Introductory STATISTICS

Prem S. Mann

9TH EDITION

WileyPLUS

Now with: ORION, An Adaptive Experience

WileyPLUS is a research-based online environment for effective teaching and learning.

WileyPLUS builds students' confidence because it takes the guesswork out of studying by providing students with a clear roadmap:

- what to do
- how to do it
- if they did it right

It offers interactive resources along with a complete digital textbook that help students learn more. With *WileyPLUS*, students take more initiative so you'll have greater impact on their achievement in the classroom and beyond.

For more information, visit www.wileyplus.com

WileyPLUS

ALL THE HELP, RESOURCES, AND PERSONAL SUPPORT YOU AND YOUR STUDENTS NEED!

www.wileyplus.com/resources

1st DAY OF CLASS ...AND BEYOND!

2-Minute Tutorials and all of the resources you and your students need to get started

WileyPLUS

Student Partner Program

Student support from an experienced student user

Wiley Faculty Network

Collaborate with your colleagues, find a mentor, attend virtual and live events, and view resources
www.WhereFacultyConnect.com

WileyPLUS

Quick Start

Pre-loaded, ready-to-use assignments and presentations created by subject matter experts

Technical Support 24/7 FAQs, online chat, and phone support
www.wileyplus.com/support

© Courtney Keating/iStockphoto

Your *WileyPLUS* Account Manager, providing personal training and support

Ninth Edition

INTRODUCTORY STATISTICS

PREM S. MANN

EASTERN CONNECTICUT STATE UNIVERSITY

WILEY

To the memory of my parents

PREFACE

Introductory Statistics is written for a one- or two-semester first course in applied statistics. This book is intended for students who do not have a strong background in mathematics. The only prerequisite for this text is knowledge of elementary algebra.

Today, college students from almost all fields of study are required to take at least one course in statistics. Consequently, the study of statistical methods has taken on a prominent role in the education of students from a variety of backgrounds and academic pursuits. From the first edition, the goal of **Introductory Statistics** has been to make the subject of statistics interesting and accessible to a wide and varied audience. Three major elements of this text support this goal:

1. Realistic content of its examples and exercises, drawing from a comprehensive range of applications from all facets of life
2. Clarity and brevity of presentation
3. Soundness of pedagogical approach

These elements are developed through the interplay of a variety of significant text features. The feedback received from the users of the eighth edition (and earlier editions) of **Introductory Statistics** has been very supportive and encouraging. Positive experiences reported by instructors and students have served as evidence that this text offers an interesting and accessible approach to statistics—the author's goal from the very first edition. The author has pursued the same goal through the refinements and updates in this ninth edition, so that **Introductory Statistics** can continue to provide a successful experience in statistics to a growing number of students and instructors.

New to the Ninth Edition

The following are some of the changes made in the ninth edition:

- New for the ninth edition, are the videos that are accessible via the *WileyPLUS* course associated with this text. These videos provide step-by-step solutions to selected examples in the book.
- A large number of the examples and exercises are new or revised, providing contemporary and varied ways for students to practice statistical concepts.
- Coverage of sample surveys, sampling techniques, and design of experiments has been moved from Appendix A to Chapter 1.
- In Chapter 3, the discussions of weighted mean, trimmed mean, and coefficient of variation have been moved from the exercises to the main part of the chapter.
- The majority of the case studies are new or revised, drawing on current uses of statistics in areas of student interest.
- New data are integrated throughout, reinforcing the vibrancy of statistics and the relevance of statistics to student lives right now.
- The *Technology Instructions* sections have been updated to support the use of the latest versions of TI-84 Color/TI-84, Minitab, and Excel.
- Many of the *Technology Assignments* at the end of each chapter are either new or have been updated.

- The data sets posted on the book companion Web site and *WileyPLUS* have been updated.
- Most of the *Uses and Misuses* sections at the end of each chapter have been updated or replaced.
- Many of the *Mini-Projects*, which are now located on the book companion Web site, are either new or have been updated.
- Many of the *Decide for Yourself* sections, also located on the book companion Web site, are either new or have been updated.

Hallmark Features of This Text

Clear and Concise Exposition The explanation of statistical methods and concepts is clear and concise. Moreover, the style is user-friendly and easy to understand. In chapter introductions and in transitions from section to section, new ideas are related to those discussed earlier.

Thorough Examples The text contains a wealth of examples. The examples are usually presented in a format showing a problem and its solution. They are well sequenced and thorough, displaying all facets of concepts. Furthermore, the examples capture students' interest because they cover a wide variety of relevant topics. They are based on situations that practicing statisticians encounter every day. Finally, a large number of examples are based on real data taken from sources such as books, government and private data sources and reports, magazines, newspapers, and professional journals.

Step-by-Step Solutions A clear, concise solution follows each problem presented in an example. When the solution to an example involves many steps, it is presented in a step-by-step format. For instance, examples related to tests of hypothesis contain five steps that are consistently used to solve such examples in all chapters. Thus, procedures are presented in the concrete settings of applications rather than as isolated abstractions. Frequently, solutions contain highlighted remarks that recall and reinforce ideas critical to the solution of the problem. Such remarks add to the clarity of presentation.

Titles for Examples Each example based on an application of concepts now contains a title that describes to what area, field, or concept the example relates.

Margin Notes for Examples A margin note appears beside each example that briefly describes what is being done in that example. Students can use these margin notes to assist them as they read through sections and to quickly locate appropriate model problems as they work through exercises.

Frequent Use of Diagrams Concepts can often be made more understandable by describing them visually with the help of diagrams. This text uses diagrams frequently to help students understand concepts and solve problems. For example, tree diagrams are used a few times in Chapters 4 and 5 to assist in explaining probability concepts and in computing probabilities. Similarly, solutions to all examples about tests of hypothesis contain diagrams showing rejection regions, nonrejection regions, and critical values.

Highlighting Definitions of important terms, formulas, and key concepts are enclosed in colored boxes so that students can easily locate them.

Cautions Certain items need special attention. These may deal with potential trouble spots that commonly cause errors, or they may deal with ideas that students often overlook. Special emphasis is placed on such items through the headings *Remember*, *An Observation*, or *Warning*. An icon is used to identify such items.

Real World Case Studies These case studies, which appear in most of the chapters, provide additional illustrations of the applications of statistics in research and statistical analysis. Most of these case studies are based on articles or data published in journals, magazines, newspapers, or Web sites. Almost all case studies are based on real data.

Variety of Exercises The text contains a variety of exercises, including technology assignments. Moreover, a large number of these exercises contain several parts. Exercise sets appearing at the

end of each section (or sometimes at the end of two or three sections) include problems on the topics of that section. These exercises are divided into two parts: **Concepts and Procedures** that emphasize key ideas and techniques and **Applications** that use these ideas and techniques in concrete settings. Supplementary exercises appear at the end of each chapter and contain exercises on all sections and topics discussed in that chapter. A large number of these exercises are based on real data taken from varied data sources such as books, government and private data sources and reports, magazines, newspapers, and professional journals. Not only do the exercises given in the text provide practice for students, but the real data contained in the exercises provide interesting information and insight into economic, political, social, psychological, and other aspects of life. The exercise sets also contain many problems that demand critical thinking skills. The answers to selected odd-numbered exercises appear in the *Answers* section at the back of the book. **Optional exercises** are indicated by an asterisk (*).

Advanced Exercises All chapters have a set of exercises that are of greater difficulty. Such exercises appear under the heading *Advanced Exercises* after the *Supplementary Exercises*.

Uses and Misuses This feature toward the end of each chapter (before the Glossary) points out common misconceptions and pitfalls students will encounter in their study of statistics and in everyday life. Subjects highlighted include such diverse topics as *do not feed the animals*.

Decide for Yourself This feature is accessible online at **www.wiley.com/college/mann**. Each Decide for Yourself discusses a real-world problem and raises questions that readers can think about and answer.

Glossary Each chapter has a glossary that lists the key terms introduced in that chapter, along with a brief explanation of each term.

Self-Review Tests Each chapter contains a *Self-Review Test*, which appears immediately after the *Supplementary* and *Advanced Exercises*. These problems can help students test their grasp of the concepts and skills presented in respective chapters and monitor their understanding of statistical methods. The problems marked by an asterisk (*) in the *Self-Review Tests* are **optional**. The answers to almost all problems of the *Self-Review Tests* appear in the *Answer* section.

Technology Usage At the end of each chapter is a section covering uses of three major technologies of statistics and probability: the TI-84 Color/TI-84, Minitab, and Excel. For each technology, students are guided through performing statistical analyses in a step-by-step fashion, showing them how to enter, revise, format, and save data in a spreadsheet, workbook, or named and unnamed lists, depending on the technology used. Illustrations and screen shots demonstrate the use of these technologies. Additional detailed technology instruction is provided in the technology manuals that are online at www.wiley.com/college/mann.

Technology Assignments Each chapter contains a few technology assignments that appear at the end of the chapter. These assignments can be completed using any of the statistical software.

Mini-projects Associated with each chapter of the text are Mini-projects posted online at **www.wiley.com/college/mann**. These Mini-projects are either very comprehensive exercises or they ask students to perform their own surveys and experiments. They provide practical applications of statistical concepts to real life.

Data Sets A large number of data sets appear on the book companion Web site at **www.wiley.com/college/mann**. These large data sets are collected from various sources, and they contain information on several variables. Many exercises and assignments in the text are based on these data sets. These large data sets can also be used for instructor-driven analyses using a wide variety of statistical software packages as well as the TI-84. **These data sets are available on the Web site of the text in numerous formats, including Minitab and Excel.**

Videos New for the Ninth Edition, videos for each text section illustrate concepts related to the topic covered in that section to more deeply engage the students. These videos are accessible via *WileyPLUS*.

GAISE Report Recommendations Adopted

In 2003, the American Statistical Association (ASA) funded the Guidelines for Assessment and Instruction in Statistics Education (GAISE) Project to develop ASA-endorsed guidelines for assessment and instruction in statistics for the introductory college statistics course. The report, which can be found at www.amstat.org/education/gaise, resulted in the following series of recommendations for the first course in statistics and data analysis.

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

Here are a few examples of how this Introductory Statistics text can assist in helping you, the instructor, in meeting the GAISE recommendations.

1. Many of the exercises require interpretation, not just answers in terms of numbers. Graphical and numeric summaries are combined in some exercises in order to emphasize looking at the whole picture, as opposed to using just one graph or one summary statistic.
2. The *Uses and Misuses* and online *Decide for Yourself* features help to develop statistical thinking and conceptual understanding.
3. All of the data sets listed in Appendix A are available on the book's Web site. They have been formatted for a variety of statistical software packages. This eliminates the need to enter data into the software. A variety of software instruction manuals also allow the instructor to spend more time on concepts and less time teaching how to use technology.
4. The *online Mini-projects* help students to generate their own data by performing an experiment and/or taking random samples from the large data sets mentioned in Appendix A.

We highly recommend that all statistics instructors take the time to read the GAISE report. There is a wealth of information in this report that can be used by everyone.

Web Site

www.wiley.com/college/mann

After you go to the page exhibited by the above URL, click on *Visit the Companion Sites*. Then click on the site that applies to you out of the two choices. This Web site provides additional resources for instructors and students. The following items are available for instructors on this Web site:

- Key Formulas
- Printed Test Bank
- Mini-Projects
- Decide for Yourself
- Power Point Lecture Slides
- Instructor's Solutions Manual
- Data Sets (see Appendix A for a complete list of these data sets)
- Chapter 14: Multiple Regression
- Chapter 15: Nonparametric Methods
- Technology Resource Manuals:
  - TI Graphing Calculator Manual
  - Minitab Manual
  - Excel Manual

These manuals provide step-by-step instructions, screen captures, and examples for using technology in the introductory statistics course. Also provided are exercise lists and indications of which exercises from the text best lend themselves to the use of the package presented.

Using *WileyPLUS*

SUCCESS: *WileyPLUS* helps to ensure that each study session has a positive outcome by putting students in control. Through instant feedback and study objective reports, students know **if they did it right** and where to focus next, so they achieve the strongest results.

Our efficacy research shows that with *WileyPLUS,* students improve their outcomes by as much as one letter grade. *WileyPLUS* helps students take more initiative, so you will have greater impact on their achievement in the classroom and beyond.

What Do Students Receive with *WileyPLUS*?

- The complete digital textbook, saving students up to 60% off the cost of a printed text.
- Question assistance, including links to relevant sections in the online digital textbook.
- Immediate feedback and proof of progress, 24/7.
- Integrated, multimedia resources—including videos—that provide multiple study paths and encourage more active learning.

What Do Instructors Receive with *WileyPLUS*?

- Reliable resources that reinforce course goals inside and outside of the classroom.
- The ability to easily identify those students who are falling behind.
- Media-rich course materials and assessment content, including Instructor's Solutions Manual, PowerPoint slides, Learning Objectives, Printed Test Bank, and much more.

  www.wileyplus.com. Learn More.

Supplements

The following supplements are available to accompany this text:

■ **Instructor's Solutions Manual (ISBN 978-1-119-14830-2).** This manual contains complete solutions to all of the exercises in the text.

■ **Printed Test Bank** The printed copy of the test bank contains a large number of multiple-choice questions, essay questions, and quantitative problems for each chapter. It can be downloaded and printed from *WileyPLUS* or from www.wiley.com/college/mann.

■ **Student Solutions Manual (ISBN 978-1-119-14829-6).** This manual contains complete solutions to all of the odd-numbered exercises in the text.

Acknowledgments

I thank the following reviewers of the ninth edition of this book, whose comments and suggestions were invaluable in improving the text.

D. P. Adhikari
Marywood University
Wendy Ahrendsen
South Dakota State University
Nan Hutchins Bailey
The University of Texas at Tyler
Les Barnhouse
Athabasca University

Hossein Behforooz
Utica College
Joleen Beltrami
University of the Incarnate Word
Bill Burgin
Gaston College
David Bush
Villanova University

Ferry Butar Butar
Sam Houston State University
Chris Chappa
University of Texas at Tyler
Jerry Chen
Suffolk County Community College,
   Ammerman Campus
A. Choudhury
Illinois State University
Robert C. Forsythe
Frostburg State University
Daesung Ha
Marshall University
Rhonda Hatcher
Texas Christian University
Joanna Jeneralczuk
University of Massachusetts, Amherst
Annette Kanko
University of the Incarnate Word
Mohammad Kazemi
University of North Carolina,
   Charlotte
Janine Keown-Gerrard
Athabasca University
Hoon Kim
California State Polytechnic
   University—Pomona
Matthew Knowlen
Horry Georgetown Technical College

Maurice LeBlanc
Kennesaw State University
Min-Lin Lo
California State University, San Bernardino
Natalya Malakhova
Johnson County Community College
Nola McDaniel
McNeese State University
Stéphane Mechoulan
Dalhousie University
Robert L. Nichols
Florida Gulf Coast University
Jonathan Oaks
Macomb Community
   College—South Campus
Marty Rhoades
West Texas A&M University
Mary Beth Rollick
Kent State University—Kent
Seema Sehgal
Athabasca University
Linda Simonsen
University of Washington, Bothell
Criselda Toto
Chapman University
Viola Vajdova
Benedictine University
Daniel Weiner
Boston University

I express my thanks to the following for their contributions to earlier editions of this book that made it better in many ways: Chris Lacke (Rowan University), Gerald Geissert (formerly of Eastern Connecticut State University), Daniel S. Miller (Central Connecticut State University), and David Santana-Ortiz (Rand Organization).

I extend my special thanks to Doug Tyson, James Bush, and Chad Cross, who contributed to this edition in many significant ways. I take this opportunity to thank Mark McKibben for working on the solutions manuals and preparing the answer section and Julie M. Clark for checking the solutions for accuracy.

It is of utmost importance that a textbook be accompanied by complete and accurate supplements. I take pride in mentioning that the supplements prepared for this text possess these qualities and much more. I thank the authors of all these supplements.

It is my pleasure to thank all the professionals at John Wiley & Sons with whom I enjoyed working during this revision. Among them are Laurie Rosatone (Vice President and Director), Joanna Dingle (Acquisitions Editor), Wendy Lai (Senior Designer), Billy Ray (Senior Photo Editor), Valerie Zaborski (Senior Content Manager), Ken Santor (Senior Production Editor), Ellen Keohane (Project Editor), Jennifer Brady (Sponsoring Editor), Carolyn DeDeo (Market Solutions Assistant), David Dietz (Senior Product Designer), and John LaVacca (Marketing Manager). I also want to thank Jackie Henry (Full Service Manager) for managing the production process and Lisa Torri (Art Development Editor) for her work on the case study art.

Any suggestions from readers for future revisions would be greatly appreciated. Such suggestions can be sent to the author at **premmann@yahoo.com** or **mann@easternct.edu**. I recently retired from Eastern Connecticut State University and, hence, will prefer that readers email me at **premmann@yahoo.com**.

Prem S. Mann
December 2015

CONTENTS

Introduction

Are you, as an American, thriving in your life? Or are you struggling? Or, even worse, are you suffering? A poll of 176,903 American adults, aged 18 and older, was conducted January 2 to December 30, 2014, as part of the Gallup-Healthways Well-Being Index survey. The poll found that while 54.1% of these Americans said that they were thriving, 42.1% indicated that they were struggling, and 3.8% mentioned that they were suffering. (See Case Study 1–2.)

The study of statistics has become more popular than ever over the past four decades. The increasing availability of computers and statistical software packages has enlarged the role of statistics as a tool for empirical research. As a result, statistics is used for research in almost all professions, from medicine to sports. Today, college students in almost all disciplines are required to take at least one statistics course. Almost all newspapers and magazines these days contain graphs and stories on statistical studies. After you finish reading this book, it should be much easier to understand these graphs and stories.

Every field of study has its own terminology. Statistics is no exception. This introductory chapter explains the basic terms and concepts of statistics. These terms and concepts will bridge our understanding of the concepts and techniques presented in subsequent chapters.

1.1   Statistics and Types of Statistics

In this section we will learn about statistics and types of statistics.

1.1.1   What Is Statistics?

The word **statistics** has two meanings. In the more common usage, *statistics* refers to numerical facts. The numbers that represent the income of a family, the age of a student, the percentage of passes completed by the quarterback of a football team, and the starting salary of a typical college graduate are examples of statistics in this sense of the word. A 1988 article in *U.S. News & World Report* mentioned that "Statistics are an American obsession."[1] During the 1988 baseball World Series between the Los Angeles Dodgers and the Oakland A's, the then NBC commentator Joe Garagiola reported to the viewers numerical facts about the players' performances. In response, fellow commentator Vin Scully said, "I love it when you talk statistics." In these examples, the word *statistics* refers to numbers.

The following examples present some statistics:

1. During March 2014, a total of 664,000,000 hours were spent by Americans watching March Madness live on TV and/or streaming (*Fortune* Magazine, March 15, 2015).
2. Approximately 30% of Google's employees were female in July 2014 (*USA TODAY*, July 24, 2014).
3. According to an estimate, an average family of four living in the United States needs $130,357 a year to live the American dream (*USA TODAY*, July 7, 2014).
4. Chicago's O'Hare Airport was the busiest airport in 2014, with a total of 881,933 flight arrivals and departures.
5. In 2013, author James Patterson earned $90 million from the sale of his books (*Forbes*, September 29, 2014).
6. About 22.8% of U.S. adults do not have a religious affiliation (*Time*, May 25, 2015).
7. Yahoo CEO Marissa Mayer was the highest paid female CEO in America in 2014, with a total compensation of $42.1 million.

The second meaning of *statistics* refers to the field or discipline of study. In this sense of the word, *statistics* is defined as follows.

> **Statistics**   **Statistics** is the science of collecting, analyzing, presenting, and interpreting data, as well as of making decisions based on such analyses.

Every day we make decisions that may be personal, business related, or of some other kind. Usually these decisions are made under conditions of uncertainty. Many times, the situations or problems we face in the real world have no precise or definite solution. Statistical methods help us make scientific and intelligent decisions in such situations. Decisions made by using statistical methods are called *educated guesses*. Decisions made without using statistical (or scientific) methods are *pure guesses* and, hence, may prove to be unreliable. For example, opening a large store in an area with or without assessing the need for it may affect its success.

Like almost all fields of study, statistics has two aspects: theoretical and applied. *Theoretical* or *mathematical statistics* deals with the development, derivation, and proof of statistical theorems, formulas, rules, and laws. *Applied statistics* involves the applications of those theorems, formulas, rules, and laws to solve real-world problems. This text is concerned with applied statistics and not with theoretical statistics. By the time you finish studying this book, you will have learned how to think statistically and how to make educated guesses.

1.1.2   Types of Statistics

Broadly speaking, applied statistics can be divided into two areas: **descriptive statistics** and **inferential statistics**.

[1]"The Numbers Racket: How Polls and Statistics Lie," *U.S. News & World Report*, July 11, 1988, pp. 44–47.
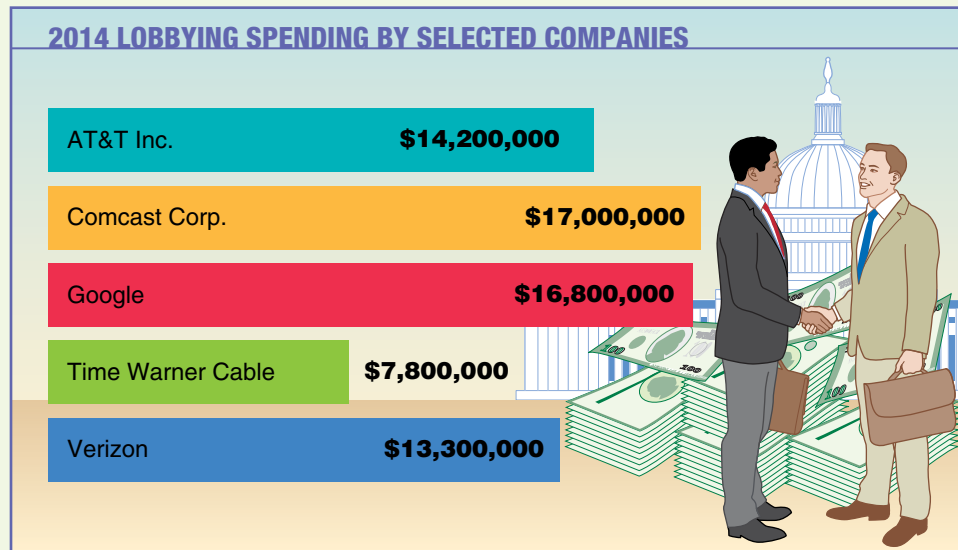
2014 LOBBYING SPENDING BY SELECTED COMPANIES

Company	Amount
AT&T Inc.	$14,200,000
Comcast Corp.	$17,000,000
Google	$16,800,000
Time Warner Cable	$7,800,000
Verizon	$13,300,000

Data source: *Fortune* Magazine, June 1, 2015

The accompanying chart shows the lobbying spending by five selected companies during 2014. Many companies spend millions of dollars to win favors in Washington. According to *Fortune* Magazine of June 1, 2015, "Comcast has remained one of the biggest corporate lobbyists in the country." In 2014, Comcast spent $17 million, Google spent $16.8 million, AT&T spent $14.2 million, Verizon spent $13.3 million, and Time Warner Cable spent $7.8 million on lobbying. These numbers simply describe the total amounts spent by these companies on lobbying. We are not drawing any inferences, decisions, or predictions from these data. Hence, this data set and its presentation is an example of descriptive statistics.

Descriptive Statistics

Suppose we have information on the test scores of students enrolled in a statistics class. In statistical terminology, the whole set of numbers that represents the scores of students is called a **data set**, the name of each student is called an **element**, and the score of each student is called an **observation**. (These terms are defined in more detail in Section 1.2.)

Many data sets in their original forms are usually very large, especially those collected by federal and state agencies. Consequently, such data sets are not very helpful in drawing conclusions or making decisions. It is easier to draw conclusions from summary tables and diagrams than from the original version of a data set. So, we summarize data by constructing tables, drawing graphs, or calculating summary measures such as averages. The portion of statistics that helps us do this type of statistical analysis is called **descriptive statistics**.

Descriptive Statistics **Descriptive statistics** consists of methods for organizing, displaying, and describing data by using tables, graphs, and summary measures.

Chapters 2 and 3 discuss descriptive statistical methods. In Chapter 2, we learn how to construct tables and how to graph data. In Chapter 3, we learn how to calculate numerical summary measures, such as averages.

Case Study 1–1 presents an example of descriptive statistics.

Inferential Statistics

In statistics, the collection of all elements of interest is called a **population**. The selection of a portion of the elements from this population is called a **sample**. (Population and sample are discussed in more detail in Section 1.5.)

AMERICANS' LIFE OUTLOOK, 2014

AMERICAN'S LIFE OUTLOOK 2014

Suffering **3.8%**

Struggling **42.1%**

Thriving **54.1%**

Margin of sampling error ±1%
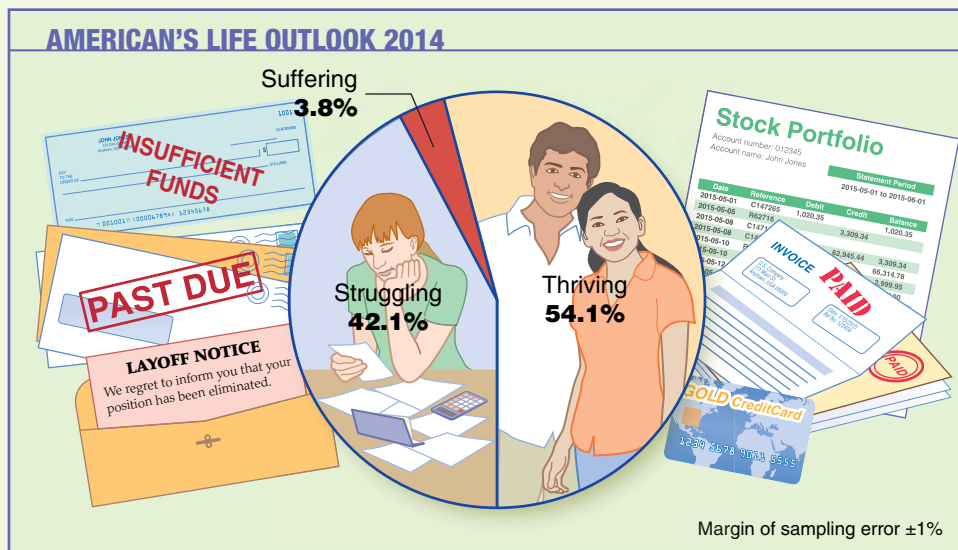
Data source: Gallup-Healthways Well-Being Index

A poll of 176,903 American adults, aged 18 and older, was conducted January 2 to December 30, 2014, as part of the Gallup-Healthways Well-Being Index survey. Gallup and Healthways have been "tracking Americans' life evaluations daily" since 2008. According to this poll, in 2014, Americans' outlook on life was the best in seven years, as 54.1% "rated their lives highly enough to be considered thriving," 42.1% said they were struggling, and 3.8% mentioned that they were suffering. As mentioned in the chart, the margin of sampling error was ±1%. In Chapter 8, we will discuss the concept of margin of error, which can be combined with these percentages when making inferences. As we notice, the results described in the chart are obtained from a poll of 176,903 adults. We will learn in later chapters how to apply these results to the entire population of adults. Such decision making about the population based on sample results is called inferential statistics.

A major portion of statistics deals with making decisions, inferences, predictions, and forecasts about populations based on results obtained from samples. For example, we may make some decisions about the political views of all college and university students based on the political views of 1000 students selected from a few colleges and universities. As another example, we may want to find the starting salary of a typical college graduate. To do so, we may select 2000 recent college graduates, find their starting salaries, and make a decision based on this information. The area of statistics that deals with such decision-making procedures is referred to as **inferential statistics**. This branch of statistics is also called *inductive reasoning* or *inductive statistics*.

Inferential Statistics **Inferential statistics** consists of methods that use sample results to help make decisions or predictions about a population.

Case Study 1–2 presents an example of inferential statistics. It shows the results of a survey in which American adults were asked about their opinions about their lives.

Chapters 8 through 15 and parts of Chapter 7 deal with inferential statistics.

Probability, which gives a measurement of the likelihood that a certain outcome will occur, acts as a link between descriptive and inferential statistics. Probability is used to make statements about the occurrence or nonoccurrence of an event under uncertain conditions. Probability and probability distributions are discussed in Chapters 4 through 6 and parts of Chapter 7.

EXERCISES

CONCEPTS AND PROCEDURES

1.1   Briefly describe the two meanings of the word *statistics*.

1.2   Briefly explain the types of statistics.

APPLICATIONS

1.3   Which of the following is an example of descriptive statistics and which is an example of inferential statistics? Explain.

    a. In a survey by *Fortune* Magazine and SurveyMonkey, participants were asked what was the most important factor when purchasing groceries (*Fortune*, June 1, 2015). The following table lists the summary of the responses of these participants. Assume that the maximum margin of error is ±1.5%.

Factor	Percent of Respondents
Price	42.4
Nutrition	36.0
Absence of additives	16.4
Number of calories	3.8
Carbon footprint	1.5

b. The following table gives the earnings of the world's top seven female professional athletes for the year 2014 (ceoworld.biz).

Female Professional Athlete	2014 Earnings (millions of dollars)
Maria Sharapova	24.4
Li Na	23.6
Serena Williams	22.0
Kim Yuna	16.3
Danica Patrick	15.0
Victoria Azarenka	11.1
Caroline Wozniacki	10.8

1.2   Basic Terms

It is very important to understand the meaning of some basic terms that will be used frequently in this text. This section explains the meaning of an element (or member), a variable, an observation, and a data set. An element and a data set were briefly defined in Section 1.1. This section defines these terms formally and illustrates them with the help of an example.

    Table 1.1 gives information, based on *Forbes* magazine, on the total wealth of the world's eight richest persons as of March 2015. Each person listed in this table is called an **element** or a **member** of this group. Table 1.1 contains information on eight elements. Note that elements are also called **observational units**.

> **Element or Member**   An **element** or **member** of a sample or population is a specific subject or object (for example, a person, firm, item, state, or country) about which the information is collected.

Table 1.1   Total Wealth of the World's Eight Richest Persons

Name	Total Wealth ← Variable (billions of dollars)
Bill Gates	79.2
Carlos Slim Helu	77.1
Warren Buffett	72.7 ← An observation or measurement
Amancio Ortega	64.5
Larry Ellison	54.3
Charles Koch	42.9
David Koch	42.9
Christy Walton	41.7

An element or member → Warren Buffett

Source: *Forbes*, March 23, 2015.

The total wealth in our example is called a variable. The total wealth is a characteristic of these persons on which information is collected.

> **Variable** A **variable** is a characteristic under study that assumes different values for different elements. In contrast to a variable, the value of a *constant* is fixed.

A few other examples of variables are household incomes, the number of houses built in a city per month during the past year, the makes of cars owned by people, the gross profits of companies, and the number of insurance policies sold by a salesperson per day during the past month.

In general, a variable assumes different values for different elements, as illustrated by the total wealth for the eight persons in Table 1.1. For some elements in a data set, however, the values of the variable may be the same. For example, if we collect information on incomes of households, these households are expected to have different incomes, although some of them may have the same income.

A variable is often denoted by x, y, or z. For instance, in Table 1.1, the total wealth for persons may be denoted by any one of these letters. Starting with Section 1.7, we will begin to use these letters to denote variables.

Each of the values representing the total wealths of the eight persons in Table 1.1 is called an **observation** or **measurement**.

> **Observation or Measurement** The value of a variable for an element is called an **observation** or **measurement**.

From Table 1.1, the total wealth of Warren Buffett was $72.7 billion. The value $72.7 billion is an observation or a measurement. Table 1.1 contains eight observations, one for each of the eight persons.

The information given in Table 1.1 on the total wealth of the eight richest persons is called the data or a **data set**.

> **Data Set** A **data set** is a collection of observations on one or more variables.

Other examples of data sets are a list of the prices of 25 recently sold homes, test scores of 15 students, opinions of 100 voters, and ages of all employees of a company.

EXERCISES

CONCEPTS AND PROCEDURES

1.4 Explain the meaning of an element, a variable, an observation, and a data set.

APPLICATIONS

1.5 The following table lists the number of deaths by cause as reported by the Centers for Disease Control and Prevention on February 6, 2015 (*Source*: www.cdc.gov).

Cause of Death	Number of Deaths
Heart disease	611,105
Cancer	584,881
Accidents	130,557
Stroke	128,978
Alzheimer's disease	84,767
Diabetes	75,578
Influenza and Pneumonia	56,979
Suicide	41,149

Briefly explain the meaning of a member, a variable, a measurement, and a data set with reference to the information in this table.

1.6 The following table lists the number of deaths by cause as reported by the Centers for Disease Control and Prevention on February 6, 2015 (*Source*: www.cdc.gov).

Cause of Death	Number of Deaths
Heart disease	611,105
Cancer	584,881
Accidents	130,557
Stroke	128,978
Alzheimer's disease	84,767
Diabetes	75,578
Influenza and Pneumonia	56,979
Suicide	41,149

a. What is the variable for this data set?
b. How many observations are in this data set?
c. How many elements does this data set contain?

1.3   Types of Variables

In Section 1.2, we learned that a variable is a characteristic under investigation that assumes different values for different elements. Family income, height of a person, gross sales of a company, price of a college textbook, make of the car owned by a family, number of accidents, and status (freshman, sophomore, junior, or senior) of a student enrolled at a university are examples of variables.

A variable may be classified as quantitative or qualitative. These two types of variables are explained next.

1.3.1   Quantitative Variables

Some variables (such as the price of a home) can be measured numerically, whereas others (such as hair color) cannot. The price of a home is an example of a **quantitative variable** while hair color is an example of a **qualitative variable**.

Quantitative Variable   A variable that can be measured numerically is called a **quantitative variable**. The data collected on a quantitative variable are called **quantitative data**.

Income, height, gross sales, price of a home, number of cars owned, and number of accidents are examples of quantitative variables because each of them can be expressed numerically. For instance, the income of a family may be $81,520.75 per year, the gross sales for a company may be $567 million for the past year, and so forth. Such quantitative variables may be classified as either *discrete variables* or *continuous variables.*

Discrete Variables

The values that a certain quantitative variable can assume may be countable or noncountable. For example, we can count the number of cars owned by a family, but we cannot count the height of a family member, as it is measured on a continuous scale. A variable that assumes countable values is called a **discrete variable**. Note that there are no possible intermediate values between consecutive values of a discrete variable.

Discrete Variable   A variable whose values are countable is called a **discrete variable**. In other words, a discrete variable can assume only certain values with no intermediate values.

For example, the number of cars sold on any given day at a car dealership is a discrete variable because the number of cars sold must be 0, 1, 2, 3, . . . and we can count it. The number of cars sold cannot be between 0 and 1, or between 1 and 2. Other examples of discrete variables are the number of people visiting a bank on any day, the number of cars in a parking lot, the number of cattle owned by a farmer, and the number of students in a class.

Continuous Variables

Some variables assume values that cannot be counted, and they can assume any numerical value between two numbers. Such variables are called **continuous variables**.

Continuous Variable   A variable that can assume any numerical value over a certain interval or intervals is called a **continuous variable**.

The time taken to complete an examination is an example of a continuous variable because it can assume any value, let us say, between 30 and 60 minutes. The time taken may be 42.6 minutes, 42.67 minutes, or 42.674 minutes. (Theoretically, we can measure time as precisely as we

want.) Similarly, the height of a person can be measured to the tenth of an inch or to the hundredth of an inch. Neither time nor height can be counted in a discrete fashion. Other examples of continuous variables are the weights of people, the amount of soda in a 12-ounce can (note that a can does not contain exactly 12 ounces of soda), and the yield of potatoes (in pounds) per acre. Note that any variable that involves money and can assume a large number of values is typically treated as a continuous variable.

1.3.2 Qualitative or Categorical Variables

Variables that cannot be measured numerically but can be divided into different categories are called **qualitative** or **categorical variables**.

> **Qualitative or Categorical Variable** A variable that cannot assume a numerical value but can be classified into two or more nonnumeric categories is called a **qualitative** or **categorical variable.** The data collected on such a variable are called **qualitative data**.

For example, the status of an undergraduate college student is a qualitative variable because a student can fall into any one of four categories: freshman, sophomore, junior, or senior. Other examples of qualitative variables are the gender of a person, the make of a computer, the opinions of people, and the make of a car.

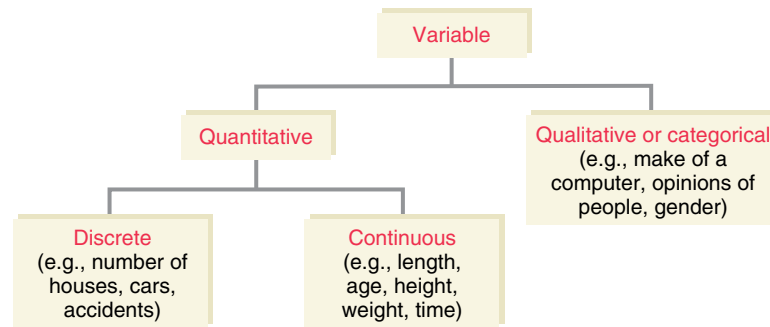Figure 1.1 summarizes the different types of variables.

Figure 1.1 Types of variables.

EXERCISES

CONCEPTS AND PROCEDURES

1.7 Explain the meaning of the following terms.

    a. Quantitative variable
    b. Qualitative variable
    c. Discrete variable
    d. Continuous variable
    e. Quantitative data
    f. Qualitative data

APPLICATIONS

1.8 Indicate which of the following variables are quantitative and which are qualitative.

    a. The amount of time a student spent studying for an exam
    b. The amount of rain last year in 30 cities
    c. The arrival status of an airline flight (early, on time, late, canceled) at an airport
    d. A person's blood type
    e. The amount of gasoline put into a car at a gas station

1.9 Classify the following quantitative variables as discrete or continuous.

    a. The amount of time a student spent studying for an exam
    b. The amount of rain last year in 30 cities
    c. The amount of gasoline put into a car at a gas station
    d. The number of customers in the line waiting for service at a bank at a given time

1.10 A survey of families living in a certain city was conducted to collect information on the following variables: age of the oldest person in the family, number of family members, number of males in the family, number of females in the family, whether or not they own a house, income of the family, whether or not the family took vacations during the past one year, whether or not they are happy with their financial situation, and the amount of their monthly mortgage or rent.

a. Which of these variables are qualitative variables?
b. Which of these variables are quantitative variables?
c. Which of the quantitative variables of part b are discrete variables?
d. Which of the quantitative variables of part b are continuous variables?

1.4 Cross-Section Versus Time-Series Data

Based on the time over which they are collected, data can be classified as either cross-section or time-series data.

1.4.1 Cross-Section Data

Cross-section data contain information on different elements of a population or sample for the *same* period of time. The information on incomes of 100 families for 2015 is an example of cross-section data. All examples of data already presented in this chapter have been cross-section data.

Cross-Section Data   Data collected on different elements at the same point in time or for the same period of time are called **cross-section data**.

Table 1.1, reproduced here as Table 1.2, shows the total wealth of each of the eight richest persons in the world. Because this table presents data on the total wealth of eight persons for the same period, it is an example of cross-section data.

Table 1.2   Total Wealth of World's Eight Richest Persons

Name	Total Wealth (billions of dollars)
Bill Gates	79.2
Carlos Slim Helu	77.1
Warren Buffett	72.7
Amancio Ortega	64.5
Larry Ellison	54.3
Charles Koch	42.9
David Koch	42.9
Christy Walton	41.7

Source: *Forbes*, March 23, 2015.

1.4.2 Time-Series Data

Time-series data contain information on the same element at *different* points in time. Information on U.S. exports for the years 2001 to 2015 is an example of time-series data.

Time-Series Data   Data collected on the same element for the same variable at different points in time or for different periods of time are called **time-series data**.