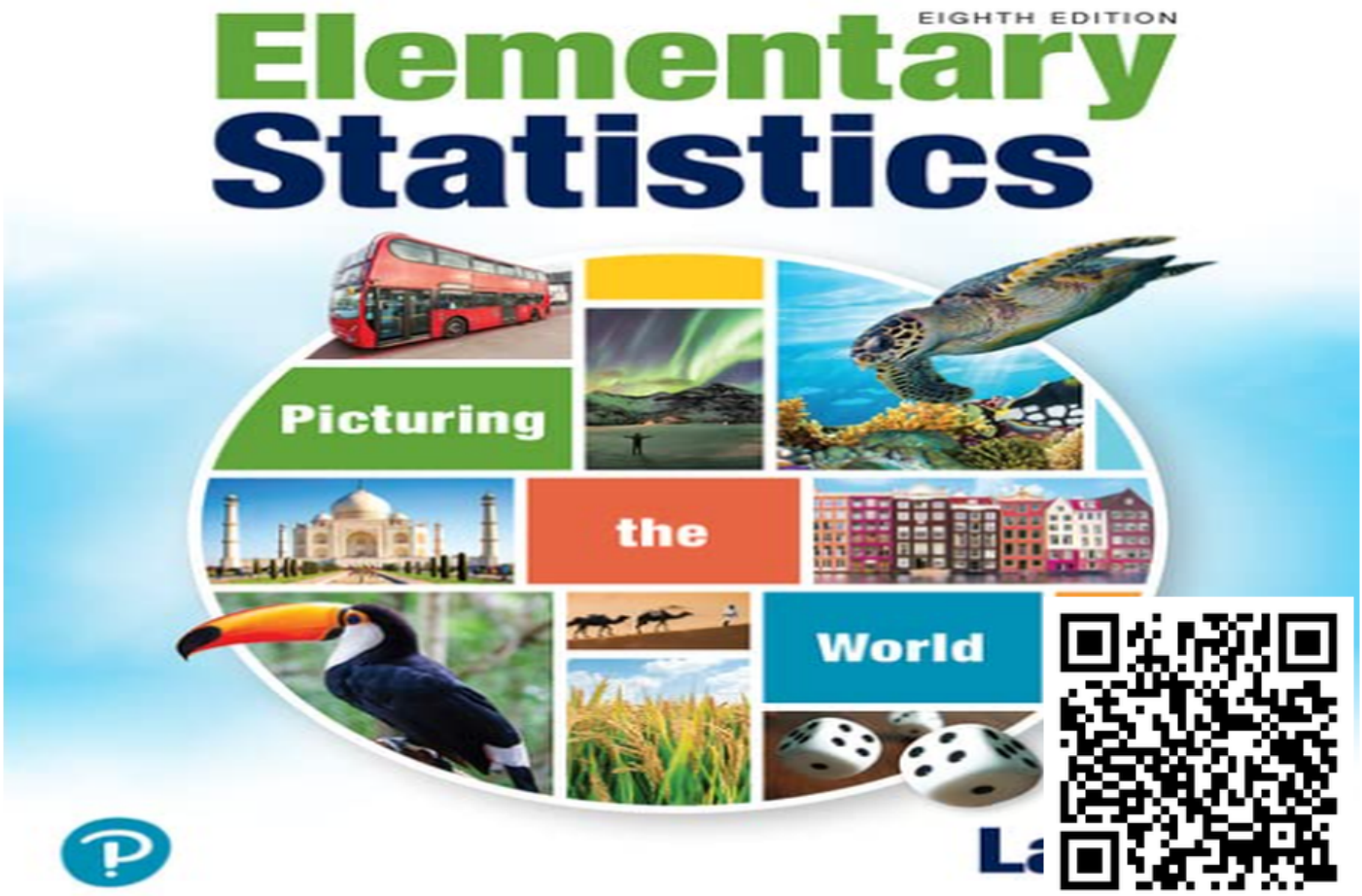


EIGHTH EDITION

Elementary Statistics

Ron Larson

This page intentionally left blank

EIGHTH EDITION

Elementary Statistics

PICTURING THE WORLD

Ron Larson
The Pennsylvania State University
The Behrend College

Pearson

Content Development: *Robert Carroll*
Content Management: *Suzanna Smith-Bainbridge, Amanda Brands Moschberger*
Content Production: *Noelle Saligumba, Peggy McMahon*
Product Management: *Karen Montgomery*
Product Marketing: *Demetrius Hall*
Rights and Permissions: *Tanvi Bhatia/Anjali Singh*
Please contact https://support.pearson.com/getsupport/s/ with any queries on this content
Cover images by Shutterstock.

Library of Congress Control Number: 2021921349

ScoutAutomatedPrintCode

ISBN 10: 0-13-749332-0
ISBN 13: 978-0-13-749332-6

Pearson's Commitment to Diversity, Equity, and Inclusion

Pearson is dedicated to creating bias-free content that reflects the diversity of all learners. We embrace the many dimensions of diversity, including but not limited to race, ethnicity, gender, socioeconomic status, ability, age, sexual orientation, and religious or political beliefs.

Education is a powerful force for equity and change in our world. It has the potential to deliver opportunities that improve lives and enable economic mobility. As we work with authors to create content for every product and service, we acknowledge our responsibility to demonstrate inclusivity and incorporate diverse scholarship so that everyone can achieve their potential through learning. As the world's leading learning company, we have a duty to help drive change and live up to our purpose to help more people create a better life for themselves and to create a better world.

Our ambition is to purposefully contribute to a world where:

- Everyone has an equitable and lifelong opportunity to succeed through learning.
- Our educational products and services are inclusive and represent the rich diversity of learners.
- Our educational content accurately reflects the histories and experiences of the learners we serve.
- Our educational content prompts deeper discussions with students and motivates them to expand their own learning (and worldview).

We are also committed to providing products that are fully accessible to all learners. As per Pearson's guidelines for accessible educational Web media, we test and retest the capabilities of our products against the highest standards for every release, following the WCAG guidelines in developing new products for copyright year 2022 and beyond. You can learn more about Pearson's commitment to accessibility at https://www.pearson.com/us/accessibility.html.

While we work hard to present unbiased, fully accessible contact, we want to hear from you about any concerns or needs with this Pearson product so that we can investigate and address them.

- Please contact us with concerns about any potential bias at https://www.pearson.com/report-bias.html.
- For accessibility-related issues, such as using assistive technology with Pearson products, alternative text requests, or accessibility documentation, email the Pearson Disability Support team at disability.support@pearson.com.

This page intentionally left blank

CONTENTS

PART 2 PROBABILITY AND PROBABILITY DISTRIBUTIONS

3 Probability 128

4 Discrete Probability Distributions 188

5 Normal Probability Distributions 232

PART 3 STATISTICAL INFERENCE

6 Confidence Intervals 296

PART 4 MORE STATISTICAL INFERENCE

11 Nonparametric Tests (Online Only)*

Where You've Been Where You're Going

* Available in MyLab Statistics

Appendices

PREFACE

Welcome to *Elementary Statistics: Picturing the World*, Eighth Edition. You will find that this textbook is written with a balance of rigor and simplicity. It combines step-by-step instructions, 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 seven 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 Eighth Edition was a thorough update of the key features, examples, and exercises:

Examples This edition has 213 examples, nearly 50% of which are new or revised. Also, several of the examples now show an alternate solution or a check using technology.

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

Picturing the World Over 70% of these are new or revised.

Screen Displays In the examples, technology tips, and other features that show screen displays from Minitab®, Excel®, the TI-84 Plus, and StatCrunch®, the displays were revised as appropriate to make them more visually appealing, easy to follow, and reflective of the most up-to-date version of the software.

Exercises Over 30% of the more than 2300 exercises are new or revised.

Extensive Chapter Feature Updates A full 50% 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

References to Co-Requisite Help Margin notes have been included at point-of-use locations throughout this edition to remind students that they can get help reviewing a particular area of mathematics in the Integrated Review in MyLab Statistics.

Applet Activities Revisions have been made to the applet activities throughout the text to reflect changes to the corresponding online applets they reference. Applet activities are discussed further on the next page.

Study Strategies At the bottom of each chapter summary page in Chapters 1 through 10, there are study strategies that students can use to help improve their performance in college. These include tips on improving reading skills, avoiding procrastination, preparing for a test, taking notes, and other areas.

Features of the Eighth 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.
- **References to Co-Requisite Help** point students to extra math help.
- **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 available in MyLab Statistics in the *Student 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 Seventh 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.

MyLab Statistics Resources for Success

MyLab Statistics is available to accompany Pearson's market-leading text options, including *Elementary Statistics: Picturing The World*, 8e (access code required).

MyLab™ is the teaching and learning platform that empowers you to reach every student. MyLab Statistics combines trusted author content—including full eText and assessment with immediate feedback—with digital tools and a flexible platform to personalize the learning experience and improve results for each student. Integrated with StatCrunch®, a web-based statistical software program, students learn the skills they need to interact with data in the real world.

MyLab Statistics supports all learners, regardless of their ability and background, to provide an equal opportunity for success. Accessible resources support learners for a more equitable experience no matter their abilities. And options to personalize learning and address individual gaps helps to provide each learner with the specific resources they need to achieve success.

Student Resources

Each student learns at a different pace. Personalized learning pinpoints the precise areas where each student needs practice, giving all students the support they need—when and where they need it—to be successful.

StatCrunch® is integrated directly into MyLab Statistics. StatCrunch® is a powerful web-based statistical software that allows users to perform complex analyses, share data sets, and generate compelling reports of their data. The vibrant online community offers tens of thousands of shared data sets for students to analyze.

- **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. 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 allows 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 can be accessed on your laptop, smartphone, or tablet when you visit the StatCrunch website from your device's browser. For more information, visit the StatCrunch website, or contact your Pearson representative.

Exercises with Immediate Feedback The exercises in MyLab Statistics reflect the approach and learning style of this text, and regenerate algorithmically to give student unlimited opportunity for practice and mastery. Most exercises include learning aids, such as guided solutions and sample problems, and they offer helpful feedback when students enter incorrect answers.

Personalized Homework With Personalized Homework, students take a quiz or test and receive a subsequent homework assignment that is personalized based on their performance. This way, students can focus on just the topics they have not yet mastered.

Integrated Review *Elementary Statistics, Picturing the World with Integrated Review* can be used in corequisite courses, or simply to help students who enter without a full understanding of prerequisite skills and concepts.

MyLab courses provide the full suite of supporting resources for the Statistics course, plus additional assignments and for study aids from select intermediate algebra topics for students who will benefit from remediation.

Assignments for the integrated review content are pre-assigned in MyLab, making it easier than ever to create your course.

Mindset videos and assignable, open-ended **exercises** foster a growth mindset in students. This material encourages them to maintain a positive attitude about learning, value their own ability to grow, and view mistakes as learning opportunities—so often a hurdle for math students.

Personal Inventory Assessments are a collection of online exercises designed to promote self reflection and metacognition in students. These 33 assessments include topics such as a Stress Management Assessment, Diagnosing Poor Performance and Enhancing Motivation, and Time Management Assessment.

Instructor Resources

Your course is unique. So whether you'd like to build your own assignments, teach multiple sections, or set prerequisites, MyLab gives you the flexibility to easily create your course to fit your needs.

MyLab Features

Performance Analytics enable instructors to see and analyze student performance across multiple courses. Based on their current course progress, the student's performance is identified as above, at, or below expectations through a variety of graphs and visualizations.

Conceptual Question Library There are 1000 questions in the Assignment Manager that require students to apply their statistical understanding.

PowerPoint Presentations include lecture content and key graphics from the textbook. Accessible PowerPoint slides are also available and are built to align with WCAG 2.0 AA standards and Section 508 guidelines.

TestGen® (www.pearsoned.com/testgen) enables instructors to build, edit, print, and administer tests using a computerized bank of questions developed to cover the objectives of the text.

Test Bank features printable PDF containing all the test exercises available in TestGen.

Accessibility Pearson works continuously to ensure our products are as accessible as possible to all students. Currently we work toward achieving WCAG 2.0 AA for our existing products (2.1 AA for future products) and Section 508 standards, as expressed in the Pearson Guidelines for Accessible Educational Web Media (https://www.pearson.com/accessibility-guidelines.html).

Minitab

Minitab™ makes 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 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

ACKNOWLEDGMENTS

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

Reviewers of the Current Edition

Chris Bendixen, Lake Michigan College
Seunghee Lee, Pellissippi State Community College
Nancy Liu, Miami Dade College
Lohuwa Mamadu, University of South Florida
Ashley Nicoloff, Glendale Community College
Jason Samuels, Borough of Manhattan Community 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
Karen Benway, University of Vermont
David P. Benzel, Montgomery College
John Bernard, University of Texas—Pan American
B.K. Brinkley, Tidewater Community College
G. Andy Chang, Youngstown State University
Keith J. Craswell, Western Washington University
Carol Curtis, Fresno City College
Christine Curtis, Hillsborough Community College–Dale Mabry
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
Carrie Elledge, San Juan 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
Jason Malozzi, Lower Columbia College
Mike McGann, Ventura Community College
Cynthia McGinnis, Northwest Florida State College
Vicki McMillian, Ocean County College
Lynn Meslinsky, Erie Community College
Larry Musolino, Pennsylvania State University
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
Cyndi Roemer, Union County College
Neal Rogness, Grand Valley State University
Jean Rowley, American Public University and DeVry 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
Heidi Webb, Horry Georgetown Technical College
Dex Whittinghall, Rowan University
Cathleen Zucco-Teveloff, Rider University

Many thanks to Betsy Farber for her significant contributions to previous editions of the text.

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

Ron Larson, odx@psu.edu

INDEX OF APPLICATIONS

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

During 2020, the fastest-growing state in the United States was Idaho. In the same year, the Idaho cities of Meridian and Nampa were among the 10 fastest-growing cities in the United States.

Where You've Been

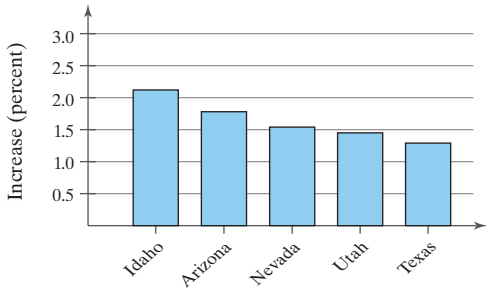
You are already familiar with many of the practices of statistics, such as taking surveys, collecting data, and describing populations. What you may not know is that collecting accurate statistical data is often difficult and costly. Consider, for instance, the monumental task of counting and describing 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

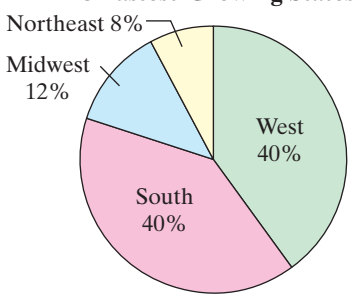
In Chapter 1, you will be introduced to the basic concepts and goals of statistics. For instance, statistics were used to construct the figures below, which show the fastest-growing U.S. states from 2019 to 2020 by the percent increase in population and by the numerical increase in population, along with the regions where these states are located.

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.

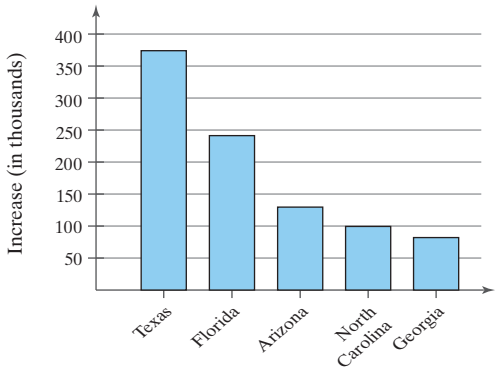
Fastest-Growing States (2019 to 2020)



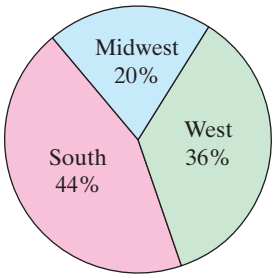
Regions of the 25 Fastest-Growing States



States with Greatest Numerical Population Increases (2019 to 2020)



Regions of the 25 States with Greatest Numerical Population Increases



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

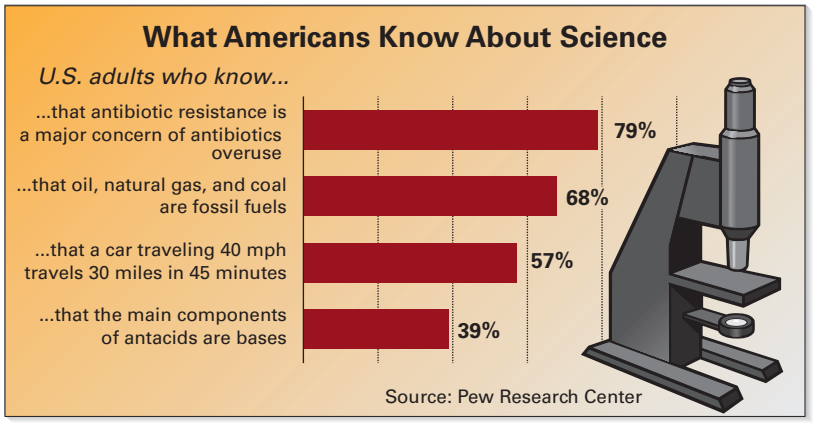
Almost every day you are exposed to statistics. For instance, consider the next two statements.

- "7 in 10 Americans believe the arts unify their communities, and 2 in 5 Americans have changed an opinion or perception based on an arts experience." *(Source: Americans for the Arts)*
- "Notably, 21% of 8–11 year-olds have a social media profile." *(Source: Smart Insights, Ltd.)*

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.

For help with percents and reading graphs, see *Integrated Review* at MyLab® Statistics



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 4464 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 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 Statistics 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.)

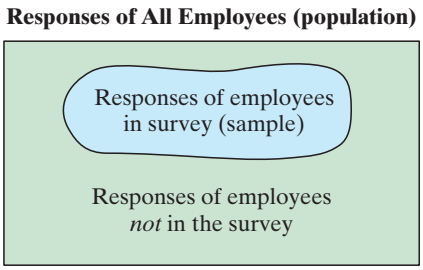
EXAMPLE 1

Identifying Data Sets

In a survey, 751 employees in the United States were asked how stressed they feel at work. Of the 751 respondents, 616 said that they feel at least a little stressed at work. Identify the population and the sample. Describe the sample data set. *(Adapted from The Marlin Company)*

SOLUTION

The population consists of the responses of all employees in the United States. The sample consists of the responses of the 751 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 616 employees who said that they feel at least a little stressed at work and 135 who said that they do not feel stressed at work.



TRY IT YOURSELF 1

In a survey of 1516 teens in the United States, 1228 said "mental health is a significant issue for young people in the U.S." Identify the population and the sample. Describe the sample data set. *(Adapted from National 4-H Council)*

Answer: Page A35

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 health care 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

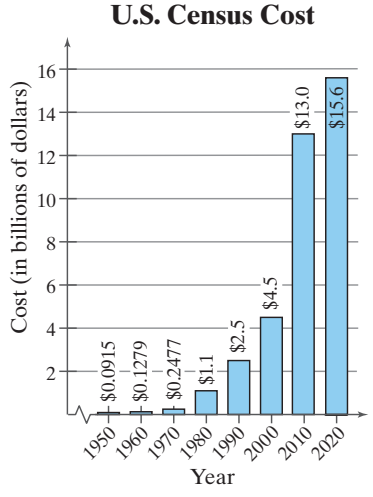
A **parameter** is a numerical description of a *population* characteristic.

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 616 of 751 employees surveyed feel at least a little stressed at work. Another sample may have a different number of employees who say they feel at least a little stressed at work. For the population, however, the number of employees who feel at least a little stressed at work does not change.

Picturing the World

What is the cost of the U.S. Census? According to estimates, it has been escalating with each decade. The cost of the 1950 Census was approximately $91.5 million. The most recent U.S. Census, taken in 2020, was estimated to cost a staggering $15.6 billion. (Source: U.S. Census Bureau and U.S. Government Accountability Office)



What are some of the costs involved in taking a census?

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. In the United States, a survey of about 9400 individuals aged 15 and over found that such individuals spent an average of 5.19 hours per day engaged in leisure and sports activities. *(Source: U.S. Bureau of Labor Statistics)*
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 5.19 hours per day 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 more than 1000 adults aged 65–80 found that 47% who report listening to loud or very loud music in their youth now report being hard of hearing. *(Source: The Harris Poll)*

Answer: Page A35

In this course, you will see how the use of statistics can help you make informed decisions. Consider the census that the U.S. government takes every decade. 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.)

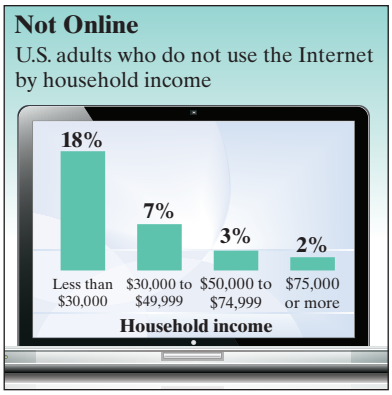
EXAMPLE 3

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 1502 U.S. adults found that 18% of adults from households earning less than $30,000 annually do not use the Internet, as shown in the figure at the left. *(Source: Pew Research Center)*

2. A study of 1000 U.S. 401(k) retirement plan participants found that the percentage who do not know how many years their retirement savings might last is 32%. *(Source: Charles Schwab & Co., Inc.)*



SOLUTION

1. The population consists of the responses of all U.S. adults, and the sample consists of the responses of the 1502 U.S. adults in the study. The part of this study that represents the descriptive branch of statistics involves the statement "18% of adults from households earning less than $30,000 annually do not use the Internet." Also, the figure represents the descriptive branch of statistics. A possible inference drawn from the study is that the Internet has been made inaccessible to lower-income households.

2. The population consists of the responses of all U.S. 401(k) retirement plan participants, and the sample consists of the responses of the 1000 U.S. 401(k) retirement plan participants in the study. The part of this study that represents the descriptive branch of statistics involves the statement "the percentage [of U.S. 401(k) retirement plan participants] who do not know how many years their retirement savings might last is 32%." A possible inference drawn from the study is that the amount of money a person needs for retirement is difficult to determine.

TRY IT YOURSELF 3

A study of 513 respondents to an Internet-wide survey found that 97% of the respondents said music is important to them, and 83% of the respondents said they actively look for new music. *(Source: Medium)*

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 A35

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.*

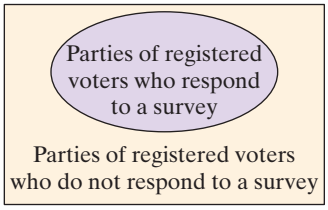
5. A statistic is a numerical description of a population characteristic.
6. A sample is a subset of a population.
7. It is impossible to obtain all the census data about the U.S. population.
8. Inferential statistics involves using a population to draw a conclusion about a corresponding sample.
9. A 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.*

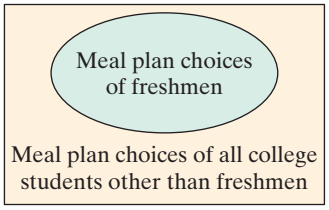
11. The salary of each employee of an advertising firm
12. The amount of energy collected from every solar panel on a photovoltaic power plant
13. A survey of 250 members from an organized union of over 20,000 members
14. The annual revenue of each team in a pro sports league
15. The carbon monoxide levels of 12 of 49 people who escaped a burning building
16. The number of electoral college votes for each state in the U.S. and the District of Columbia
17. The number of guests in each room of a hotel
18. The amount spent by every tenth person cashing out at a store
19. The nationality of every person passing through a customs station
20. The precipitation amounts at 15 locations in a county

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

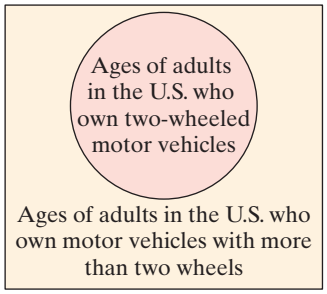
21. **Parties of Registered Voters**



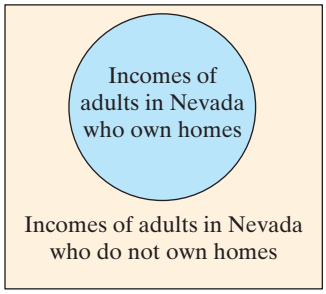
22. **Meal Plan Choices of College Students**



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



24. **Incomes of Adults in Nevada**



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 1021 U.S. adults found that 45% have a favorable view of Cuba. *(Source: Gallup)*

26. A study of 227 U.S. infants was conducted to explore norms of the gut microbiomes of healthy infants. *(Source: Scientific Reports)*

27. A survey of 1500 U.S. adults found that 59% have never had a vaccine reaction. *(Source: SingleCare)*

28. A survey of 1028 U.S. adults found that 7% of respondents have never heard of organ and tissue donation. *(Source: Research!America)*

29. A survey of 2111 U.S. small business owners found that 54% oppose increasing the minimum wage. *(Source: CNBC)*

30. A survey of 214 of the seniors graduating with a bachelor of science degree from a university found that 15% planned to obtain entry-level jobs in the health field.

31. A survey of 1001 U.S. adults found that 47% of respondents typically feel well rested on weekdays. *(Source: National Sleep Foundation)*

32. A survey of 366 automobile owners who purchased extended warranties found that 44% never used the warranty.

33. To gather information about starting salaries at companies listed in the Standard & Poor's 500, a researcher contacts 74 of the 500 companies.

34. In a survey of 679 members of a local children's museum about parenting attitudes, 575 of the participants were female and 423 of the participants were parents of two or more children. *(Source: University of California Press)*

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 24 of a hospital's 82 registered nurses is $71,000.

36. A survey of 919 college board members found that 89% think that their institution is a good place for members of racial and ethnic minorities. *(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 2021, 54% of the governors of the 50 states in the United States were Republicans. *(Source: National Governors Association)*

39. In a survey of automobile owners, 6% said they had to change their engine control module at least once.

40. Voter registration records show that 47% of all voters in a county are registered as Democrats.

41. A survey of 1000 U.S. adults found that 79% think that the spread of infectious diseases 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 20.2. *(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 Immunity A study showed the same level of T cell production in senior citizens who are amateur cyclists as in young adults, but a significantly lower level of T cell production in senior citizens who do not exercise regularly. Is it appropriate to infer that exercise stimulates T cell production? Explain. *(Source: University of Birmingham)*

48. Weight Loss and High Blood Pressure A study showed an association between intentional weight loss and a decreased risk of high blood pressure. Is it appropriate to infer from this study that weight loss causes a decreased risk of high blood pressure? Explain. *(Source: European Association for the Study of Obesity)*

49. Sleep and Student Achievement A study of college students showed that participants earned higher scores on quizzes and midterm exams with better sleep. *(Source: The American Journal of Managed Care)*

(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 a partial list of vulnerable, endangered, or critically endangered species and the approximate numbers of each species remaining. Which data are qualitative data and which are quantitative data? Explain your reasoning. *(Source: World Wildlife Fund)*

Vulnerable, Endangered, or Critically Endangered Species

Common species name	Number remaining
African elephant	415,000
Black-footed ferret	370
Giant panda	1864
Indus river dolphin	1816
Javan rhinoceros	60
North Atlantic right whale	400
Sunda tiger	400
Tapanuli orangutan	800
Vaquita	10

SOLUTION

The information shown in the table can be separated into two data sets. One data set contains the common species names and the other contains the numbers remaining. The names are nonnumerical entries, so these are qualitative data. The numbers remaining 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	593,490
Chicago, IL	2,693,976
Glendale, AZ	252,381
Denver, CO	727,211
Portland, OR	654,741
San Francisco, CA	881,549

Answer: Page A35

Levels of Measurement

Another characteristic of data is their 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 30 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: The Harris Poll)

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

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 2029)
1. Home health and personal care aides
2. Fast food and counter workers
3. Restaurant cooks
4. Software developers and software quality assurance analysts and testers
5. Registered nurses

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 A35

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. In contrast, 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. In contrast, 2°C is not twice as warm as 1°C, so these data are at the interval level.

For help with basic mathematical symbols and Greek letters and addition and subtraction of integers, see *Integrated Review* at

MyLab Statistics

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

2020 American League home run totals (by team)	
Baltimore	77
Boston	81
Chicago	96
Cleveland	59
Detroit	62
Houston	69
Kansas City	68
Los Angeles	85
Minnesota	91
New York	94
Oakland	71
Seattle	60
Tampa Bay	80
Texas	62
Toronto	88

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 22 more home runs than Cleveland hit because $81 - 59 = 22$ home runs. Also, Chicago hit about 1.25 times as many home runs as Baltimore hit because

$$\frac{96}{77} \approx 1.25.$$

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 A35

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 30.9, Jul 73.6 Feb 32.8, Aug 71.5 Mar 40.0, Sep 62.4 Apr 47.5, Oct 50.3 May 57.2, Nov 38.6 Jun 67.0, Dec 30.0 *(Source: National Oceanic and Atmospheric Administration)*	Put in a category, put in order, and *find differences between data entries.* For instance, $71.5 - 62.4 = 9.1°F$. 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 Oceanic and Atmospheric Administration)*	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. Nationalities of passengers on a plane
8. Zip codes
9. Ages of dogs at a rescue facility
10. Capacities of commercial freezers
11. Types of flowers
12. Names of towns where branch campuses of a college are located
13. Distances of track events
14. Response times for a customer service representative

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 colleges in terms of value for the money according to *U.S. News & World Report* are listed. *(Source: U.S. News & World Report)*

1. Harvard
2. Princeton
3. Gallaudet
4. Yale
5. MIT
6. Columbia
7. Stanford
8. Rice
9. Vanderbilt
10. Dartmouth

17. Automobiles The lengths (in centimeters) of 22 Ford automobiles are listed. *(Source: Automobiledimension.com)*

404	407	410	416	421	438	440	443
461	467	471	478	480	483	483	485
487	487	497	505	534	536		

18. Classrooms The room numbers of the classrooms in a college science building are listed.

112	113	114	116	117	118	122
212	213	214	215	216	217	219

19. Best Sellers List The top ten fiction hardcover books on *The New York Times* Best Sellers List based on sales in the week ending March 6, 2021, are listed. *(Source: The New York Times)*

1. *Life After Death*
2. *The Four Winds*
3. *Klara and the Sun*
4. *Dark Sky*
5. *The Affair*
6. *The Midnight Library*
7. *The Lost Apothecary*
8. *The Vanishing Half*
9. *Infinite Country*
10. *A Court of Silver Flames*

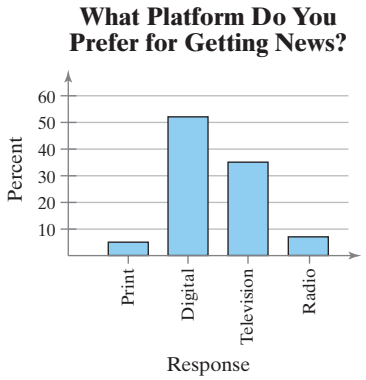
20. Bell Schedule The times from a high school bell schedule are listed.

8:00 A.M.	8:52 A.M.	8:56 A.M.	9:48 A.M.	9:52 A.M.
10:44 A.M.	10:48 A.M.	11:40 A.M.	11:44 A.M.	1:08 P.M.
1:12 P.M.	2:04 P.M.	2:08 P.M.	3:00 P.M.	

Graphical Analysis

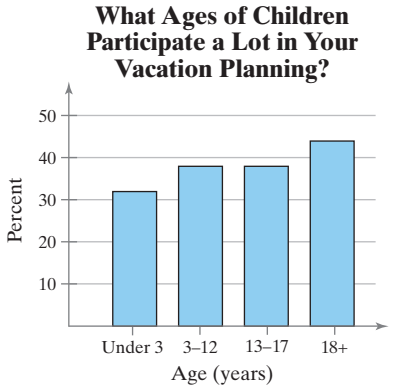
In Exercises 21–24, determine the level of measurement of the data listed on the horizontal and vertical axes in the figure.

21.



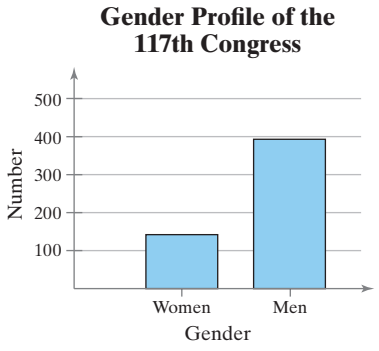
(Source: Pew Research Center)

22.



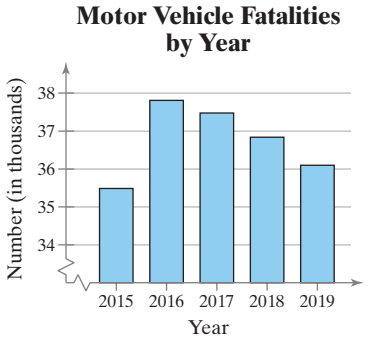
(Source: Marriott Bonvoy Bold from Chase)

23.



(Source: Center for American Women and Politics)

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 2021 are listed. *(Source: Associated Press)*

1. Alabama
2. Ohio State
3. Clemson
4. Texas A&M
5. Notre Dame
6. Oklahoma
7. Georgia
8. Cincinnati
9. Iowa State
10. Northwestern

28. Coffee Four principal types of coffee beans are listed.

Arabica Robusta Liberica Excelsa

29. Census Regions The four geographical regions of the United States recognized by the U.S. Census Bureau are listed.

Northeast South
Midwest West

30. Figure Skating The top six final scores at the 2021 U.S. Women's Figure Skating Championships are listed. *(Source: NBC Sports)*

232.61 215.33 214.98
213.39 199.95 178.89

31. Richest People The ten richest people in the world as of March 30, 2021, are listed. *(Source: Bloomberg Reporting)*

1. Jeff Bezos
2. Elon Musk
3. Bill Gates
4. Bernard Arnault
5. Mark Zuckerberg
6. Warren Buffett
7. Larry Page
8. Sergey Brin
9. Steve Ballmer
10. Larry Ellison

32. Numbers of Performances The numbers of performances for the 10 longest-running original runs of Broadway shows as of March 15, 2020, are listed. *(Source: Playbill)*

13,370 9692 9302 7485 6836
6680 6137 5959 5758 5461

Extending Concepts

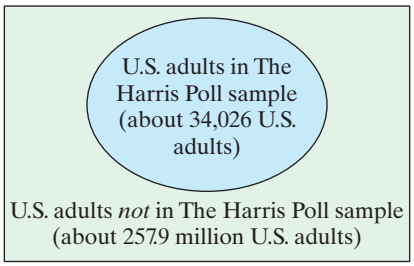
33. Writing What is an inherent zero? Describe three examples of data sets that have inherent zeros and three that do not.

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 used a sample of 34,026 U.S. adults for the 2020 survey. The survey respondents rate companies according to key attributes that are classified into seven categories: (1) trust, (2) vision, (3) growth, (4) products and services, (5) culture, (6) ethics, and (7) citizenship. This information is used to determine the reputation of a company as Excellent, Very Good, Good, Fair, Poor, or Very Poor. 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	798,000
Netflix, Inc.	1999	Very Good	Digital television	8,600
Apple, Inc.	1977	Very Good	Computers and peripherals	147,000
The Kraft Heinz Co.	2015	Very Good	Food products	37,000
Chipotle Mexican Grill, Inc.	1993	Good	Restaurant	83,000
Exxon Mobil Corp.	1999	Good	Petroleum (integrated)	71,000
The Boeing Co.	1916	Fair	Aircraft	161,000
Comcast Corp.	1963	Poor	Cable television	168,000
Wells Fargo & Co.	1998	Poor	Banking	263,000
Facebook, Inc.	2004	Poor	Internet	45,000

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

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 257.9 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 volunteering, 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 rats in a treatment group were given trimethylamine oxide, a substance present in seafood, while rats in a control group were given water. After performing testing, researchers concluded that trimethylamine oxide reduced mortality related to heart disease in rats that had heart disease. *(Source: eLife)*

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 who were newly diagnosed with a viral infection. To perform the study, researchers give 2700 U.S. adults either a daily vitamin D_3 supplement or a placebo for four weeks. *(Source: U.S. National Library of Medicine)*
2. Researchers conduct a study to determine how confident Americans are in the U.S. economy. To perform the study, researchers call 1019 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 determine the percentage of the Pennsylvania elk population in each age and sex class. The commission captured and released elk during each year of the study and found an overall average of 16% branched bulls, 7% spike bulls, 56% adult cows, and 21% calves. Is this study an observational study or an experiment? *(Source: Pennsylvania Game Commission)*

Answer: Page A35

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 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 their 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.

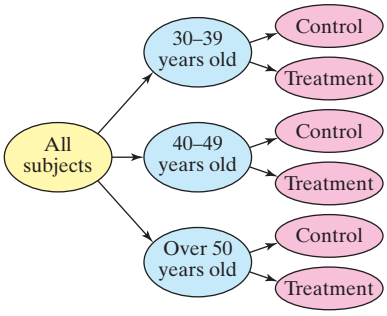
DEFINITION

Blinding is a technique in which 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 are 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,** in which 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 1000 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 A35

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.

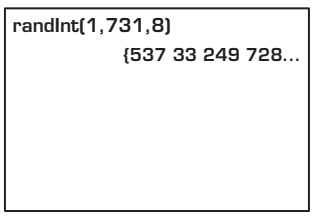
MATH

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 such as 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

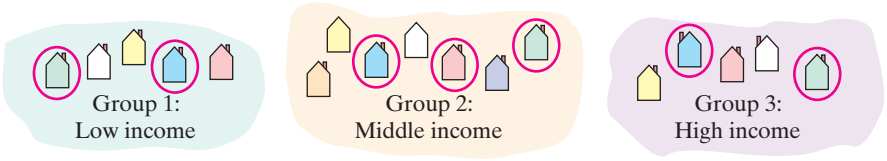
A company employs 79 people. Choose a simple random sample of five to survey.

Answer: Page A35

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 categories and then randomly select households from each category. 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.



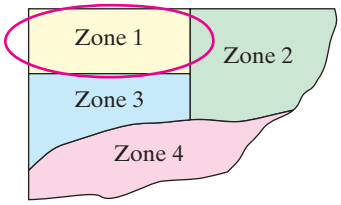
Stratified Sampling

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.

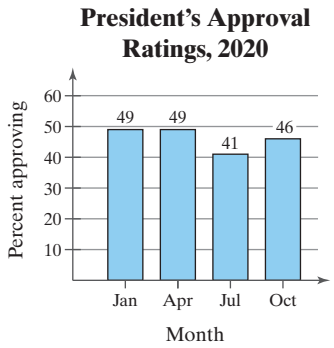
Zip Code Zones in West Ridge County



Cluster Sampling

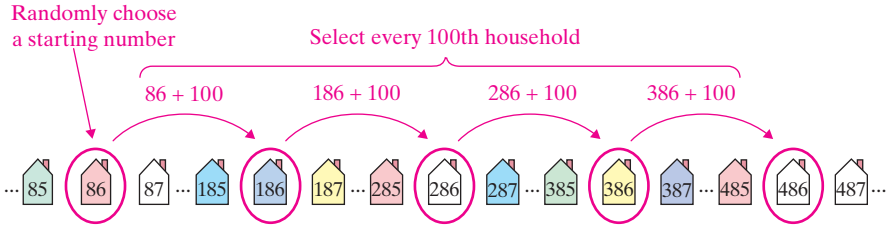
Picturing the World

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 Donald Trump for selected months in 2020 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 access.

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 who 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 A35

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. A research study compared the memory retention of subjects when a learning activity was followed by a brief period of wakeful rest and when a learning activity was followed by a brief period of distraction. *(Source: Springer Nature)*
12. In a survey of U.S. employees, 42% of black women say they are uncomfortable sharing thoughts about racial inequality. *(Source: McKinsey & Company)*
13. A study used periodic blood pressure readings and brain MRIs of adults to find that adults with long periods of high blood pressure were more likely to develop cerebral small blood vessel disease. *(Source: American Heart Association)*
14. To study the effects of music on body image, researchers played a song with a body-positive message or a song with a body-negative message to different groups of women. *(Source: Psychology of Popular Media)*
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. Acne Treatment A company wants to test the effectiveness of a new acne cream. The company recruits 500 girls ages 13 to 17 who have acne. The subjects are randomly assigned into two groups. One group is given the acne cream and the other is given a placebo that looks exactly like the acne cream. Both groups apply the cream daily for two months. Facial photos are taken at the beginning and end of the treatment to compare results.

(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. Social Anxiety Disorder Researchers in Japan tested the effect of cannabidiol (CBD) in treating late teenagers with social anxiety disorder (SAD). Thirty-seven 18- to 19-year-old teenagers with SAD took part in the study. The patients were assigned at random to receive a daily dose of either CBD or a placebo for 4 weeks. Symptoms were measured using The Fear of Negative Evaluation Questionnaire and the Liebowitz Social Anxiety Scale at the beginning and end of the treatment. *(Source: Frontiers in Psychology)*

(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, Arya, Xavier, Nyla, Shaniece, Chen, Juan, Hana, Nia, Ansel, Liam, Bruno, Mei, Zoey, Kayla, Liam, Sofia, and Kai. 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 students at random from a campus directory, researchers contact 300 people and ask what obstacles (partying is mentioned as an example) keep them from completing their homework.

24. Questioning university students as they leave a college cafeteria, a researcher asks 342 students about their eating 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 using a mall entrance is asked to name their 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. By making calls to 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 GPA of the 85 students on a college football team roster

30. The average distance traveled to a stadium by 55,000 spectators

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 a subject that interests 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, such as a "yes" or "no" response, or a multiple choice response. An open question and a closed question are given below. List an advantage and a disadvantage of each type of question.

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

Closed Question Would a mandatory nutrition course be an effective way to get students to eat healthier foods?

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