpful teaching tips. All answers to the section and review exercises are provided in the margins next to the exercise. (ISBN-13: 978-0-13-468358-4; ISBN-10: 0-13-468358-7)

Instructor's Solutions Manual (downloadable)

Includes complete solutions to all of the exercises (including exercises in *Try it Yourself, Case Study, Technology, Uses and Abuses,* and *Real Statistics—Real Decisions* sections). It can be downloaded from within MyLab Statistics or from Pearson's online catalog, *www.pearson.com/us/higher-education.*

PowerPoint Lecture Slides (downloadable)

Classroom presentation slides feature key concepts, examples, and definitions from this text, along with notes with suggestions for presenting the material in class. They can be downloaded from within MyLab Statistics or from Pearson's online catalog, *www.pearson.com/us/higher-education.*

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's online catalog, *www.pearson.com/us/higher-education.* 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 pre-built exercises created specifically for statistics.

Student Resources

Video Resources

A comprehensive set of videos tied to the textbook contain short video clips with solutions to *Try It Yourself* exercises, Chapter Quiz Prep Videos, and Section Lecture Videos. Also, StatTalk Videos, hosted by fun-loving statistician Andrew Vickers, demonstrate important statistical concepts through interesting stories and real-life events. StatTalk Videos include assessment questions and an instructor's guide.

Student's Solutions Manual

(softcover and downloadable)

This manual includes complete worked-out solutions to all of the *Try It Yourself* exercises, the odd-numbered exercises, and all of the Chapter Quiz exercises. This manual is available in print and can be downloaded from MyLab Statistics. (ISBN-13: 978-0-13-468361-4; ISBN-10: 0-13-468361-7)

Technology Manuals for Elementary Statistics (downloadable)

Technology-specific manuals for Graphing Calculator, Excel®, and Minitab® include tutorial instruction and worked-out examples from the book. Each manual can be downloaded from within MyLab Statistics.

pearson.com/mylab/statistics

ACKNOWLEDGMENTS

I owe a debt of gratitude to the many reviewers who helped me shape and refine *Elementary Statistics: Picturing the World,* Seventh Edition.

Reviewers of the Current Edition

Karen Benway, University of Vermont
B.K. Brinkley, Tidewater Community College
Christine Curtis, Hillsborough Community College–Dale Mabry
Carrie Elledge, San Juan College
Jason Malozzi, Lower Columbia College
Cynthia McGinnis, Northwest Florida State College
Larry Musolino, Pennsylvania State University
Cyndi Roemer, Union County College
Jean Rowley, American Public University and DeVry University
Heidi Webb, Horry Georgetown Technical College

Reviewers of the Previous Editions

Rosalie Abraham, Florida Community College at Jacksonville
Ahmed Adala, Metropolitan Community College
Olcay Akman, College of Charleston
Polly Amstutz, University of Nebraska, Kearney
John J. Avioli, Christopher Newport University
David P. Benzel, Montgomery College
John Bernard, University of Texas—Pan American
G. Andy Chang, Youngstown State University
Keith J. Craswell, Western Washington University
Carol Curtis, Fresno City College
Dawn Dabney, Northeast State Community College
Cara DeLong, Fayetteville Technical Community College
Ginger Dewey, York Technical College
David DiMarco, Neumann College
Gary Egan, Monroe Community College
Charles Ehler, Anne Arundel Community College
Harold W. Ellingsen, Jr., SUNY—Potsdam
Michael Eurgubian, Santa Rosa Jr. College
Jill Fanter, Walters State Community College
Patricia Foard, South Plains College
Douglas Frank, Indiana University of Pennsylvania
Frieda Ganter, California State University
David Gilbert, Santa Barbara City College
Donna Gorton, Butler Community College
Larry Green, Lake Tahoe Community College
Sonja Hensler, St. Petersburg Jr. College
Sandeep Holay, Southeast Community College, Lincoln Campus
Lloyd Jaisingh, Morehead State
Nancy Johnson, Manatee Community College
Martin Jones, College of Charleston
David Kay, Moorpark College
Mohammad Kazemi, University of North Carolina—Charlotte
Jane Keller, Metropolitan Community College
Susan Kellicut, Seminole Community College
Hyune-Ju Kim, Syracuse University
Rita Kolb, Cantonsville Community College
Rowan Lindley, Westchester Community College
Jeffrey Linek, St. Petersburg Jr. College
Benny Lo, DeVry University, Fremont
Diane Long, College of DuPage
Austin Lovenstein, Pulaski Technical College
Rhonda Magel, North Dakota State University
Mike McGann, Ventura Community College
Vicki McMillian, Ocean County College
Lynn Meslinsky, Erie Community College
Lyn A. Noble, Florida Community College at Jacksonville—South Campus
Julie Norton, California State University—Hayward
Lynn Onken, San Juan College
Lindsay Packer, College of Charleston
Nishant Patel, Northwest Florida State
Jack Plaggemeyer, Little Big Horn College
Eric Preibisius, Cuyamaca Community College
Melonie Rasmussen, Pierce College
Neal Rogness, Grand Valley State University
Elisabeth Schuster, Benedictine University
Jean Sells, Sacred Heart University
John Seppala, Valdosta State University
Carole Shapero, Oakton Community College
Abdullah Shuaibi, Harry S. Truman College
Aileen Solomon, Trident Technical College
Sandra L. Spain, Thomas Nelson Community College
Michelle Strager-McCarney, Penn State—Erie, The Behrend College
Jennifer Strehler, Oakton Community College
Deborah Swiderski, Macomb Community College
William J. Thistleton, SUNY—Institute of Technology, Utica
Millicent Thomas, Northwest University
Agnes Tuska, California State University—Fresno
Clark Vangilder, DeVry University
Ting-Xiu Wang, Oakton Community
Dex Whittinghall, Rowan University
Cathleen Zucco-Teveloff, Rider University

Many thanks to Betsy Farber for her significant contributions to previous editions of the text. Sadly, Betsy passed away in 2013.

I would also like to thank the staff of Larson Texts, Inc., who assisted with the production of the book. On a personal level, I am grateful to my spouse, Deanna Gilbert Larson, for her love, patience, and support. Also, a special thanks goes to R. Scott O'Neil.

I have worked hard to make this text a clean, clear, and enjoyable one from which to teach and learn statistics. Despite my best efforts to ensure accuracy and ease of use, many users will undoubtedly have suggestions for improvement. I welcome your suggestions.

Ron Larson, odx@psu.edu

INDEX OF APPLICATIONS

Earth Science

Economics and Finance

Education

Engineering

Entertainment

Food and Nutrition

Government and Political Science

Health and Medicine

Housing and Construction

Law

Miscellaneous

Mortality

Motor Vehicles and Transportation

Psychology

Sports

Work

CHAPTER 1

Introduction to Statistics

For the first 10 months of 2016, construction completions of privately-owned housing units in the U.S. was greatest in the south.

Where You've Been

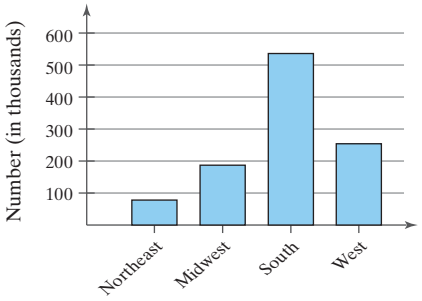
You are already familiar with many of the practices of statistics, such as taking surveys, collecting data, and describing populations. What you may not know is that collecting accurate statistical data is often difficult and costly. Consider, for instance, the monumental task of counting and describing the entire population of the United States. If you were in charge of such a census, how would you do it? How would you ensure that your results are accurate? These and many more concerns are the responsibility of the United States Census Bureau, which conducts the census every decade.

Where You're Going

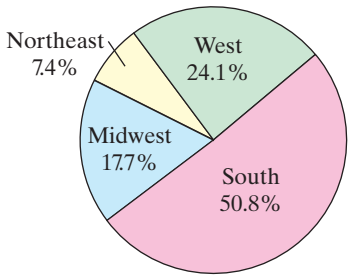
In Chapter 1, you will be introduced to the basic concepts and goals of statistics. For instance, statistics were used to construct the figures below, which show the numbers, by region in the U.S., of construction completions of privately-owned housing units for October of 2016 and for the first 10 months of 2016, as numbers in thousands and as percents of the total.

For the 2010 Census, the Census Bureau sent short forms to every household. Short forms ask all members of every household such things as their gender, age, race, and ethnicity. Previously, a long form, which covered additional topics, was sent to about 17% of the population. But for the first time since 1940, the long form was replaced by the American Community Survey, which surveys more than 3.5 million households a year throughout the decade. These households form a sample. In this course, you will learn how the data collected from a sample are used to infer characteristics about the entire population.

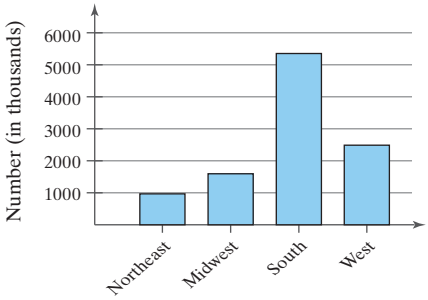
Housing Units Completed in the U.S. (October 2016)



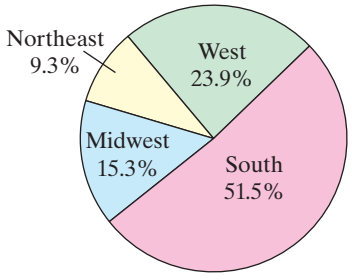
Housing Units Completed in the U.S. (October 2016)



Housing Units Completed in the U.S. (January–October 2016)



Housing Units Completed in the U.S. (January–October 2016)



1.1 An Overview of Statistics

What You Should Learn

- A definition of statistics
- How to distinguish between a population and a sample and between a parameter and a statistic
- How to distinguish between descriptive statistics and inferential statistics

A Definition of Statistics ■ Data Sets ■ Branches of Statistics

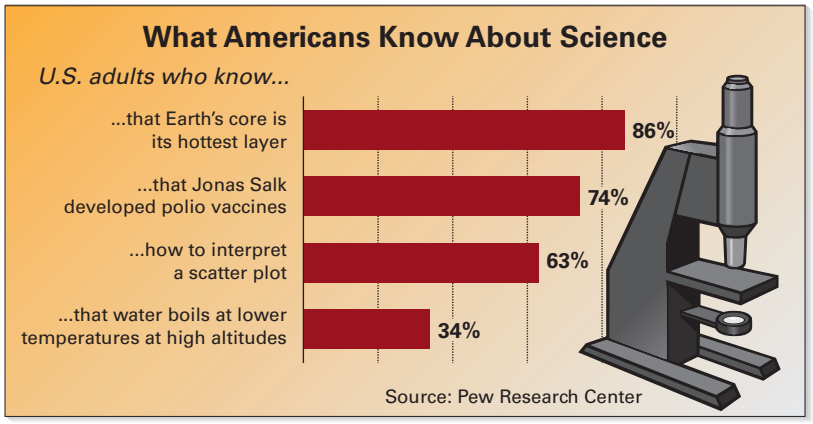
A Definition of Statistics

Almost every day you are exposed to statistics. For instance, consider the next two statements.

- According to a survey, more than 7 in 10 Americans say a nursing career is a prestigious occupation. *(Source: The Harris Poll)*
- "Social media consumes kids today as well, as more score their first social media accounts at an average age of 11.4 years old." *(Source: Influence Central's 2016 Digital Trends Study)*

By learning the concepts in this text, you will gain the tools to become an informed consumer, understand statistical studies, conduct statistical research, and sharpen your critical thinking skills.

Many statistics are presented graphically. For instance, consider the figure shown below.



The information in the figure is based on the collection of **data.** In this instance, the data are based on the results of a science quiz given to 3278 U.S. adults.

DEFINITION

Data consist of information coming from observations, counts, measurements, or responses.

The use of statistics dates back to census taking in ancient Babylonia, Egypt, and later in the Roman Empire, when data were collected about matters concerning the state, such as births and deaths. In fact, the word *statistics* is derived from the Latin word *status*, meaning "state." The modern practice of statistics involves more than counting births and deaths, as you can see in the next definition.

DEFINITION

Statistics is the science of collecting, organizing, analyzing, and interpreting data in order to make decisions.

Data Sets

There are two types of data sets you will use when studying statistics. These data sets are called **populations** and **samples.**

Study Tip

A *census* consists of data from an entire population. But, unless a population is small, it is usually impractical to obtain all the population data. In most studies, information must be obtained from a random sample.

DEFINITION

A **population** is the collection of *all* outcomes, responses, measurements, or counts that are of interest. A **sample** is a subset, or part, of a population.

A sample is used to gain information about a population. For instance, to estimate the unemployment rate for the *population* of the United States, the U.S. Bureau of Labor uses a *sample* of about 60,000 households.

A sample should be representative of a population so that sample data can be used to draw conclusions about that population. Sample data must be collected using an appropriate method, such as *random sampling.* When sample data are collected using an *inappropriate* method, the data cannot be used to draw conclusions about the population. (You will learn more about random sampling and data collection in Section 1.3.)

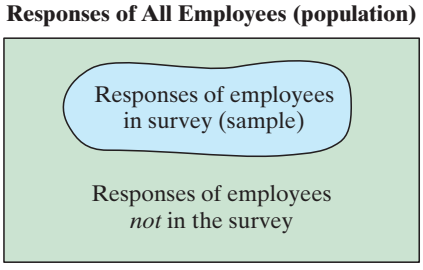
EXAMPLE 1

Identifying Data Sets

In a survey, 834 employees in the United States were asked whether they thought their jobs were highly stressful. Of the 834 respondents, 517 said yes. Identify the population and the sample. Describe the sample data set. *(Source: CareerCast Job Stress Report)*

SOLUTION

The population consists of the responses of all employees in the United States. The sample consists of the responses of the 834 employees in the survey. In the Venn diagram below, notice that the sample is a subset of the responses of all employees in the United States. Also, the sample data set consists of 517 people who said yes and 317 who said no.



TRY IT YOURSELF 1

In a survey of 1501 ninth to twelfth graders in the United States, 1215 said "leaders today are more concerned with their own agenda than with achieving the overall goals of the organization they serve." Identify the population and the sample. Describe the sample data set. *(Source: National 4-H Council)*

Answer: Page A31

Whether a data set is a population or a sample usually depends on the context of the real-life situation. For instance, in Example 1, the population is the set of responses of all employees in the United States. Depending on the purpose of the survey, the population could have been the set of responses of all employees who live in California or who work in the healthcare industry.

Study Tip

To remember the terms parameter and statistic, try using the mnemonic device of matching the first letters in *population parameter* and the first letters in *sample statistic.*

Two important terms that are used throughout this course are **parameter** and **statistic.**

DEFINITION

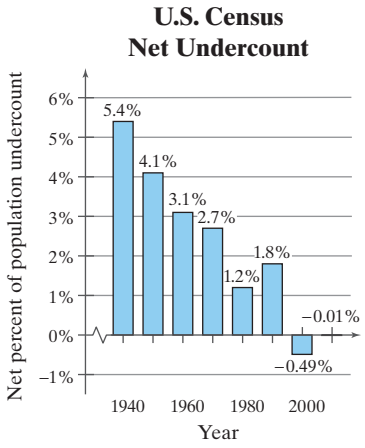
A **parameter** is a numerical description of a *population* characteristic.

A **statistic** is a numerical description of a *sample* characteristic.

It is important to note that a sample statistic can differ from sample to sample, whereas a population parameter is constant for a population. For instance, consider the survey in Example 1. The results showed that 517 of 834 employees surveyed think their jobs are highly stressful. Another sample may have a different number of employees that say their jobs are highly stressful. For the population, however, the number of employees who think that their jobs are highly stressful does not change.

Picturing the World

How accurate is the count of the U.S. population taken each decade by the Census Bureau? According to estimates, the net undercount of the U.S. population by the 1940 census was 5.4%. The accuracy of the census has improved greatly since then. The net undercount in the 2010 census was –0.01%. (This means that the 2010 census overcounted the U.S. population by 0.01%, which is about 36,000 people.) (Source: U.S. Census Bureau)



What are some difficulties in collecting population data?

EXAMPLE 2

Distinguishing Between a Parameter and a Statistic

Determine whether each number describes a population parameter or a sample statistic. Explain your reasoning.

1. A survey of several hundred collegiate student-athletes in the United States found that, during the season of their sport, the average time spent on athletics by student-athletes is 50 hours per week. *(Source: Penn Schoen Berland)*
2. The freshman class at a university has an average SAT math score of 514.
3. In a random check of several hundred retail stores, the Food and Drug Administration found that 34% of the stores were not storing fish at the proper temperature.

SOLUTION

1. Because the average of 50 hours per week is based on a subset of the population, it is a sample statistic.
2. Because the average SAT math score of 514 is based on the entire freshman class, it is a population parameter.
3. Because 34% is based on a subset of the population, it is a sample statistic.

TRY IT YOURSELF 2

Determine whether each number describes a population parameter or a sample statistic. Explain your reasoning.

a. Last year, a small company spent a total of $5,150,694 on employees' salaries.
b. In the United States, a survey of a few thousand adults with hearing loss found that 43% have difficulty remembering conversations. *(Source: The Harris Poll)*

Answer: Page A31

In this course, you will see how the use of statistics can help you make informed decisions that affect your life. Consider the census that the U.S. government takes every decade. When taking the census, the Census Bureau attempts to contact everyone living in the United States. Although it is impossible to count everyone, it is important that the census be as accurate as it can be because public officials make many decisions based on the census information. Data collected in the census will determine how to assign congressional seats and how to distribute public funds.

Branches of Statistics

The study of statistics has two major branches: **descriptive statistics** and **inferential statistics.**

DEFINITION

Descriptive statistics is the branch of statistics that involves the organization, summarization, and display of data.

Inferential statistics is the branch of statistics that involves using a sample to draw conclusions about a population. A basic tool in the study of inferential statistics is probability. (You will learn more about probability in Chapter 3.)

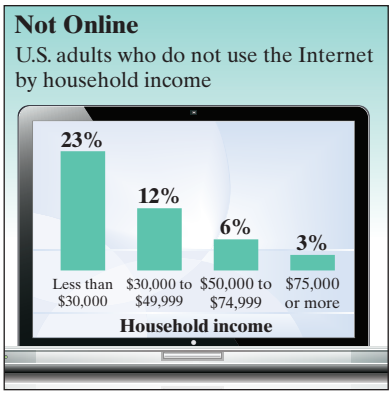
EXAMPLE 3

Descriptive and Inferential Statistics

For each study, identify the population and the sample. Then determine which part of the study represents the descriptive branch of statistics. What conclusions might be drawn from the study using inferential statistics?

1. A study of 2560 U.S. adults found that of adults not using the Internet, 23% are from households earning less than $30,000 annually, as shown in the figure at the left. *(Source: Pew Research Center)*

2. A study of 300 Wall Street analysts found that the percentage who incorrectly forecasted high-tech earnings in a recent year was 44%. *(Adapted from Bloomberg News)*



SOLUTION

1. The population consists of the responses of all U.S. adults, and the sample consists of the responses of the 2560 U.S. adults in the study. The part of this study that represents the descriptive branch of statistics involves the statement "23% [of U.S. adults not using the Internet] are from households earning less than $30,000 annually." Also, the figure represents the descriptive branch of statistics. A possible inference drawn from the study is that lower-income households cannot afford access to the Internet.

2. The population consists of the high-tech earnings forecasts of all Wall Street analysts, and the sample consists of the forecasts of the 300 Wall Street analysts in the study. The part of this study that represents the descriptive branch of statistics involves the statement "the percentage [of Wall Street analysts] who incorrectly forecasted high-tech earnings in a recent year was 44%." A possible inference drawn from the study is that the stock market is difficult to forecast, even for professionals.

TRY IT YOURSELF 3

A study of 1000 U.S. adults found that when they have a question about their medication, three out of four adults will consult with their physician or pharmacist and only 8% visit a medication-specific website. *(Source: Finn Futures™ Health poll)*

a. Identify the population and the sample.
b. Determine which part of the study represents the descriptive branch of statistics.
c. What conclusions might be drawn from the study using inferential statistics?

Answer: Page A31

Study Tip

Throughout this course you will see applications of both branches of statistics. A major theme in this course will be how to use sample statistics to make inferences about unknown population parameters.

1.1 EXERCISES

For Extra Help: MyLab Statistics

Building Basic Skills and Vocabulary

1. How is a sample related to a population?
2. Why is a sample used more often than a population?
3. What is the difference between a parameter and a statistic?
4. What are the two main branches of statistics?

True or False? *In Exercises 5–10, determine whether the statement is true or false. If it is false, rewrite it as a true statement.*

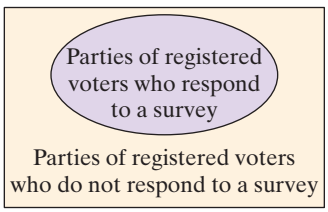
5. A statistic is a numerical description of a population characteristic.
6. A sample is a subset of a population.
7. It is impossible to obtain all the census data about the U.S. population.
8. Inferential statistics involves using a population to draw a conclusion about a corresponding sample.
9. A population is the collection of some outcomes, responses, measurements, or counts that are of interest.
10. A sample statistic will not change from sample to sample.

Classifying a Data Set *In Exercises 11–20, determine whether the data set is a population or a sample. Explain your reasoning.*

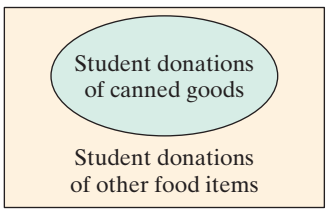
11. The salary of each member of a Major League Baseball team
12. The amount of energy collected from every solar panel on a photovoltaic power plant
13. A survey of 300 people from an auditorium with 13,000 people
14. The annual revenue of each store in a shopping mall
15. The triglyceride levels of 10 patients in a clinic with 50 patients
16. The number of wireless devices in each U.S. household
17. The final score of each gamer in a tournament
18. The age of every fourth person entering a grocery store
19. The political party of every U.S. senator
20. The air contamination levels at 20 locations near a factory

Graphical Analysis *In Exercises 21–24, use the Venn diagram to identify the population and the sample.*

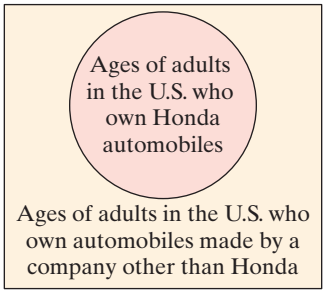
21. **Parties of Registered Voters**



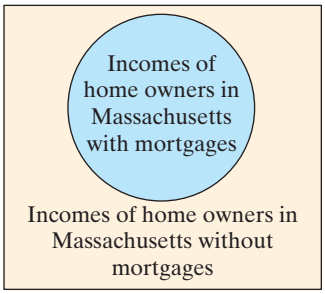
22. **Student Donations at a Food Drive**



23. **Ages of Adults in the United States Who Own Automobiles**



24. **Incomes of Home Owners in Massachusetts**



Using and Interpreting Concepts

Identifying Data Sets *In Exercises 25–34, identify the population and the sample. Describe the sample data set.*

25. A survey of 1020 U.S. adults found that 42% trust their political leaders. *(Source: Gallup)*

26. A study of 203 infants was conducted to find a link between fetal tobacco exposure and focused attention in infancy. *(Source: Infant Behavior and Development)*

27. A survey of 3301 U.S. adults found that 39% received an influenza vaccine for a recent flu season. *(Source: U.S. Centers for Disease Control and Prevention)*

28. A survey of 1100 travelers worldwide found that 53% of respondents with pets travel with their pets.

29. A survey of 159 U.S. law firms found that the average hourly billing rate for partners was $604. *(Source: The National Law Journal)*

30. A survey of 496 students at a high school found that 95% planned on going to college.

31. A survey of 1029 U.S. adults found that 23% of those suffering with chronic pain had been diagnosed with a sleep disorder. *(Source: National Sleep Foundation)*

32. A survey of 1254 preowned automobile shoppers found that 5% bought extended warranties.

33. To gather information about starting salaries at companies listed in the Standard & Poor's 500, a researcher contacts 54 of the 500 companies.

34. A survey of 1060 parents of 13- to 17-year-olds found that 636 of the 1060 parents have checked their teen's social media profile. *(Source: Pew Research Center)*

Distinguishing Between a Parameter and a Statistic *In Exercises 35–42, determine whether the number describes a population parameter or a sample statistic. Explain your reasoning.*

35. The average salary for 45 of a consulting firm's 300 engineers is $72,000.

36. A survey of 1058 college board members found that 56.3% think that college completion is a major priority or the most important priority for their board. *(Source: Association of Governing Boards of Universities and Colleges)*

37. Sixty-two of the 97 passengers aboard the Hindenburg airship survived its explosion.

38. In January 2016, 62% of the governors of the 50 states in the United States were Republicans. *(Source: National Governors Association)*

39. In a survey of 400 computer users, 7% said their computers had malfunctions that needed to be repaired by service technicians.

40. Voter registration records show that 87% of all voters in a county are registered as Democrats.

41. A survey of 2008 U.S. adults found that 80% think that the militant group known as ISIS is a major threat to the well-being of the United States. *(Source: Pew Research Center)*

42. In a recent year, the average math score on the ACT for all graduates was 20.6. *(Source: ACT, Inc.)*

43. Descriptive and Inferential Statistics Which part of the survey described in Exercise 31 represents the descriptive branch of statistics? What conclusions might be drawn from the survey using inferential statistics?

44. Descriptive and Inferential Statistics Which part of the survey described in Exercise 32 represents the descriptive branch of statistics? What conclusions might be drawn from the survey using inferential statistics?

Extending Concepts

45. Identifying Data Sets in Articles Find an article that describes a survey.

(a) Identify the sample used in the survey.

(b) What is the population?

(c) Make an inference about the population based on the results of the survey.

46. Writing Write an essay about the importance of statistics for one of the following.

- A study on the effectiveness of a new drug
- An analysis of a manufacturing process
- Drawing conclusions about voter opinions using surveys

47. Exercise and Cognitive Ability A study of 876 senior citizens shows that participants who exercise regularly exhibit less of a decline in cognitive ability than those who barely exercise at all. From this study, a researcher infers that your cognitive ability increases the more your exercise. What is wrong with this type of reasoning? *(Source: Neurology)*

48. Increase in Obesity Rates A study shows that the obesity rate among adolescents has steadily increased since 1988. From this study, a researcher infers that this trend will continue in future years. What is wrong with this type of reasoning? *(Source: Journal of the American Medical Association)*

49. Sleep and Student Achievement A study shows the closer that participants were to an optimal sleep duration target, the better they performed on a standardized test. *(Source: Eastern Economics Journal)*

(a) Identify the sample used in the study.

(b) What is the population?

(c) Which part of the study represents the descriptive branch of statistics?

(d) Make an inference about the population based on the results of the study.

1.2 Data Classification

What You Should Learn

- How to distinguish between qualitative data and quantitative data
- How to classify data with respect to the four levels of measurement: nominal, ordinal, interval, and ratio

Types of Data ■ Levels of Measurement

Types of Data

When conducting a study, it is important to know the kind of data involved. The type of data you are working with will determine which statistical procedures can be used. In this section, you will learn how to classify data by type and by level of measurement. Data sets can consist of two types of data: **qualitative data** and **quantitative data.**

DEFINITION

Qualitative data consist of attributes, labels, or nonnumerical entries.

Quantitative data consist of numbers that are measurements or counts.

EXAMPLE 1

Classifying Data by Type

The table shows sports-related head injuries treated in U.S. emergency rooms during a recent five-year span for several sports. Which data are qualitative data and which are quantitative data? Explain your reasoning. *(Source: BMC Emergency Medicine)*

Sports-Related Head Injuries Treated in U.S. Emergency Rooms

Sport	Head injuries treated
Basketball	131,930
Baseball	83,522
Football	220,258
Gymnastics	33,265
Hockey	41,450
Soccer	98,710
Softball	41,216
Swimming	44,815
Volleyball	13,848

SOLUTION

The information shown in the table can be separated into two data sets. One data set contains the names of sports, and the other contains the numbers of head injuries treated. The names are nonnumerical entries, so these are qualitative data. The numbers of head injuries treated are numerical entries, so these are quantitative data.

TRY IT YOURSELF 1

The populations of several U.S. cities are shown in the table. Which data are qualitative data and which are quantitative data? Explain your reasoning. *(Source: U.S. Census Bureau)*

City	Population
Baltimore, MD	621,849
Chicago, IL	2,720,546
Glendale, AZ	240,126
Miami, FL	441,003
Portland, OR	632,309
San Francisco, CA	864,816

Answer: Page A31

Levels of Measurement

Another characteristic of data is its level of measurement. The level of measurement determines which statistical calculations are meaningful. The four levels of measurement, in order from lowest to highest, are **nominal, ordinal, interval,** and **ratio.**

DEFINITION

Data at the **nominal level of measurement** are qualitative only. Data at this level are categorized using names, labels, or qualities. No mathematical computations can be made at this level.

Data at the **ordinal level of measurement** are qualitative or quantitative. Data at this level can be arranged in order, or ranked, but differences between data entries are not meaningful.

When numbers are at the nominal level of measurement, they simply represent a label. Examples of numbers used as labels include Social Security numbers and numbers on sports jerseys. For instance, it would not make sense to add the numbers on the players' jerseys for the Chicago Bears.

Picturing the World

For more than 25 years, the Harris Poll has conducted an annual study to determine the strongest brands, based on consumer response, in several industries. A recent study determined the top five health nonprofit brands, as shown in the table. (Source: Harris Poll)

Top five health nonprofit brands
1. St Jude Children's Research Hospital
2. Shriners Hospital for Children
3. Make-A-Wish
4. The Jimmy Fund
5. American Cancer Society

In this list, what is the level of measurement?

EXAMPLE 2

Classifying Data by Level

For each data set, determine whether the data are at the nominal level or at the ordinal level. Explain your reasoning. *(Source: U.S. Bureau of Labor Statistics)*

1.

Top five U.S. occupations with the most job growth (projected 2024)
1. Personal care aides
2. Registered nurses
3. Home health aides
4. Combined food preparation and serving workers, including fast food
5. Retail salespersons

2.

Movie genres
Action
Adventure
Comedy
Drama
Horror

SOLUTION

1. This data set lists the ranks of the five fastest-growing occupations in the U.S. over the next few years. The data set consists of the ranks 1, 2, 3, 4, and 5. Because the ranks can be listed in order, these data are at the ordinal level. Note that the difference between a rank of 1 and 5 has no mathematical meaning.

2. This data set consists of the names of movie genres. No mathematical computations can be made with the names, and the names cannot be ranked, so these data are at the nominal level.

TRY IT YOURSELF 2

For each data set, determine whether the data are at the nominal level or at the ordinal level. Explain your reasoning.

1. The final standings for the Pacific Division of the National Basketball Association

2. A collection of phone numbers

Answer: Page A31

The two highest levels of measurement consist of quantitative data only.

DEFINITION

Data at the **interval level of measurement** can be ordered, and meaningful differences between data entries can be calculated. At the interval level, a zero entry simply represents a position on a scale; the entry is not an inherent zero.

Data at the **ratio level of measurement** are similar to data at the interval level, with the added property that a zero entry is an inherent zero. A ratio of two data entries can be formed so that one data entry can be meaningfully expressed as a multiple of another.

An *inherent zero* is a zero that implies "none." For instance, the amount of money you have in a savings account could be zero dollars. In this case, the zero represents no money; it is an inherent zero. On the other hand, a temperature of 0°C does not represent a condition in which no heat is present. The 0°C temperature is simply a position on the Celsius scale; it is not an inherent zero.

To distinguish between data at the interval level and at the ratio level, determine whether the expression "twice as much" has any meaning in the context of the data. For instance, \$2 is twice as much as \$1, so these data are at the ratio level. On the other hand, 2°C is not twice as warm as 1°C, so these data are at the interval level.

EXAMPLE 3

Classifying Data by Level

Two data sets are shown at the left. Which data set consists of data at the interval level? Which data set consists of data at the ratio level? Explain your reasoning. *(Source: Major League Baseball)*

New York Yankees' World Series victories (years)
1923, 1927, 1928, 1932, 1936, 1937, 1938, 1939, 1941, 1943, 1947, 1949, 1950, 1951, 1952, 1953, 1956, 1958, 1961, 1962, 1977, 1978, 1996, 1998, 1999, 2000, 2009

2016 American League home run totals (by team)	
Baltimore	253
Boston	208
Chicago	168
Cleveland	185
Detroit	211
Houston	198
Kansas City	147
Los Angeles	156
Minnesota	200
New York	183
Oakland	169
Seattle	223
Tampa Bay	216
Texas	215
Toronto	221

SOLUTION

Both of these data sets contain quantitative data. Consider the dates of the Yankees' World Series victories. It makes sense to find differences between specific dates. For instance, the time between the Yankees' first and last World Series victories is

$$2009 - 1923 = 86 \text{ years.}$$

But it does not make sense to say that one year is a multiple of another. So, these data are at the interval level. However, using the home run totals, you can find differences *and* write ratios. For instance, Boston hit 23 more home runs than Cleveland hit because $208 - 185 = 23$ home runs. Also, Baltimore hit about 1.5 times as many home runs as Chicago hit because

$$\frac{253}{168} \approx 1.5.$$

So, these data are at the ratio level.

TRY IT YOURSELF 3

For each data set, determine whether the data are at the interval level or at the ratio level. Explain your reasoning.

1. The body temperatures (in degrees Fahrenheit) of an athlete during an exercise session
2. The heart rates (in beats per minute) of an athlete during an exercise session

Answer: Page A31

The tables below summarize which operations are meaningful at each of the four levels of measurement. When identifying a data set's level of measurement, use the highest level that applies.

Level of measurement	Put data in categories	Arrange data in order	Subtract data entries	Determine whether one data entry is a multiple of another
Nominal	Yes	No	No	No
Ordinal	Yes	Yes	No	No
Interval	Yes	Yes	Yes	No
Ratio	Yes	Yes	Yes	Yes

Summary of Four Levels of Measurement

	Example of a data set	Meaningful calculations
Nominal level (Qualitative data)	*Types of Shows Televised by a Network* Comedy — Documentaries Drama — Cooking Reality Shows — Soap Operas Sports — Talk Shows	*Put in a category.* For instance, a show televised by the network could be put into one of the eight categories shown.
Ordinal level (Qualitative or quantitative data)	*Motion Picture Association of America Ratings Description* G — General Audiences PG — Parental Guidance Suggested PG-13 — Parents Strongly Cautioned R — Restricted NC-17 — No One 17 and Under Admitted	Put in a category and *put in order.* For instance, a PG rating has a stronger restriction than a G rating.
Interval level (Quantitative data)	*Average Monthly Temperatures (in degrees Fahrenheit) for Denver, CO* Jan 30.7 — Jul 74.2 Feb 32.5 — Aug 72.5 Mar 40.4 — Sep 63.4 Apr 47.4 — Oct 50.9 May 57.1 — Nov 38.3 Jun 67.4 — Dec 30.0 *(Source: National Climatic Data Center)*	Put in a category, put in order, and *find differences between data entries.* For instance, $72.5 - 63.4 = 9.1$°F. So, August is 9.1°F warmer than September.
Ratio level (Quantitative data)	*Average Monthly Precipitation (in inches) for Orlando, FL* Jan 2.35 — Jul 7.27 Feb 2.38 — Aug 7.13 Mar 3.77 — Sep 6.06 Apr 2.68 — Oct 3.31 May 3.45 — Nov 2.17 Jun 7.58 — Dec 2.58 *(Source: National Climatic Data Center)*	Put in a category, put in order, find differences between data entries, and *find ratios of data entries.* For instance, $\frac{7.58}{3.77} \approx 2.$ So, there is about twice as much precipitation in June as in March.

1.2 EXERCISES

For Extra Help: MyLab Statistics

Building Basic Skills and Vocabulary

1. Name each level of measurement for which data can be qualitative.
2. Name each level of measurement for which data can be quantitative.

True or False? *In Exercises 3–6, determine whether the statement is true or false. If it is false, rewrite it as a true statement.*

3. Data at the ordinal level are quantitative only.
4. For data at the interval level, you cannot calculate meaningful differences between data entries.
5. More types of calculations can be performed with data at the nominal level than with data at the interval level.
6. Data at the ratio level cannot be put in order.

Using and Interpreting Concepts

Classifying Data by Type *In Exercises 7–14, determine whether the data are qualitative or quantitative. Explain your reasoning.*

7. Weights of dogs at an animal rescue facility
8. Carrying capacities of flatbed trucks
9. Hair colors of classmates
10. Student ID numbers
11. Heights of infants in a maternity ward
12. Species of mammals in a rain forest
13. Responses on an opinion poll
14. Wait times at a the Department of Motor Vehicles

Classifying Data By Level *In Exercises 15–20, determine the level of measurement of the data set. Explain your reasoning.*

15. **Comedy Series** The years that a television show on ABC won the Emmy for best comedy series are listed. *(Source: Academy of Television Arts and Sciences)*

1955	1979	1980	1981	1982	1988
2010	2011	2012	2013	2014	

16. **Business Schools** The top ten business schools in the United States for a recent year according to Forbes are listed. *(Source: Forbes Media LLC)*

1. Stanford
2. Harvard
3. Northwestern (Kellogg)
4. Columbia
5. Dartmouth (Tuck)
6. Chicago (Booth)
7. Pennsylvania (Wharton)
8. UC Berkeley (Haas)
9. MIT (Sloan)
10. Cornell (Johnson)

17. Flight Departures The flight numbers of 21 departing flights from Chicago O'Hare International Airport on an afternoon in October of 2016 are listed. *(Source: Chicago O'Hare International Airport)*

1785	5159	4509	1575	6827	3486	7676
1989	522	6868	1893	3133	3337	3266
3458	334	6320	8385	3112	2110	7664

18. Songs The lengths (in seconds) of songs on an album are listed.

228	233	268	265	252
335	103	338	252	371
586	290	532	282	

19. Best Sellers List The top ten fiction books on The New York Times Best Sellers List on October 9, 2016, are listed. *(Source: The New York Times)*

1. The Girl on the Train
2. Home
3. The Kept Woman
4. Magic Binds
5. Commonwealth
6. The Light Between Oceans
7. Immortal Nights
8. A Man Called Ove
9. Thrice the Brinded Cat Hath Mew'd
10. The Woman in Cabin 10

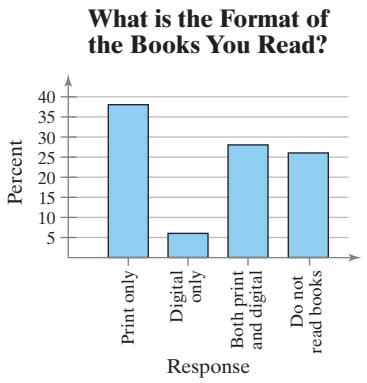
20. Cell Phone The times of the day when a person checks his or her cell phone are listed.

8:28 A.M.	9:30 A.M.	9:43 A.M.	10:18 A.M.
11:25 A.M.	11:46 A.M.	12:27 P.M.	2:18 P.M.
2:26 P.M.	2:49 P.M.	3:05 P.M.	4:18 P.M.
5:28 P.M.	5:57 P.M.	8:17 P.M.	

Graphical Analysis

In Exercises 21–24, determine the level of measurement of the data listed on the horizontal and vertical axes in the figure.

21.

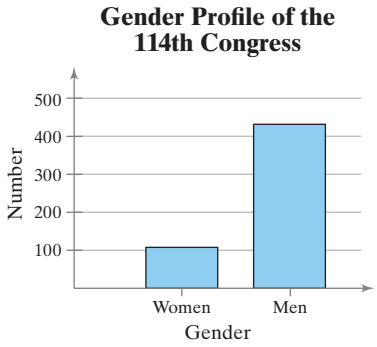


(Source: Pew Research Center)

22.

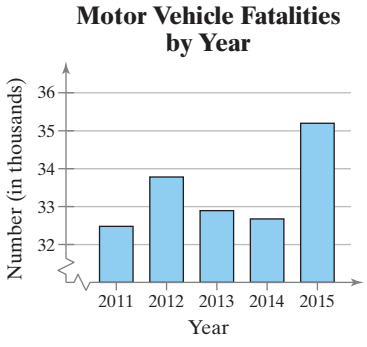
(Source: The Harris Poll)

23.



(Source: Congressional Research Service)

24.



(Source: National Highway Traffic Safety Administration)

25. The items below appear on a physician's intake form. Determine the level of measurement of the data for each category.

(a) Temperature (b) Allergies

(c) Weight (d) Pain level (scale of 0 to 10)

26. The items below appear on an employment application. Determine the level of measurement of the data for each category.

(a) Highest grade level completed (b) Gender

(c) Year of college graduation (d) Number of years at last job

Classifying Data by Type and Level *In Exercises 27–32, determine whether the data are qualitative or quantitative, and determine the level of measurement of the data set.*

27. Football The top ten teams in the final college football poll released in January 2017 are listed. *(Source: Associated Press)*

1. Clemson
2. Alabama
3. USC
4. Washington
5. Oklahoma
6. Ohio State
7. Penn State
8. Florida State
9. Wisconsin
10. Michigan

28. Politics The three political parties in the 114th Congress are listed.

Republican Democrat Independent

29. Top Salespeople The regions representing the top salespeople in a corporation for the past six years are listed.

Southeast	Northwest
Northeast	Southeast
Southwest	Southwest

30. Diving The scores for the gold medal winning diver in the men's 10-meter platform event from the 2016 Summer Olympics are listed. *(Source: International Olympic Committee)*

91.80	91.00	88.20
97.20	99.90	91.80

31. Concert Tours The top ten highest grossing worldwide concert tours for 2016 are listed. *(Source: Pollstar)*

1. Bruce Springsteen & the E Street Band
2. Beyoncé
3. Coldplay
4. Guns N' Roses
5. Adele
6. Justin Bieber
7. Paul McCartney
8. Garth Brooks
9. The Rolling Stones
10. Celine Dion

32. Numbers of Performances The numbers of performances for the 10 longest-running Broadway shows at the end of the 2016 season are listed. *(Source: The Broadway League)*

11,782	8107	7705	7485	6680
6137	5959	5758	5461	5238

Extending Concepts

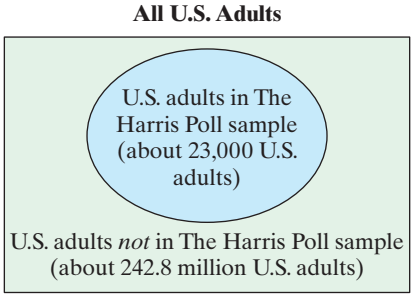
33. Writing What is an inherent zero? Describe three examples of data sets that have inherent zeros and three that do not.

34. Describe two examples of data sets for each of the four levels of measurement. Justify your answer.

CASE STUDY Reputations of Companies in the U.S.

For more than 50 years, The Harris Poll has conducted surveys using a representative sample of people in the United States. The surveys have been used to represent the opinions of people in the United States on many subjects, such as health, politics, the U.S. economy, and sports.

Since 1999, The Harris Poll has conducted an annual survey to measure the reputations of the most visible companies in the United States, as perceived by U.S. adults. The Harris Poll uses a sample of about 23,000 U.S. adults for the survey. The survey respondents rate companies according to 20 attributes that are classified into six categories: (1) social responsibility, (2) vision and leadership, (3) financial performance, (4) products and services, (5) emotional appeal, and (6) workplace environment. This information is used to determine the reputation of a company as Excellent, Very Good, Good, Fair, Poor, Very Poor, or Critical. The reputations (along with some additional information) of 10 companies are shown in the table.



Reputations of 10 Companies in the U.S.

Company Name	Year Company Formed	Reputation	Industry	Number of Employees
Amazon.com	1994	Excellent	Retail	230,800
Apple, Inc.	1977	Excellent	Computers and peripherals	116,000
Netflix, Inc.	1999	Very Good	Internet television	4,700
The Kraft Heinz Co.	2015	Very Good	Food products	41,000
Facebook, Inc.	2004	Good	Internet	17,048
Ford Motor Co.	1903	Good	Automotive	201,000
Chipotle Mexican Grill, Inc.	1993	Fair	Restaurant	64,570
Comcast Corp.	1963	Poor	Cable television	136,000
Exxon Mobil Corp.	1999	Poor	Petroleum (integrated)	71,100
Wells Fargo & Co.	1998	Critical	Banking	265,000

(Source: The Harris Poll; Amazon.com; Apple, Inc.; Netflix, Inc.; The Kraft Heinz Co.; Facebook, Inc.; Ford Motor Co.; Chipotle Mexican Grill, Inc.; Comcast Corp.; Exxon Mobil Corp.; Wells Fargo & Co.)

EXERCISES

1. **Sampling Percent** What percentage of the total number of U.S. adults did The Harris Poll sample for its survey? (Assume the total number of U.S. adults is 242.8 million.)
2. **Nominal Level of Measurement** Identify any column in the table with data at the nominal level.
3. **Ordinal Level of Measurement** Identify any column in the table with data at the ordinal level. Describe two ways that the data can be ordered.
4. **Interval Level of Measurement** Identify any column in the table with data at the interval level. How can these data be ordered?
5. **Ratio Level of Measurement** Identify any column in the table with data at the ratio level.
6. **Inferences** What decisions can be made on the basis of The Harris Poll survey that measures the reputations of the most visible companies in the United States?

1.3 Data Collection and Experimental Design

What You Should Learn

- How to design a statistical study and how to distinguish between an observational study and an experiment
- How to collect data by using a survey or a simulation
- How to design an experiment
- How to create a sample using random sampling, simple random sampling, stratified sampling, cluster sampling, and systematic sampling and how to identify a biased sample

Design of a Statistical Study ■ Data Collection ■ Experimental Design ■ Sampling Techniques

Design of a Statistical Study

The goal of every statistical study is to collect data and then use the data to make a decision. Any decision you make using the results of a statistical study is only as good as the process used to obtain the data. When the process is flawed, the resulting decision is questionable.

Although you may never have to develop a statistical study, it is likely that you will have to interpret the results of one. Before interpreting the results of a study, however, you should determine whether the results are reliable. In other words, you should be familiar with how to design a statistical study.

GUIDELINES

Designing a Statistical Study

1. Identify the variable(s) of interest (the focus) and the population of the study.
2. Develop a detailed plan for collecting data. If you use a sample, make sure the sample is representative of the population.
3. Collect the data.
4. Describe the data, using descriptive statistics techniques.
5. Interpret the data and make decisions about the population using inferential statistics.
6. Identify any possible errors.

A statistical study can usually be categorized as an observational study or an experiment. In an **observational study,** a researcher does not influence the responses. In an **experiment,** a researcher deliberately applies a treatment before observing the responses. Here is a brief summary of these types of studies.

- In an **observational study,** a researcher observes and measures characteristics of interest of part of a population but does not change existing conditions. For instance, an observational study was conducted in which researchers measured the amount of time people spent doing various activities, such as paid work, childcare, and socializing. *(Source: U.S. Bureau of Labor Statistics)*
- In performing an **experiment,** a **treatment** is applied to part of a population, called a **treatment group,** and responses are observed. Another part of the population may be used as a **control group,** in which no treatment is applied. (The subjects in both groups are called **experimental units.**) In many cases, subjects in the control group are given a **placebo,** which is a harmless, fake treatment that is made to look like the real treatment. The responses of both groups can then be compared and studied. In most cases, it is a good idea to use the same number of subjects for each group. For instance, an experiment was performed in which overweight subjects in a treatment group were given the artificial sweetener sucralose to drink while a control group drank water. After performing a glucose test, researchers concluded that "sucralose affects the glycemic and insulin responses" in overweight people who do not normally consume artificial sweeteners. *(Source: Diabetes Care)*

EXAMPLE 1

Distinguishing Between an Observational Study and an Experiment

Determine whether each study is an observational study or an experiment.

1. Researchers study the effect of vitamin D_3 supplementation among patients with antibody deficiency or frequent respiratory tract infections. To perform the study, 70 patients receive 4000 IU of vitamin D_3 daily for a year. Another group of 70 patients receive a placebo daily for one year. *(Source: British Medical Journal)*
2. Researchers conduct a study to determine how confident Americans are in the U.S. economy. To perform the study, researchers call 3040 U.S. adults and ask them to rate current U.S. economic conditions and whether the U.S. economy is getting better or worse. *(Source: Gallup)*

SOLUTION

1. Because the study applies a treatment (vitamin D_3) to the subjects, the study is an experiment.
2. Because the study does not attempt to influence the responses of the subjects (there is no treatment), the study is an observational study.

TRY IT YOURSELF 1

The Pennsylvania Game Commission conducted a study to count the number of elk in Pennsylvania. The commission captured and released 636 elk, which included 350 adult cows, 125 calves, 110 branched bulls, and 51 spikes. Is this study an observational study or an experiment? *(Source: Pennsylvania Game Commission)*

Answer: Page A31

Data Collection

There are several ways to collect data. Often, the focus of the study dictates the best way to collect data. Here is a brief summary of two methods of data collection.

- A **simulation** is the use of a mathematical or physical model to reproduce the conditions of a situation or process. Collecting data often involves the use of computers. Simulations allow you to study situations that are impractical or even dangerous to create in real life, and often they save time and money. For instance, automobile manufacturers use simulations with dummies to study the effects of crashes on humans. Throughout this course, you will have the opportunity to use applets that simulate statistical processes on a computer.
- A **survey** is an investigation of one or more characteristics of a population. Most often, surveys are carried out on *people* by asking them questions. The most common types of surveys are done by interview, Internet, phone, or mail. In designing a survey, it is important to word the questions so that they do not lead to biased results, which are not representative of a population. For instance, a survey is conducted on a sample of female physicians to determine whether the primary reason for their career choice is financial stability. In designing the survey, it would be acceptable to make a list of reasons and ask each individual in the sample to select her first choice.

Experimental Design

To produce meaningful unbiased results, experiments should be carefully designed and executed. It is important to know what steps should be taken to make the results of an experiment valid. Three key elements of a well-designed experiment are *control*, *randomization*, and *replication*.

Because experimental results can be ruined by a variety of factors, being able to control these influential factors is important. One such factor is a **confounding variable.**

DEFINITION

A **confounding variable** occurs when an experimenter cannot tell the difference between the effects of different factors on the variable.

Study Tip

The *Hawthorne effect* occurs in an experiment when subjects change their behavior simply because they know they are participating in an experiment.

For instance, to attract more customers, a coffee shop owner experiments by remodeling the shop using bright colors. At the same time, a shopping mall nearby has its grand opening. If business at the coffee shop increases, it cannot be determined whether it is because of the new colors or the new shopping mall. The effects of the colors and the shopping mall have been confounded.

Another factor that can affect experimental results is the *placebo effect*. The **placebo effect** occurs when a subject reacts favorably to a placebo when in fact the subject has been given a fake treatment. To help control or minimize the placebo effect, a technique called **blinding** can be used.

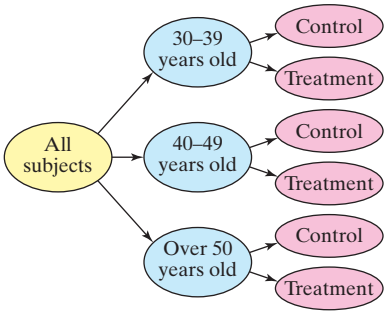
DEFINITION

Blinding is a technique where the subjects do not know whether they are receiving a treatment or a placebo. In a **double-blind experiment,** neither the experimenter nor the subjects know whether the subjects are receiving a treatment or a placebo. The experimenter is informed after all the data have been collected. This type of experimental design is preferred by researchers.

One challenge for experimenters is assigning subjects to groups so the groups have similar characteristics (such as age, height, weight, and so on). When treatment and control groups are similar, experimenters can conclude that any differences between groups is due to the treatment. To form groups with similar characteristics, experimenters use **randomization.**

DEFINITION

Randomization is a process of randomly assigning subjects to different treatment groups.



Randomized Block Design

In a **completely randomized design,** subjects are assigned to different treatment groups through random selection. In some experiments, it may be necessary for the experimenter to use **blocks,** which are groups of subjects with similar characteristics. A commonly used experimental design is a **randomized block design.** To use a randomized block design, the experimenter divides the subjects with similar characteristics into blocks, and then, within each block, randomly assign subjects to treatment groups. For instance, an experimenter who is testing the effects of a new weight loss drink may first divide the subjects into age categories such as 30–39 years old, 40–49 years old, and over 50 years old, and then, within each age group, randomly assign subjects to either the treatment group or the control group (see figure at the left).

Another type of experimental design is a **matched-pairs design,** where subjects are paired up according to a similarity. One subject in each pair is randomly selected to receive one treatment while the other subject receives a different treatment. For instance, two subjects may be paired up because of their age, geographical location, or a particular physical characteristic.

Sample size, which is the number of subjects in a study, is another important part of experimental design. To improve the validity of experimental results, **replication** is required.

Study Tip

The *validity* of an experiment refers to the accuracy and reliability of the experimental results. The results of a valid experiment are more likely to be accepted in the scientific community.

DEFINITION

Replication is the repetition of an experiment under the same or similar conditions.

For instance, suppose an experiment is designed to test a vaccine against a strain of influenza. In the experiment, 10,000 people are given the vaccine and another 10,000 people are given a placebo. Because of the sample size, the effectiveness of the vaccine would most likely be observed. But, if the subjects in the experiment are not selected so that the two groups are similar (according to age and gender), the results are of less value.

EXAMPLE 2

Analyzing an Experimental Design

A company wants to test the effectiveness of a new gum developed to help people quit smoking. Identify a potential problem with each experimental design and suggest a way to improve it.

1. The company identifies ten adults who are heavy smokers. Five of the subjects are given the new gum and the other five subjects are given a placebo. After two months, the subjects are evaluated and it is found that the five subjects using the new gum have quit smoking.
2. The company identifies one thousand adults who are heavy smokers. The subjects are divided into blocks according to gender. Females are given the new gum and males are given the placebo. After two months, a significant number of the female subjects have quit smoking.

SOLUTION

1. The sample size being used is not large enough to validate the results of the experiment. The experiment must be replicated to improve the validity.
2. The groups are not similar. The new gum may have a greater effect on women than on men, or vice versa. The subjects can be divided into blocks according to gender, but then, within each block, they should be randomly assigned to be in the treatment group or in the control group.

TRY IT YOURSELF 2

The company in Example 2 identifies 240 adults who are heavy smokers. The subjects are randomly assigned to be in a gum treatment group or in a control group. Each subject is also given a DVD featuring the dangers of smoking. After four months, most of the subjects in the treatment group have quit smoking. Identify a potential problem with the experimental design and suggest a way to improve it.

Answer: Page A31

Study Tip

A *biased sample* is one that is not representative of the population from which it is drawn. For instance, a sample consisting of only 18- to 22-year-old U.S. college students would not be representative of the entire 18- to 22-year-old population in the United States.

1.3 To explore this topic further, see **Activity 1.3** on page 27.

Tech Tip

You can use technology such as Minitab, Excel, StatCrunch, or the TI-84 Plus to generate random numbers. (Detailed instructions for using Minitab, Excel, and the TI-84 Plus are shown in the technology manuals that accompany this text.) For instance, here are instructions for using the random integer generator on a TI-84 Plus for Example 3.

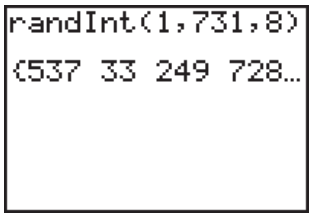
MATH

Choose the PRB menu.

5: randInt(

1 , 7 3 1 , 8)

ENTER



Continuing to press ENTER will generate more random samples of 8 integers.

Sampling Techniques

A **census** is a count or measure of an *entire* population. Taking a census provides complete information, but it is often costly and difficult to perform. A **sampling** is a count or measure of *part* of a population and is more commonly used in statistical studies. To collect unbiased data, a researcher must ensure that the sample is representative of the population. Appropriate sampling techniques must be used to ensure that inferences about the population are valid. Remember that when a study is done with faulty data, the results are questionable. Even with the best methods of sampling, a **sampling error** may occur. A sampling error is the difference between the results of a sample and those of the population. When you learn about inferential statistics, you will learn techniques of controlling sampling errors.

A **random sample** is one in which every member of the population has an equal chance of being selected. A **simple random sample** is a sample in which every possible sample of the same size has the same chance of being selected. One way to collect a simple random sample is to assign a different number to each member of the population and then use a random number table like Table 1 in Appendix B. Responses, counts, or measures for members of the population whose numbers correspond to those generated using the table would be in the sample. Calculators and computer software programs are also used to generate random numbers (see page 36).

Table 1—Random Numbers

92630	78240	19267	95457	53497	23894	37708	79862
79445	78735	71549	44843	26104	67318	00701	34986
59654	71966	27386	50004	05358	94031	29281	18544
31524	49587	76612	39789	13537	48086	59483	60680
06348	76938	90379	51392	55887	71015	09209	79157

Portion of Table 1 found in Appendix B

Consider a study of the number of people who live in West Ridge County. To use a simple random sample to count the number of people who live in West Ridge County households, you could assign a different number to each household, use a technology tool or table of random numbers to generate a sample of numbers, and then count the number of people living in each selected household.

EXAMPLE 3

Using a Simple Random Sample

There are 731 students currently enrolled in a statistics course at your school. You wish to form a sample of eight students to answer some survey questions. Select the students who will belong to the simple random sample.

SOLUTION

Assign numbers 1 to 731 to the students in the course. In the table of random numbers, choose a starting place at random and read the digits in groups of three (because 731 is a three-digit number). For instance, if you started in the third row of the table at the beginning of the second column, you would group the numbers as follows:

719 | 66 2 | 738 | 6 50 | 004 | 053 | 58 9 | 403 | 1 29 | 281 | 185 | 44

Ignoring numbers greater than 731, the first eight numbers are 719, 662, 650, 4, 53, 589, 403, and 129. The students assigned these numbers will make up the sample. To find the sample using a TI-84 Plus, follow the instructions shown at the left.

TRY IT YOURSELF 3

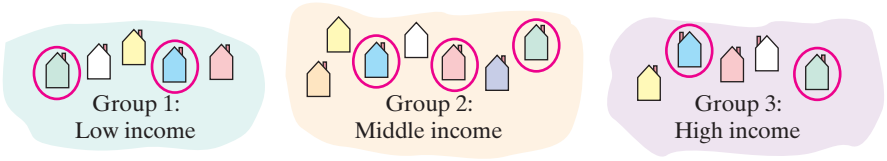
A company employs 79 people. Choose a simple random sample of five to survey.

Answer: Page A31

When you choose members of a sample, you should decide whether it is acceptable to have the same population member selected more than once. If it is acceptable, then the sampling process is said to be *with replacement*. If it is not acceptable, then the sampling process is said to be *without replacement*.

There are several other commonly used sampling techniques. Each has advantages and disadvantages.

- ***Stratified Sample*** When it is important for the sample to have members from each segment of the population, you should use a stratified sample. Depending on the focus of the study, members of the population are divided into two or more subsets, called *strata*, that share a similar characteristic such as age, gender, ethnicity, or even political preference. A sample is then randomly selected from each of the strata. Using a stratified sample ensures that each segment of the population is represented. For instance, to collect a stratified sample of the number of people who live in West Ridge County households, you could divide the households into socioeconomic levels and then randomly select households from each level. In using a stratified sample, care must be taken to ensure that all strata are sampled in proportion to their actual percentages of occurrence in the population. For instance, if 40% of the people in West Ridge County belong to the low-income group, then the proportion of the sample should have 40% from this group.



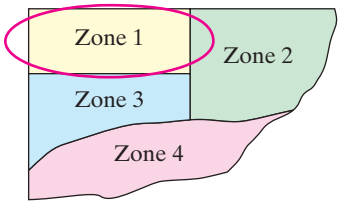
Stratified Sampling

Study Tip

Be sure you understand that stratified sampling randomly selects a *sample of members* from *all* strata. Cluster sampling uses *all members* from a randomly selected sample of *clusters* (but not all, so some clusters will not be part of the sample). For instance, in the figure for "Stratified Sampling" at the right, a *sample of households* in West Ridge County is randomly selected from *all* three income groups. In the figure for "Cluster Sampling," *all households* in a randomly selected *cluster* (Zone 1) are used. (Notice that the other zones are not part of the sample.)

- ***Cluster Sample*** When the population falls into naturally occurring subgroups, each having similar characteristics, a cluster sample may be the most appropriate. To select a cluster sample, divide the population into groups, called *clusters*, and select all of the members in one or more (but not all) of the clusters. Examples of clusters could be different sections of the same course or different branches of a bank. For instance, to collect a cluster sample of the number of people who live in West Ridge County households, divide the households into groups according to zip codes, then select all the households in one or more, but not all, zip codes and count the number of people living in each household. In using a cluster sample, care must be taken to ensure that all clusters have similar characteristics. For instance, if one of the zip code clusters has a greater proportion of high-income people, the data might not be representative of the population.

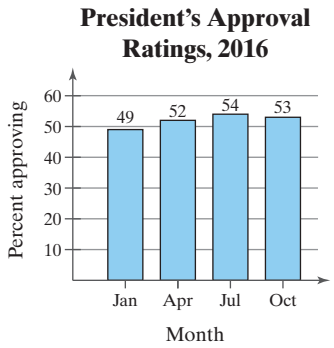
Zip Code Zones in West Ridge County



Cluster Sampling

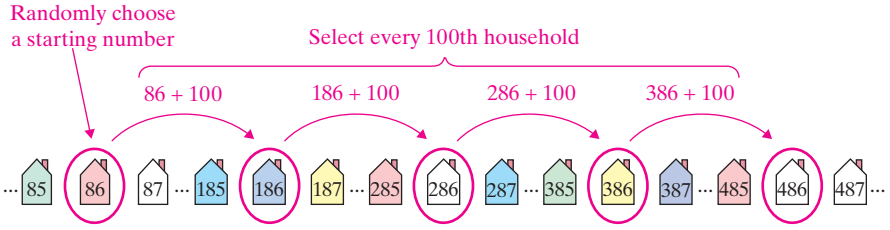
Picturing the World

The research firm Gallup conducts many polls (or surveys) regarding the president, Congress, and political and nonpolitical issues. A commonly cited Gallup poll is the public approval rating of the president. For instance, the approval ratings for President Barack Obama for selected months in 2016 are shown in the figure. (Each rating is from the poll conducted at the end of the indicated month.)



Discuss some ways that Gallup could select a biased sample to conduct a poll. How could Gallup select a sample that is unbiased?

- ***Systematic Sample*** A systematic sample is a sample in which each member of the population is assigned a number. The members of the population are ordered in some way, a starting number is randomly selected, and then sample members are selected at regular intervals from the starting number. (For instance, every 3rd, 5th, or 100th member is selected.) For instance, to collect a systematic sample of the number of people who live in West Ridge County households, you could assign a different number to each household, randomly choose a starting number, select every 100th household, and count the number of people living in each. An advantage of systematic sampling is that it is easy to use. In the case of any regularly occurring pattern in the data, however, this type of sampling should be avoided.



Systematic Sampling

A type of sample that often leads to biased studies (so it is not recommended) is a **convenience sample.** A convenience sample consists only of members of the population that are easy to get.

EXAMPLE 4

Identifying Sampling Techniques

You are doing a study to determine the opinions of students at your school regarding stem cell research. Identify the sampling technique you are using when you select the samples listed. Discuss potential sources of bias (if any).

1. You divide the student population with respect to majors and randomly select and question some students in each major.
2. You assign each student a number and generate random numbers. You then question each student whose number is randomly selected.
3. You select students who are in your biology class.

SOLUTION

1. Because students are divided into strata (majors) and a sample is selected from each major, this is a stratified sample.
2. Each sample of the same size has an equal chance of being selected and each student has an equal chance of being selected, so this is a simple random sample.
3. Because the sample is taken from students that are readily available, this is a convenience sample. The sample may be biased because biology students may be more familiar with stem cell research than other students and may have stronger opinions.

TRY IT YOURSELF 4

You want to determine the opinions of students regarding stem cell research. Identify the sampling technique you are using when you select these samples.

1. You select a class at random and question each student in the class.
2. You assign each student a number and, after choosing a starting number, question every 25th student.

Answer: Page A31

1.3 EXERCISES

For Extra Help: MyLab Statistics

Building Basic Skills and Vocabulary

1. What is the difference between an observational study and an experiment?
2. What is the difference between a census and a sampling?
3. What is the difference between a random sample and a simple random sample?
4. What is replication in an experiment? Why is replication important?

True or False? *In Exercises 5–10, determine whether the statement is true or false. If it is false, rewrite it as a true statement.*

5. A placebo is an actual treatment.
6. A double-blind experiment is used to increase the placebo effect.
7. Using a systematic sample guarantees that members of each group within a population will be sampled.
8. A convenience sample is always representative of a population.
9. The method for selecting a stratified sample is to order a population in some way and then select members of the population at regular intervals.
10. To select a cluster sample, divide a population into groups and then select all of the members in at least one (but not all) of the groups.

Distinguishing Between an Observational Study and an Experiment *In Exercises 11–14, determine whether the study is an observational study or an experiment. Explain.*

11. In a survey of 1033 U.S. adults, 51% said U.S. presidents should release all medical information that might affect their ability to serve. *(Source: Gallup)*
12. Researchers demonstrated that adults using an intensive program to lower systolic blood pressure to less than 120 millimeters of mercury reduce the risk of death from all causes by 27%. *(Source: American Heart Association)*
13. To study the effects of social media on teenagers' brains, researchers showed a few dozen teenagers photographs that had varying numbers of "likes" while scanning the reactions in their brains. *(Source: NPR)*
14. In a study designed to research the effect of music on driving habits, 1000 motorists ages 17–25 years old were asked whether the music they listened to influenced their driving. *(Source: More Th>n)*
15. **Random Number Table** Use the sixth row of Table 1 in Appendix B to generate 12 random numbers between 1 and 99.
16. **Random Number Table** Use the tenth row of Table 1 in Appendix B to generate 10 random numbers between 1 and 920.

Random Numbers *In Exercises 17 and 18, use technology to generate the random numbers.*

17. Fifteen numbers between 1 and 150
18. Nineteen numbers between 1 and 1000

Using and Interpreting Concepts

19. Allergy Drug A pharmaceutical company wants to test the effectiveness of a new drug used to treat migraine headaches. The company identifies 500 females ages 25 to 45 years old who suffer from migraine headaches. The subjects are randomly assigned into two groups. One group is given the drug and the other is given a placebo that looks exactly like the drug. After three months, the subjects' symptoms are studied and compared.

(a) Identify the experimental units and treatments used in this experiment.

(b) Identify a potential problem with the experimental design being used and suggest a way to improve it.

(c) How could this experiment be designed to be double-blind?

20. Dietary Supplement Researchers in Germany tested the effect of a dietary supplement designed to control metabolism in patients with type 2 diabetes. Thirty-one patients with type 2 diabetes completed the study. The patients were assigned at random either the supplement or a placebo for 12 weeks. After a subsequent "wash-out" period of 12 weeks, the patients were assigned the other product. At the conclusion of the study, the patients' glycated hemoglobin, fasting blood glucose, and fructosamine levels were checked, as well as their lipid parameters. *(Source: Food and Nutrition Research)*

(a) Identify the experimental units and treatments used in this experiment.

(b) Identify a potential problem with the experimental design being used and suggest a way to improve it.

(c) The experiment is described as a placebo-controlled, double-blind study. Explain what this means.

(d) How could blocking be used in designing this experiment?

21. Sleep Deprivation A researcher wants to study the effects of sleep deprivation on motor skills. Eighteen people volunteer for the experiment: Jake, Maria, Mike, Lucy, Ron, Adam, Bridget, Carlos, Steve, Susan, Vanessa, Rick, Dan, Kate, Pete, Judy, Mary, and Connie. Use a random number generator to choose nine subjects for the treatment group. The other nine subjects will go into the control group. List the subjects in each group. Tell which method you used to generate the random numbers.

22. Using a Simple Random Sample Volunteers for an experiment are numbered from 1 to 90. The volunteers are to be randomly assigned to two different treatment groups. Use a random number generator different from the one you used in Exercise 21 to choose 45 subjects for the treatment group. The other 45 subjects will go into the control group. List the subjects, according to number, in each group. Tell which method you used to generate the random numbers.

Identifying Sampling Techniques *In Exercises 23–28, identify the sampling technique used, and discuss potential sources of bias (if any). Explain.*

23. Selecting employees at random from an employee directory, researchers contact 300 people and ask what obstacles (such as computer problems) keep them from accomplishing tasks at work.

24. Questioning university students as they leave a fraternity party, a researcher asks 463 students about their study habits.

25. After a hurricane, a disaster area is divided into 200 equal grids. Thirty of the grids are selected, and every occupied household in the grid is interviewed to help focus relief efforts on what residents require the most.

26. Every tenth person entering a mall is asked to name his or her favorite store.

27. Soybeans are planted on a 48-acre field. The field is divided into one-acre subplots. A sample is taken from each subplot to estimate the harvest.

28. From calls made with randomly generated telephone numbers, 1012 respondents are asked if they rent or own their residences.

Choosing Between a Census and a Sampling *In Exercises 29 and 30, determine whether you would take a census or use a sampling. If you would use a sampling, determine which sampling technique you would use. Explain.*

29. The average age of the 115 residents of a retirement community

30. The most popular type of movie among 100,000 online movie rental subscribers

Recognizing a Biased Question *In Exercises 31–34, determine whether the survey question is biased. If the question is biased, suggest a better wording.*

31. Why does eating whole-grain foods improve your health?

32. Why does text messaging while driving increase the risk of a crash?

33. How much do you exercise during an average week?

34. How does the media influence the opinions of voters?

Extending Concepts

35. Analyzing a Study Find an article or a news story that describes a statistical study.

(a) Identify the population and the sample.

(b) Classify the data as qualitative or quantitative. Determine the level of measurement.

(c) Is the study an observational study or an experiment? If it is an experiment, identify the treatment.

(d) Identify the sampling technique used to collect the data.

36. Designing and Analyzing a Study Design a study for some subject that is of interest to you. Answer parts (a)–(d) of Exercise 35 for this study.

37. Open and Closed Questions Two types of survey questions are open questions and closed questions. An open question allows for any kind of response; a closed question allows for only a fixed response. An open question and a closed question with its possible choices are given below. List an advantage and a disadvantage of each question.

Open Question What can be done to get students to eat healthier foods?

Closed Question How would you get students to eat healthier foods?

1. Mandatory nutrition course

2. Offer only healthy foods in the cafeteria and remove unhealthy foods

3. Offer more healthy foods in the cafeteria and raise the prices on unhealthy foods

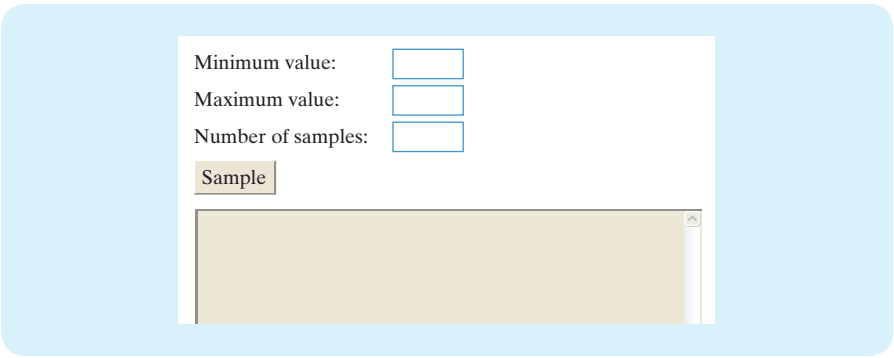
38. Natural Experiments Observational studies are sometimes referred to as *natural experiments*. Explain, in your own words, what this means.

1.3 ACTIVITY Random Numbers

APPLET

You can find the interactive applet for this activity within MyLab Statistics or at *www.pearsonhighered.com/mathstatsresources.*

The *random numbers* applet is designed to allow you to generate random numbers from a range of values. You can specify integer values for the minimum value, maximum value, and the number of samples in the appropriate fields. You should not use decimal points when filling in the fields. When SAMPLE is clicked, the applet generates random values, which are displayed as a list in the text field.



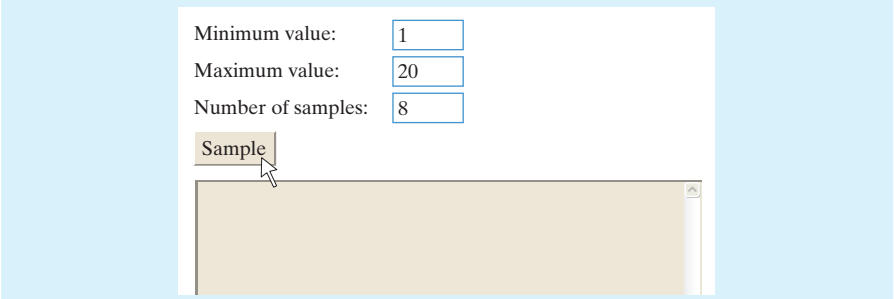
EXPLORE

Step 1 Specify a minimum value.
Step 2 Specify a maximum value.
Step 3 Specify the number of samples.
Step 4 Click SAMPLE to generate a list of random values.

DRAW CONCLUSIONS

APPLET

1. Specify the minimum, maximum, and number of samples to be 1, 20, and 8, respectively, as shown. Run the applet. Continue generating lists until you obtain one that shows that the random sample is taken with replacement. Write down this list. How do you know that the list is a random sample taken with replacement?



2. Use the applet to repeat Example 3 on page 21. What values did you use for the minimum, maximum, and number of samples? Which method do you prefer? Explain.

USES AND ABUSES

Statistics in the Real World

Uses

An experiment studied 321 women with advanced breast cancer. All of the women had been previously treated with other drugs, but the cancer had stopped responding to the medications. The women were then given the opportunity to take a new drug combined with a chemotherapy drug.

The subjects were divided into two groups, one that took the new drug combined with a chemotherapy drug, and one that took only the chemotherapy drug. After three years, results showed that the new drug in combination with the chemotherapy drug delayed the progression of cancer in the subjects. The results were so significant that the study was stopped, and the new drug was offered to all women in the study. The Food and Drug Administration has since approved use of the new drug in conjunction with a chemotherapy drug.

Abuses

For four years, one hundred eighty thousand teenagers in Norway were used as subjects to test a new vaccine against the deadly bacteria *meningococcus b*. A brochure describing the possible effects of the vaccine stated, "it is unlikely to expect serious complications," while information provided to the Norwegian Parliament stated, "serious side effects can not be excluded." The vaccine trial had some disastrous results: More than 500 side effects were reported, with some considered serious, and several of the subjects developed serious neurological diseases. The results showed that the vaccine was providing immunity in only 57% of the cases. This result was not sufficient for the vaccine to be added to Norway's vaccination program. Compensations have since been paid to the vaccine victims.

Ethics

Experiments help us further understand the world that surrounds us. But, in some cases, they can do more harm than good. In the Norwegian experiments, several ethical questions arise. Was the Norwegian experiment unethical if the best interests of the subjects were neglected? When should the experiment have been stopped? Should it have been conducted at all? When serious side effects are not reported and are withheld from subjects, there is no ethical question here, it is just wrong.

On the other hand, the breast cancer researchers would not want to deny the new drug to a group of patients with a life-threatening disease. But again, questions arise. How long must a researcher continue an experiment that shows better-than-expected results? How soon can a researcher conclude a drug is safe for the subjects involved?

EXERCISES

1. Find an example of a real-life experiment other than the one described above that may be considered an "abuse." What could have been done to avoid the outcome of the experiment?

2. ***Stopping an Experiment*** In your opinion, what are some problems that may arise when clinical trials of a new experimental drug or vaccine are stopped early and then the drug or vaccine is distributed to other subjects or patients?

1 Chapter Summary

What Did You Learn?	Example(s)	Review Exercises
Section 1.1		
How to distinguish between a population and a sample	1	1–4
How to distinguish between a parameter and a statistic	2	5–8
How to distinguish between descriptive statistics and inferential statistics	3	9, 10
Section 1.2		
How to distinguish between qualitative data and quantitative data	1	11–14
How to classify data with respect to the four levels of measurement: nominal, ordinal, interval, and ratio	2, 3	15–18
Section 1.3		
How to design a statistical study and how to distinguish between an observational study and an experiment	1	19, 20
How to design an experiment	2	21, 22
How to create a sample using random sampling, simple random sampling, stratified sampling, cluster sampling, and systematic sampling and how to identify a biased sample	3, 4	23–29

Level of measurement	Put data in categories	Arrange data in order	Subtract data entries	Determine whether one data entry is a multiple of another
Nominal	Yes	No	No	No
Ordinal	Yes	Yes	No	No
Interval	Yes	Yes	Yes	No
Ratio	Yes	Yes	Yes	Yes

Sampling Techniques
Random: A sample in which every member of a population has an equal chance of being selected.
Simple random: A sample in which every possible sample of the same size has the same chance of being selected from a population.
Stratified: Members of a population are divided into two or more subsets, called strata, that share a similar characteristic. A *sample* is then randomly selected from *each* of the strata. Using a stratified sample ensures that each segment of the population is represented.
Cluster: The population is divided into groups (or clusters) and *all of the members in one or more* (but not all) of the clusters are selected. To avoid a biased sample, care must be taken to ensure that all clusters have similar characteristics.
Systematic: Each member of a population is assigned a number. The members of the population are ordered in some way, a starting number is randomly selected, and then sample members are selected at regular intervals from the starting number. (For instance, every 3rd, 5th, or 100th member is selected.)

1 Review Exercises

Section 1.1

In Exercises 1–4, identify the population and the sample. Describe the sample data set.

1. A survey of 4787 U.S. adults found that 15% use ride-hailing applications. *(Source: Pew Research Center)*

2. Eighty-three doctors working in the St. Louis area were surveyed concerning their opinions of health care reform.

3. A survey of 2223 U.S. adults found that 62% would encourage a child to pursue a career as a video game developer or designer. *(Source: The Harris Poll)*

4. A survey of 1601 U.S. children and adults ages 16 years and older found that 48% have visited a public library or a bookmobile over a recent span of 12 months. *(Source: Pew Research Center)*

In Exercises 5–8, determine whether the number describes a population parameter or a sample statistic. Explain your reasoning.

5. In 2016, the National Science Foundation announced $22.7 million in infrastructure-strengthening investments. *(Source: National Science Foundation)*

6. In a survey of 1000 likely U.S. voters, 29% trust media fact-checking of candidates' comments. *(Source: Rasmussen Reports)*

7. In a study of math majors at a university, 10 students minored in physics.

8. Thirty percent of a sample of 521 U.S. workers say that they worry about having their benefits reduced. *(Source: Gallup)*

9. Which part of the survey described in Exercise 3 represents the descriptive branch of statistics? Make an inference based on the results of the survey.

10. Which part of the survey described in Exercise 4 represents the descriptive branch of statistics? Make an inference based on the results of the survey.

Section 1.2

In Exercises 11–14, determine whether the data are qualitative or quantitative. Explain your reasoning.

11. The ages of a sample of 350 employees of a software company

12. The zip codes of a sample of 200 customers at a sporting goods store

13. The revenues of the companies on the Fortune 500 list

14. The marital statuses of all engineers at an electric utility

In Exercises 15–18, determine the level of measurement of the data set. Explain.

15. The daily high temperatures (in degrees Fahrenheit) for Sacramento, California, for a week in September are listed. *(Source: National Climatic Data Center)*

90 80 76 84 91 94 97

16. The vehicle size classes for a sample of sedans are listed.

Minicompact Subcompact Compact Mid-size Large

17. The four departments of a printing company are listed.

Administration Sales Production Billing

18. The total compensations (in millions of dollars) of the ten highest-paid CEOs at U.S. public companies are listed. *(Source: Equilar, Inc.)*

94.6 56.4 54.1 53.2 53.2 51.6 47.5 43.5 39.2 37.0

Section 1.3

In Exercises 19 and 20, determine whether the study is an observational study or an experiment. Explain.

19. Researchers conduct a study to determine whether a drug used to treat hypertension in patients with obstructive sleep apnea works better when taken in the morning or in the evening. To perform the study, 78 patients are given one pill to take in the morning and one pill to take in the evening (one containing the drug and the other a placebo). After 6 weeks, researchers collected blood pressure information on the patients. *(Source: American Thoracic Society)*

20. Researchers conduct a study to determine the effect of coffee consumption on the development of multiple sclerosis. To perform the study, researchers asked 4408 adults in Sweden and 2331 adults in the United States how many cups of coffee they drink per day. *(Source: American Association for the Advancement of Science)*

In Exercises 21 and 22, two hundred students volunteer for an experiment to test the effects of sleep deprivation on memory recall. The students will be placed in one of five different treatment groups, including the control group.

21. Explain how you could design an experiment so that it uses a randomized block design.

22. Explain how you could design an experiment so that it uses a completely randomized design.

In Exercises 23–28, identify the sampling technique used, and discuss potential sources of bias (if any). Explain.

23. Using random digit dialing, researchers ask 1201 U.S. adults whether enough is being done to fight opioid addiction. *(Source: Kaiser Family Foundation)*

24. A student asks 18 friends to participate in a psychology experiment.

25. A study in a town in northwest Ethiopia designed to determine prevalence and predictors of depression among pregnant women randomly selects four districts of the town, then interviews all pregnant women in these districts. *(Source: Public Library of Science)*

26. Law enforcement officials stop and check the driver of every third vehicle for blood alcohol content.

27. Twenty-five students are randomly selected from each grade level at a high school and surveyed about their study habits.

28. A journalist interviews 154 people waiting at an airport baggage claim and asks them how safe they feel during air travel.

29. You want to know the favorite spring break destination among 15,000 students at a university. Determine whether you would take a census or use a sampling. If you would use a sampling, determine which sampling technique you would use. Explain your reasoning.

1 Chapter Quiz

Take this quiz as you would take a quiz in class. After you are done, check your work against the answers given in the back of the book.

1. A study of the dietary habits of 359,264 Korean adolescents was conducted to find a link between dietary habits and school performance. Identify the population and the sample in the study. *(Source: Wolters Kluwer Health, Inc.)*

2. Determine whether each number describes a population parameter or a sample statistic. Explain your reasoning.

(a) A survey of 1000 U.S. adults found that 52% think that the introduction of driverless cars will make roads less safe. *(Source: Rasmussen Reports)*

(b) At a college, 90% of the members of the Board of Trustees approved the contract of the new president.

(c) A survey of 727 small business owners found that 25% reported job openings they could not fill. *(Source: National Federation of Independent Business)*

3. Determine whether the data are qualitative or quantitative. Explain.

(a) A list of debit card personal identification numbers

(b) The final scores on a video game

4. Determine the level of measurement of the data set. Explain your reasoning.

(a) A list of badge numbers of police officers at a precinct

(b) The horsepowers of racing car engines

(c) The top 10 grossing films released in a year

(d) The years of birth for the runners in the Boston marathon

5. Determine whether the study is an observational study or an experiment. Explain.

(a) Researchers conduct a study to determine whether body mass index (BMI) influences mortality. To conduct the study, researchers obtained the BMIs of 3,951,455 people. *(Source: Elsevier, Ltd.)*

(b) Researchers conduct a study to determine whether taking a multivitamin daily affects cognitive health among men as they age. To perform the study, researchers studied 5947 male physicians ages 65 years or older and had one group take a multivitamin daily and had another group take a placebo daily. *(Source: American College of Physicians)*

6. An experiment is performed to test the effects of a new drug on high blood pressure. The experimenter identifies 320 people ages 35–50 years old with high blood pressure for participation in the experiment. The subjects are divided into equal groups according to age. Within each group, subjects are then randomly selected to be in either the treatment group or the control group. What type of experimental design is being used for this experiment?

7. Identify the sampling technique used in each study. Explain your reasoning.

(a) A journalist asks people at a campground about air pollution.

(b) For quality assurance, every tenth machine part is selected from an assembly line and measured for accuracy.

(c) A study on attitudes about smoking is conducted at a college. The students are divided by class (freshman, sophomore, junior, and senior). Then a random sample is selected from each class and interviewed.

8. Which technique used in Exercise 7 could lead to a biased study? Explain.