GLOBAL EDITION

STATISTICS for BUSINESS and ECONOMICS

14TH EDITION

James T. McClave • P. George Benson • Terry Sincich

APPLET CORRELATION

Applet	Concept Illustrated	Description	Applet Activity
Random numbers	Uses a random number generator to determine the experimental units to be included in a sample.	Generates random numbers from a range of integers specified by the user.	**1.1**, 44; **1.2**, 44; **3.6**, 198; **4.1**, 219; **4.2**, 219; **4.8**, 271
Sample from a population	Assesses how well a sample represents the population and the role that sample size plays in the process.	Produces random sample from population from specified sample size and population distribution shape. Reports mean, median, and standard deviation; applet creates plot of sample.	**4.4**, 233; **4.6**, 257; **4.7**, 271
Sampling distributions	Compares means and standard deviations of distributions; assesses effect of sample size; illustrates undbiasedness.	Simulates repeatedly choosing samples of a fixed size n from a population with specified sample size, number of samples, and shape of population distribution. Applet reports means, medians, and standard deviations; creates plots for both.	**5.1**, 310; **5.2**, 310

Long-run probability demonstrations illustrate the concept that theoretical probabilities are long-run experimental probabilities.

Applet	Concept Illustrated	Description	Applet Activity
Simulating probability of rolling a 6	Investigates relationship between theoretical and experimental probabilities of rolling 6 as number of die rolls increases.	Reports and creates frequency histogram for each outcome of each simulated roll of a fair die. Students specify number of rolls; applet calculates and plots proportion of 6s.	**3.1**, 162; **3.3**, 174; **3.4**, 175; **3.5**, 188
Simulating probability of rolling a 3 or 4	Investigates relationship between theoretical and experimental probabilities of rolling 3 or 4 as number of die rolls increases.	Reports outcome of each simulated roll of a fair die; creates frequency histogram for outcomes. Students specify number of rolls; applet calculates and plots proportion of 3s and 4s.	**3.3**, 174; **3.4**, 175
Simulating the probability of heads: fair coin	Investigates relationship between theoretical and experimental probabilities of getting heads as number of fair coin flips increases.	Reports outcome of each fair coin flip and creates a bar graph for outcomes. Students specify number of flips; applet calculates and plots proportion of heads.	**3.2**, 162; **4.2**, 219
Simulating probability of heads: unfair coin $(P(H) = .2)$	Investigates relationship between theoretical and experimental probabilities of getting heads as number of unfair coin flips increases.	Reports outcome of each flip for a coin where heads is less likely to occur than tails and creates a bar graph for outcomes. Students specify number of flips; applet calculates and plots the proportion of heads.	**4.3**, 233
Simulating probability of heads: unfair coin $(P(H) = .8)$	Investigates relationship between theoretical and experimental probabilities of getting heads as number of unfair coin flips increases.	Reports outcome of each flip for a coin where heads is more likely to occur than tails and creates a bar graph for outcomes. Students specify number of flips; applet calculates and plots the proportion of heads.	**4.3**, 233
Simulating the stock market	Theoretical probabilities are long run experimental probabilities.	Simulates stock market fluctuation. Students specify number of days; applet reports whether stock market goes up or down daily and creates a bar graph for outcomes. Calculates and plots proportion of simulated days stock market goes up.	**4.5**, 234
Mean versus median	Investigates how skewedness and outliers affect measures of central tendency.	Students visualize relationship between mean and median by adding and deleting data points; applet automatically updates mean and median.	**2.1**, 88; **2.2**, 88; **2.3**, 88

(*Continued*)

Applet	**Concept Illustrated**	**Description**	**Applet Activity**
Standard deviation	Investigates how distribution shape and spread affect standard deviation.	Students visualize relationship between mean and standard deviation by adding and deleting data points; applet updates mean and standard deviation.	**2.4**, 96; **2.5**, 96; **2.6**, 96; **2.7**, 118
Confidence intervals for a mean (the impact of confidence level)	Not all confidence intervals contain the population mean. Investigates the meaning of 95% and 99% confidence.	Simulates selecting 100 random samples from population; finds 95% and 99% confidence intervals for each. Students specify sample size, distribution shape, and population mean and standard deviation; applet plots confidence intervals and reports number and proportion containing true mean.	**6.1**, 336; **6.2**, 336
Confidence intervals for a mean (not knowing standard deviation)	Confidence intervals obtained using the sample standard deviation are different from those obtained using the population standard deviation. Investigates effect of not knowing the population standard deviation.	Simulates selecting 100 random samples from the population and finds the 95% *z*-interval and 95% *t*-interval for each. Students specify sample size, distribution shape, and population mean and standard deviation; applet plots confidence intervals and reports number and proportion containing true mean.	**6.3**, 346; **6.4**, 346
Confidence intervals for a proportion	Not all confidence intervals contain the population proportion. Investigates the meaning of 95% and 99% confidence.	Simulates selecting 100 random samples from the population and finds the 95% and 99% confidence intervals for each. Students specify population proportion and sample size; applet plots confidence intervals and reports number and proportion containing true proportion.	**6.5**, 354; **6.6**, 354
Hypothesis tests for a mean	Not all tests of hypotheses lead correctly to either rejecting or failing to reject the null hypothesis. Investigates the relationship between the level of confidence and the probabilities of making Type I and Type II errors.	Simulates selecting 100 random samples from population; calculates and plots *t* statistic and *P*-value for each. Students specify population distribution shape, mean, and standard deviation; sample size, and null and alternative hypotheses; applet reports number and proportion of times null hypothesis is rejected at both 0.05 and 0.01 levels.	**7.1**, 397; **7.2**, 408; **7.3**, 408; **7.4**, 408
Hypothesis tests for a proportion	Not all tests of hypotheses lead correctly to either rejecting or failing to reject the null hypothesis. Investigates the relationship between the level of confidence and the probabilities of making Type I and Type II errors.	Simulates selecting 100 random samples from population; calculates and plots *z*-statistic and *P*-value for each. Students specify population proportion, sample size, and null and alternative hypotheses; applet reports number and proportion of times null hypothesis is rejected at 0.05 and 0.01 levels.	**7.5**, 424; **7.6**, 425
Correlation by eye	Correlation coefficient measures strength of linear relationship between two variables. Teaches user how to assess strength of a linear relationship from a scattergram.	Computes correlation coefficient *r* for a set of bivariate data plotted on a scattergram. Students add or delete points and guess value of *r*; applet compares guess to calculated value.	**11.2**, 682
Regression by eye	The least squares regression line has a smaller SSE than any other line that might approximate a set of bivariate data. Teaches students how to approximate the location of a regression line on a scattergram.	Computes least squares regression line for a set of bivariate data plotted on a scattergram. Students add or delete points and guess location of regression line by manipulating a line provided on the scattergram; applet plots least squares line and displays the equations and the SSEs for both lines.	**11.1**, 657

Statistics

for Business and Economics

This page is intentionally left blank

14 EDITION

GLOBAL EDITION

Statistics

for Business and Economics

James T.
MCCLAVE
Info Tech, Inc.
University of Florida

P. George
BENSON
College of
Charleston

Terry
SINCICH
University of South
Florida

Pearson

Harlow, England • London • New York • Boston • San Francisco • Toronto • Sydney • Dubai • Singapore • Hong Kong
Tokyo • Seoul • Taipei • New Delhi • Cape Town • São Paulo • Mexico City • Madrid • Amsterdam • Munich • Paris • Milan

Credits and acknowledgments borrowed from other sources and reproduced, with permission, in this textbook appear on the appropriate page of appearance or in the Credits on pages

Cover image by MikeDotta / Shutterstock

Pearson Education Limited
KAO Two
KAO Park
Hockham Way
Harlow
Essex
CM17 9SR
United Kingdom

and Associated Companies throughout the world

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

Authorized adaptation from the United States edition, entitled *Statistics for Business and Economics*, 14th Edition, ISBN 978-0-13-685535-4 by James T. McClave, P. George Benson, and Terry Sincich, published by Pearson Education © 2022.

ISBN 10: 1-292-41339-5 (print)
ISBN 13: 978-1-292-41339-6 (print)
eBook ISBN 13: 978-1-292-41352-5 (eBook/uPDF)

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

1 22

Typeset in Times NR MT Pro by B2R Technologies Pvt. Ltd.

Contents

Preface

This 14th edition of *Statistics for Business and Economics* is an introductory text emphasizing inference, with extensive coverage of data collection and analysis as needed to evaluate the reported results of statistical studies and make good decisions. As in earlier editions, the text stresses the development of statistical thinking, the assessment of credibility and value of the inferences made from data, both by those who consume and by those who produce them. It assumes a mathematical background of basic algebra.

The text incorporates the following features, developed from the American Statistical Association (ASA) sponsored conferences on *Making Statistics More Effective in Schools of Business* (MSMESB) and ASA's Guidelines for Assessment and Instruction in Statistics Education (GAISE) Project:

- Emphasize statistical literacy and develop statistical thinking
- Use real data in applications
- Use technology for developing conceptual understanding and analyzing data
- Foster active learning in the classroom
- Stress conceptual understanding rather than mere knowledge of procedures
- Emphasize intuitive concepts of probability

New in the 14th Edition

- **More than 1,200 exercises, with revisions and updates to 30%.** Many new and updated exercises, based on contemporary business-related studies and real data, have been added. Most of these exercises foster critical thinking skills.
- **Data Informed Development.** The authors analyzed aggregated student usage and performance data from MyLab Statistics for the previous edition of this text. The results of this analysis helped improve the quality and quantity of exercises that matter most to instructors and students.
- **Updated technology.** All printouts from statistical software (Excel 2019/XLSTAT, StatCrunch 3.0, Minitab 19, and the TI-84 Graphing Calculator) and corresponding instructions for use have been revised to reflect the latest versions of the software.
- **Statistics in Action Cases Updated.** Three of the 14 Statistics in Action cases have been updated. All cases are based on real data from a recent business study.
- **Continued Emphasis on Ethics.** Where appropriate, boxes have been added to emphasize the importance of ethical behavior when collecting, analyzing, and interpreting data with statistics.
- **Business Analytics.** The importance of statistical thinking to successful business analytics is established early in the text.
- **Short Video Tutorials.** New videos guide students through real-life applications of chapter topics to illustrate how these concepts translate to everyday life.

Hallmark Strengths

We have maintained the pedagogical features of *Statistics for Business and Economics* that we believe make it unique among introductory business statistics texts. These features, which assist the student in achieving an overview of statistics and an understanding of its relevance in both the business world and everyday life, are as follows:

- **Use of Examples as a Teaching Device** Almost all new ideas are introduced and illustrated by data-based applications and examples. We believe that students better understand definitions, generalizations, and theoretical concepts *after* seeing an application. All examples have three components: (1) "Problem," (2) "Solution," and (3) "Look Back" (or "Look Ahead"). This step-by-step process provides students with a defined structure by which to approach problems and enhances their problem-solving skills. The "Look Back/Look Ahead" feature often gives helpful hints to solving the problem and/or provides a further reflection or insight into the concept or procedure that is covered.
- **Now Work** A "Now Work" exercise suggestion follows each example. The Now Work exercise (marked with the NW icon in the exercise sets) is similar in style and concept to the text example. This provides students with an opportunity to immediately test and confirm their understanding.
- **Statistics in Action** Each chapter begins with a case study based on an actual contemporary, controversial or high-profile issue in business. Relevant research questions and data from the study are presented and the proper analysis is demonstrated in short "Statistics in Action Revisited" sections throughout the chapter. These motivate students to critically evaluate the findings and think through the statistical issues involved.
- **"Hands-On" Activities for Students** At the end of each chapter, students are provided with an opportunity to participate in hands-on classroom activities, ranging from real data collection to formal statistical analysis. These activities are designed to be performed by students individually or as a class.
- **Applet Exercises.** The text is accompanied by applets (short computer programs), available on the student resource site (www.pearsonglobaleditions.com) and in MyLab Statistics. These point-and-click applets allow students to easily run simulations that visually demonstrate some of the more difficult statistical concepts (e.g., sampling distributions and confidence intervals.) Each chapter contains several optional applet exercises in the exercise sets. They are denoted with the following Applet icon: .
- **Real-World Business Cases** Seven extensive business problem-solving cases, with real data and assignments for the student, are provided. Each case serves as a good capstone and review of the material that has preceded it. Typically, these cases follow a group of two or three chapters and require the student to apply the methods presented in these chapters.
- **Real Data–Based Exercises** The text includes more than 1,200 exercises based on applications in a variety of business disciplines and research areas. All applied exercises use current real data extracted from current publications (e.g., newspapers, magazines, current journals, and the Internet). Some students have difficulty learning the mechanics of statistical techniques when all problems are couched in terms of realistic applications. For this reason, all exercise sections are divided into at least four parts:

  Learning the Mechanics. Designed as straightforward applications of new concepts, these exercises allow students to test their ability to comprehend a mathematical concept or a definition.

  Applying the Concepts—Basic. Based on applications taken from a wide variety of business journals, newspapers, and other sources, these short exercises help students to begin developing the skills necessary to diagnose and analyze real-world problems.

  Applying the Concepts—Intermediate. Based on more detailed real-world applications, these exercises require students to apply their knowledge of the technique presented in the section.

Applying the Concepts—Advanced. These more difficult real-data exercises require students to use their critical thinking skills.

Critical Thinking Challenges. Placed at the end of the "Supplementary Exercises" section only, this feature presents students with one or two challenging business problems.

- **Exploring Data with Statistical Computer Software and the Graphing Calculator** Each statistical analysis method presented is demonstrated using output from three leading Windows-based statistical software packages: Excel/XLSTAT, StatCrunch, and Minitab. Students are exposed early and often to computer printouts they will encounter in today's hi-tech business world.

- **"Using Technology" Tutorials** At the end of each chapter are statistical software tutorials with point-and-click instructions (with screen shots) for Minitab, StatCrunch, and Excel/XLSTAT. These tutorials are easily located and show students how to best use and maximize statistical software. In addition, output and keystroke instructions for the TI-84 Graphing Calculator are presented.

- **Profiles of Statisticians in History (Biography)** Brief descriptions of famous statisticians and their achievements are presented in side boxes. In reading these profiles, students will develop an appreciation for the statistician's efforts and the discipline of statistics as a whole.

- **Data and Applets** The text is accompanied by a website (www.pearsonglobaleditions.com) that contains files for all of the data sets marked with an icon in the text. These include data sets for text examples, exercises, Statistics in Action, and Real-World cases. Data files are available in multiple formats: Excel and Minitab. This website also contains the applets that are used to illustrate statistical concepts.

Flexibility in Coverage

The text is written to allow the instructor flexibility in coverage of topics. Suggestions for two topics, probability and regression, are given below.

- **Probability and Counting Rules** One of the most troublesome aspects of an introductory statistics course is the study of probability. Probability poses a challenge for instructors because they must decide on the level of presentation, and students find it a difficult subject to comprehend. We believe that one cause for these problems is the mixture of probability and counting rules that occurs in most introductory texts. Consequently, we have included the counting rules (with examples) in an appendix (Appendix B) rather than in the body of Chapter 3. Thus, the instructor can control the level of coverage of probability.

- **Multiple Regression and Model Building** This topic represents one of the most useful statistical tools for the solution of applied problems. Although an entire text could be devoted to regression modeling, we feel that we have presented coverage that is understandable, usable, and much more comprehensive than the presentations in other introductory statistics texts. We devote two full chapters to discussing the major types of inferences that can be derived from a regression analysis, showing how these results appear in the output from statistical software, and, most important, selecting multiple regression models to be used in an analysis. Thus, the instructor has the choice of a one-chapter coverage of simple linear regression (Chapter 11), a two-chapter treatment of simple and multiple regression (excluding the sections on model building in Chapter 12), or complete coverage of regression analysis, including model building and regression diagnostics. This extensive coverage of such useful statistical tools will provide added evidence to the student of the relevance of statistics to real-world problems.

- **Role of Calculus in Footnotes** Although the text is designed for students with a noncalculus background, **footnotes** explain the role of calculus in various derivations. Footnotes are also used to inform the student about some of the theory underlying certain methods of analysis. These footnotes allow additional flexibility in the mathematical and theoretical level at which the material is presented.

Acknowledgments

This book reflects the efforts of a great many people over a number of years. First, we would like to thank the following professors, whose reviews and comments on this and prior editions have contributed to the 14th edition:

Reviewers of the 14th Edition of *Statistics for Business and Economics*

Anna Errore, University of Minnesota Twin Cities
Alka Gandhi, University of Maryland-College Park
Stacey Hachigian, Humber College
Seunghee Lee, Pellissippi State Community College
Amit Mitra, Auburn University
Yu Yue, Baruch College

Reviewers of Previous Editions

ALABAMA Volodymyr Melnykov, *University of Alabama–Tuscaloosa* **ARKANSAS** Julie Trivitt, *University of Arkansas* **CALIFORNIA** Joyce Curley-Daly, Jim Daly, Robert K. Smidt, *California Polytechnic State University* • Jim Davis, *Golden Gate University* • Carol Eger, *Stanford University* • Paul W. Guy, *California State University, Chico* • Judd Hammack, P. Kasliwal, *California State University, Los Angeles* • Mabel T. King, *California State University, Fullerton* • James Lackritz, *California State University, San Diego* • Beth Rose, *University of Southern California* • Daniel Sirvent, *Vanguard University* **COLORADO** Rick L. Edgeman, Charles F. Warnock, *Colorado State University* • Eric Huggins, *Fort Lewis College* • William J. Weida, *United States Air Force Academy* **CONNECTICUT** Alan E. Gelfand, Joseph Glaz, Timothy J. Killeen, *University of Connecticut* **DELAWARE** Christine Ebert, *University of Delaware* **DISTRICT OF COLUMBIA** Phil Cross, Jose Luis Guerrero-Cusumano, *Georgetown University* • Gaminie Meepagala, *Howard University* **FLORIDA** John M. Charnes, *University of Miami* • C. Brad Davis, *Clearwater Christian College* • Vivian Jones, *Bethune-Cookman University* • Fred Leysieffer, Pi-Erh Lin, Doug Zahn, *Florida State University* • P. V. Rao, *University of Florida* • Laura Reisert, *Florida International University* • Jeffrey W. Steagall, *University of North Florida* • Edna White, *Florida Atlantic University* **GEORGIA** Robert Elrod, *Georgia State University* • Karen Smith, *West Georgia University* **HAWAII** Steve Hora, *University of Hawaii, Hilo* **ILLINOIS** Arunas Dagys, *St. Xavier University* • Edward Minieka, *University of Illinois at Chicago* • Don Robinson, *Illinois State University* • Chipei Tseng, *Northern Illinois University* • Pankaj Vaish, *Arthur Andersen & Company* **IOWA** Dileep Dhavale, *University of Northern Iowa* • William Duckworth II, William Q. Meeker, *Iowa State University* • Tim E. McDaniel, *Buena Vista University* **KANSAS** Paul I. Nelson, *Kansas State University* • Lawrence A. Sherr, *University of Kansas* **KENTUCKY** Richard N. McGrath, *Bowling Green State University* **LOUISIANA** James Willis, *Louisiana State University* **MARYLAND** John F. Beyers, Michael Kulansky, *University of Maryland–University College* • Glenn J. Browne, Mary C. Christman, *University of Maryland* **MASSACHUSETTS** Warren M. Holt, *Southeastern Massachusetts University* • Remus Osan, *Boston University* **MICHIGAN** Atul Agarwal, Petros Ioannatos, *GMI Engineering and Management Institute* • Richard W. Andrews, Peter Lenk, Benjamin Lev, *University of Michigan* • Leszek Gawarecki, *Kettering University* • Toni M. Somers, *Wayne State University* • William Welch, *Saginaw Valley State University* • T. J. Wharton, *Oakland University* **MINNESOTA** Gordon J. Alexander, Donald W. Bartlett, David M. Bergman, Atul Bhatia, Steve Huchendorf, Benny Lo, Karen Lundquist, Vijay Pisharody, Donald N. Steinnes, Robert W. Van Cleave, Steve Wickstrom, *University of Minnesota* • Daniel G. Brick, Leigh Lawton, *University of St. Thomas* • Susan Flach, *General Mills, Inc.* • David D. Krueger, Ruth K. Meyer, Jan Saraph, Gary Yoshimoto, *St. Cloud State University* • Paula M. Oas, *General Office Products* • Fike Zahroom, *Moorhead State University* **MISSISSIPPI** Eddie M. Lewis, *University of Southern Mississippi* • Alireza Tahai, *Mississippi State University* **MISSOURI** James Holstein, Lawrence D. Ries,

University of Missouri, Columbia • Marius Janson, L. Douglas Smith, *University of Missouri, St. Louis* • Farroll Tim Wright, *University of Missouri* • Stephanie Schartel-Dunn *Missouri Western University* **NEBRASKA** James Wright, *Chadron State College* **NEW HAMPSHIRE** Ken Constantine, *University of New Hampshire* **NEW JERSEY** Lewis Coopersmith, Cengiz Haksever, *Rider University* • Lei Lei, Xuan Li, Zina Taran, *Rutgers University* • Philip Levine, Leonard Presby, *William Paterson University* **NEW MEXICO** S. Howard Kraye, *University of New Mexico* **NEW YORK** James Czachor, *Fordham-Lincoln Center, AT&T* • Bernard Dickman, *Hofstra University* • Joshua Fogel, *Brooklyn College of City University of New York* • Martin Labbe, *State University of New York, College at New Paltz* • Kenneth Leong, *College of New Rochelle* • Mark R. Marino, *Niagara University/Erie Community College* • G. E. Martin, *Clarkson University* • Thomas J. Pfaff, *Ithaca College* • Gary Simon, *New York University, Stern School of Business* • Rungrudee Suetorsak, *SUNY-Fredonia* **NORTH CAROLINA** Golam Azam, *North Carolina Agricultural & Technical University* • Edward Carlstein, Douglas A. Elvers, *University of North Carolina at Chapel Hill* • Barry P. Cuffe, *Wingate University* • Don Holbert, *East Carolina University* • J. Morgan Jones, *University of North Carolina* • Douglas S. Shafer, *University of North Carolina, Charlotte* **OHIO** William H. Beyer, *University of Akron* • Michael Broida, Tim Krehbiel, *Miami University of Ohio* • Chih-Hsu Cheng, Douglas A. Wolfe, *Ohio State University* • Ronald L. Coccari, *Cleveland State University* • Richard W. Culp, *Wright-Patterson AFB, Air Force Institute of Technology* **OKLAHOMA** Larry Claypool, Brenda Masters, Rebecca Moore, *Oklahoma State University* • Robert Curley, *University of Central Oklahoma* **PENNSYLVANIA** Mohammed Albohali, Douglas H. Frank, *Indiana University of Pennsylvania* • Sukhwinder Bagi, *Bloomsburg University* • Carl Bedell, *Philadelphia College of Textiles and Science* • Ann Hussein, *Philadelphia University* • Behnam Nakhai, *Millersville University* • Rose Prave, *University of Scranton* • Farhad Saboori, *Albright College* • Kathryn Szabet, *LaSalle University* • Pandu Tadikamalla *University of Pittsburgh* • Christopher J. Zappe, *Bucknell University* **SOUTH CAROLINA** Iris Fetta, Robert Ling, *Clemson University* • Kathleen M. Whitcomb, *University of South Carolina* **TENNESSEE** Francis J. Brewerton, *Middle Tennessee State University* **TEXAS** Larry M. Austin, *Texas Tech University* • Jim Branscome, Robert W. Brobst, Mark Eakin, Grace Esimai, Michael E. Hanna, Craig W. Slinkman, *University of Texas at Arlington* • Katarina Jegdic, *University of Houston–Downtown* • Virgil F. Stone, *Texas A & M University* **VIRGINIA** Edward R. Clayton, *Virginia Polytechnic Institute and State University* **WASHINGTON** June Morita, Kim Tamura, *University of Washington* **WISCONSIN** Ross H. Johnson, *Madison College* **WASHINGTON, D.C.** Keith Ord, *Georgetown University* • Balaji Srinivasan, *George Washington University* **CANADA** Clarence Bayne, *Concordia University* • Edith Gombay, *University of Alberta* **TURKEY** Dilek Onkal, *Bilkent University, Ankara* **OTHER** Michael P. Wegmann, *Keller Graduate School of Management*

Other Contributors

Special thanks are due to our supplements author, Mark Dummeldinger, who has worked with us for many years. Accuracy checker Engin Sungur helped ensure a highly accurate, clean text. Nati Jain and Shantel Vargas diligently researched new data and cases for the exercise sets. Finally, the Pearson Education staff of Amanda Brands, Suzanna Bainbridge, Karen Montgomery, Alicia Wilson, Demetrius Hall, Joe Vetere, Peggy McMahon, Jean Choe, and Bob Carroll, who helped greatly with all phases of the text development, production, and marketing effort. Our gratitude also goes to Heidi Aguiar of Spi Global for overseeing the production process.

Global Edition Acknowledgments

Pearson would like to thank Alicia Tan Yiing Fei, Taylor's University Malaysia, for developing content for this Global Edition. Pearson would also like to thank Simon Trimborn, City University of Hong Kong; and John Alexander Wright, The Chinese University of Hong Kong, for sharing suggestions that were valuable to us in developing the content for the Global Edition.

MyLab Statistics Resources for Success

MyLab Statistics is available to accompany Pearson's market-leading text options, including Statistics for Business and Economics, 14th Edition (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.

Student Resources

Excel Technology Manual, by Mark Dummeldinger (University of South Florida), provides tutorial instruction and worked-out text examples for Excel. The *Excel Technology Manual* is available for download at www.pearsonglobaleditions.com or within MyLab Statistics.

NEW! Videos by author Terry Sincich have been added to the Video and Resource Library. These videos illustrate the solutions to select exercises in the book, demonstrate key topics in detail, and include tutorials on real-world statistics in action.

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.

For the Instructor

Instructor's Solutions Manual, by Mark Dummeldinger (University of South Florida), provides detailed, worked-out solutions to all of the book's exercises. Careful attention has been paid to ensure that all methods of solution and notation are consistent with those used in the core text. Available for download at www.pearsonglobaleditions.com.

TestGen® (www.pearsoned.com/testgen) enables instructors to build, edit, print, and administer tests using a computerized bank of questions developed to cover all the objectives of the text. TestGen is algorithmically based, allowing instructors to create multiple but equivalent versions of the same question or test with the click of a button. Instructors can also modify test bank questions or add new questions. The software and testbank are available for download from Pearson's Instructor Resource Center.

Student's Solutions Manual, by Mark Dummeldinger (University of South Florida), provides detailed, worked-out solutions to all odd-numbered text exercises. This item is available within MyLab Statistics and can be shared by the instructor only.

Data Analytics

Instructors have a comprehensive gradebook with enhanced reporting functionality that makes it easier to understand which students are struggling, and which topics they struggle with most.

1

CONTENTS

WHERE WE'RE GOING

- Introduce the field of statistics **(1.1)**
- Demonstrate how statistics applies to business **(1.2)**
- Introduce the language of statistics and the key elements of any statistical problem **(1.3)**
- Differentiate between population and sample data **(1.3)**
- Differentiate between descriptive and inferential statistics **(1.3)**
- Introduce the key elements of a process **(1.4)**
- Identify the different types of data and data-collection methods **(1.5–1.6)**
- Discover how critical thinking through statistics can help improve our quantitative literacy **(1.7)**

Statistics, Data, and Statistical Thinking

STATISTICS IN ACTION: A 20/20 View of Surveys and Studies: Facts or Fake News?

Several years ago, the popular ABC television program 20/20 aired a story titled "Fact or Fiction?—Exposés of So-Called Surveys." The focus of the program segment was on what we now call "fake news" or "alternative facts," that is, false information that is often highly publicized in the media (newspapers, magazines, TV shows, Twitter, Instagram, etc.). Several misleading (and possibly unethical) surveys, conducted by businesses or special interest groups with specific objectives in mind, were presented on the ABC program. Several are listed in Table SIA1.1, as well as some recent misleading studies used in product advertisements.

For this *Statistics in Action*, we consider research sponsored by Mars Corp. and published in the journal *Nature Neuroscience* (Dec. 2014). Researchers from Columbia University wondered whether taking cocoa supplements would enhance a region of the brain that deteriorates with age and is associated with age-related memory loss. They concluded that cocoa supplements can indeed boost cognition in older adults. Other similar studies claim that chocolate will reduce cardiovascular disease risk and help with weight loss. These results were reported on by media outlets such as the *New York Times*, with headlines like, "*To improve a memory, consider chocolate!*", "*Good news for chocolate lovers: The more you eat, the lower your risk of heart disease*," or simply "*Chocolate is good for you*." These reported "facts" have likely grown consumer demand for chocolate. At a time when candy sales overall has declined, chocolate retail sales in the United States have risen from $14.2 billion in 2007 to $18.9 billion in 2017.

Critical thinkers may question chocolate as a powerful health food. One such group, Vox Media (November 7, 2017), investigated this chocolate phenomenon.

STATISTICS IN ACTION

(continued)

First, Vox discovered that chocolate companies like Mars, "have spent millions of dollars for scientific studies and research grants that support cocoa science. And, of 100 [recent] Mars-sponsored studies on cocoa, chocolate, and health, 98 had conclusions that were favorable to the candy maker in some way." This unusually high percentage of favorable studies, according to Vox, "raises questions about the quality of the studies, given that Mars and other chocolate makers can use the positive findings to market their products." This finding motivated Vox Media to critically analyze the *Nature Neuroscience* chocolate study.

To conduct this study, the Vox Media researchers randomly assigned 37 people to one of four groups. Each subject in group 1 was given a high daily dose (900 mg) of cocoa flavanol supplements and assigned one hour of aerobic exercise four times per week. Subjects in group 2 received the same high dose of cocoa flavanol supplements but were not assigned to exercise. Group 3 subjects received a low dose of cocoa flavanols (10 mg) and were assigned one hour of aerobic exercise four times per week. Finally, the last group received a low dose of cocoa flavanols but was not assigned exercise (See Figure SIA1.1). After a 3-month period, the researchers tested whether cocoa flavanol supplements staved off cognitive decline in a region of the brain associated with age-related memory loss. They did this by measuring brain waves in an MRI machine and by using an object-recognition task to test memory and reaction time. The researchers also tested if exercise had any effect on memory.

Group 1: High Cocoa / Exercise (8 subjects)	*Group 2*: High Cocoa / No Exercise (11 subjects)
Group 3: Low Cocoa / Exercise (9 subjects)	*Group 4*: Low Cocoa / No Exercise (9 subjects)

Figure SIA1.1
Schematic of Chocolate Study

The researchers reported that exercise had no impact on brain function—but cocoa flavanols did. Subjects receiving a high cocoa flavanol dosage demonstrated a greater improvement in cognitive performance than those in the low dosage groups. However, as reported by Vox Media, the researchers drew conclusions that went beyond the scope of the study. For example, the researchers claimed that the effects they saw in the high-flavanol group demonstrated that cocoa could reverse age-related memory decline by 30 years. Vox also discovered problems with the study's small sample size and design.

Henry Drysdale, a doctor and fellow at Oxford University's Center for Evidence-Based Medicine elaborated on the design issue. First, he warned that eating cocoa supplements in order to improve memory in three months is not relevant to real-world age-related memory decline. Second, the doctor pointed out the need for a much larger (than 37) group of study participants, and to conduct the trial for several years. Finally, Drysdale commented on the study variables: "Instead of only tracking the study participants' brain waves in an MRI machine (which is not a measure of cognitive ability), or using an object recognition task to test memory, you'd also want to measure outcomes that matter in people's lives, like, whether those taking cocoa could remember what they did that morning or that they had a doctor's appointment next week better than the people who didn't take the cocoa."

Ultimately, Vox Media stated that, "this trial only demonstrated that supplements seem to enhance brain function over a period of weeks, and only according to a very specific (and not very widely used) test of cognitive function. That is far from valid proof that cocoa is a memory enhancer."

To conclude the introduction to this *Statistics in Action*, we return to the ABC *20/20* TV program segment. The segment ended with an interview of Cynthia Crossen, author of *Tainted Truth: The Manipulation of Fact in America*, an exposé of misleading and biased surveys.

Some 20 years before the term "fake news" was coined, Crossen warned, "If everybody is misusing numbers and scaring us with numbers to get us to do something, however good [that something] is, we've lost the power of numbers. Now, we know certain things from research. For example, we know that smoking cigarettes is hard on your lungs and heart, and because we know that, many people's lives have been extended or saved. We don't want to lose the power of information to help us make decisions, and that's what I worry about."

Table SIA1.1: Examples of "Fake News"

Fake News (Source)	Actual Study Information/Flaw
1. Eating oat bran is a cheap and easy way to reduce your cholesterol. (*Quaker Oats*)	Diet must consist of nothing but oat bran to reduce your cholesterol count.
2. One in four American children under age 12 is hungry or at risk of hunger. (*Food Research and Action Center*)	Based on responses to questions: "Do you ever cut the size of meals?" "Do you ever eat less than you feel you should?" "Did you ever rely on limited numbers of foods to feed your children because you were running out of money to buy food for a meal?"
3. There is a strong correlation between a CEO's golf handicap and the company's stock performance: The lower the CEO's handicap (i.e., the better the golfer), the better the stock performs. (*New York Times*, May 31, 1998)	Survey sent to CEOs of 300 largest US companies; only 51 revealed their golf handicaps. Data for several top-ranking CEOs were excluded from the analysis.
4. Prior to the passing of the federal government's health reform act, 30% of employers are predicted to "definitely" or "probably" stop offering health coverage. (*McKinsey & Company Survey*, February 2011)	Online survey of 1,329 private-sector employers in the United States. Respondents were asked leading questions that made it logical to stop offering health insurance.
5. In an advertisement, "more than 80% of dentists surveyed recommend Colgate toothpaste to patients." (*Colgate-Palmolive Company*, January 2007)	The survey allowed each dentist to recommend more than one toothpaste.The Advertising Standards Authority cited and fined Colgate for a misleading ad (implying 80% of dentists recommend Colgate toothpaste in preference to all other brands) and banned the advertisement.
6. An advertisement for Kellogg's Frosted Mini-Wheats claimed that the cereal was "clinically shown to improve kids' attentiveness by nearly 20%." (*Kellogg Company*, 2009)	Only half of the kids in the study showed any improvement in attentiveness; only 1 in 7 improved by 18% or more, and only 1 in 9 improved by 20% or more; kids who ate Frosted Mini-Wheats were compared against kids who had only water for breakfast. (The Kellogg Company agreed to pay $4 million to settle suit over false ad claim.)
7. On the basis of a commissioned study, Walmart advertised that it "was responsible for an overall 3.1% decline in consumer prices" and it "saves customers over $700 per year." (*Global Insight*, 2005)	The Economic Policy Institute noted that the Global Insight study was based on the retailer's impact on the Consumer Price Index (CPI)—but 60% of the items in the CPI are services, not commodities that can be purchased at Walmart. (Walmart was forced to withdraw the misleading advertisement.)
8. In a survey commissioned by cable provider Comcast, respondents were asked to decide which cable provider, Comcast or DIRECTV, offered more HD channels. Respondents were shown channel lists for DIRECTV (List #387) and Comcast (List #429). (*NAD Case Report No. 5208*, August 25, 2010).	The National Advertising Division (NAD) of the Council of Better Business Bureaus rejected the survey after finding that the higher list number (#429) "served as a subtle, yet effective cue" that Comcast's list contained more channels.

STATISTICS IN ACTION (*continued*)

Fake News (Source)	Actual Study Information/Flaw
9. NPR reported on a recent study that found that teens who spend five or more hours per day on their smartphones are 71 percent more likely to have one risk factor for suicide/depression (*Journal of Abnormal Psychology*, March 2019).	Data was based on teen's perceived use of their smartphones. Studies have shown that perceived use is poorly related to actual use measured with an app. Also, the measures of addiction or suicide risk were based on unreliable scales.

In the following *Statistics in Action Revisited* sections, we discuss several key statistical concepts covered in this chapter that are relevant to misleading surveys and studies.

STATISTICS IN ACTION REVISITED

- Identifying the population, sample, and inference **(p. 29)**
- Identifying the data-collection method and data type **(p. 39)**
- Critically assessing the ethics of a statistical study **(p. 42)**

1.1 The Science of Statistics

What does *statistics* mean to you? Does it bring to mind batting averages? Gallup polls, unemployment figures, or numerical distortions of facts (lying with statistics)? Or is it simply a college requirement you have to complete? We hope to persuade you that statistics is a meaningful, useful science with a broad scope of applications to business, government, and the physical and social sciences that is almost limitless. We also want to show that statistics can lie only when they are misapplied. Finally, we wish to demonstrate the key role statistics play in critical thinking—whether in the classroom, on the job, or in everyday life. Our objective is to leave you with the impression that the time you spend studying this subject will repay you in many ways.

Although the term can be defined in many ways, a broad definition of *statistics* is the science of collecting, classifying, analyzing, and interpreting information. Thus, a statistician isn't just someone who calculates batting averages at baseball games or tabulates the results of a Gallup poll. Professional statisticians are trained in *statistical science*—that is, they are trained in collecting information in the form of **data,** evaluating it, and drawing conclusions from it. Furthermore, statisticians determine what information is relevant in a given problem and whether the conclusions drawn from a study are to be trusted.

> **Statistics** is the science of data. It involves collecting, classifying, summarizing, organizing, analyzing, and interpreting numerical and categorical information.

In the next section, you'll see several real-life examples of statistical applications in business and government that involve making decisions and drawing conclusions.

1.2 Types of Statistical Applications in Business

Statistics means "numerical descriptions" to most people. Monthly unemployment figures, the failure rate of startup companies, and the proportion of female executives in a particular industry all represent statistical descriptions of large sets of data collected on some phenomenon. Often the data are selected from some larger set of data that has characteristics we wish to estimate. We call this selection process *sampling*. For example,

BIOGRAPHY

FLORENCE NIGHTINGALE (1820–1910)

The Passionate Statistician

In Victorian England, the "Lady of the Lamp" had a mission to improve the squalid field hospital conditions of the British army during the Crimean War. Today, most historians consider Florence Nightingale to be the founder of the nursing profession. To convince members of the British Parliament of the need for supplying nursing and medical care to soldiers in the field, Nightingale compiled massive amounts of data from the army files. Through a remarkable series of graphs (which included the first "pie chart"), she demonstrated that most of the deaths in the war were due to illnesses contracted outside the battlefield or long after battle action from wounds that went untreated. Florence Nightingale's compassion and self-sacrificing nature, coupled with her ability to collect, arrange, and present large amounts of data, led some to call her the "Passionate Statistician."

you might collect the ages of a sample of customers of a video streaming services company to estimate the average age of *all* customers of the company. Then you could use your estimate to target the firm's advertisements to the appropriate age group. Notice that statistics involves two different processes: (1) describing sets of data and (2) drawing conclusions (making estimates, decisions, predictions, etc.) about the sets of data based on sampling. So, the applications of statistics can be divided into two broad areas: *descriptive statistics* and *inferential statistics*.

Descriptive statistics utilizes numerical and graphical methods to explore data, i.e., to look for patterns in a data set, to summarize the information revealed in a data set, and to present the information in a convenient form for the user.

Inferential statistics utilizes sample data to make estimates, decisions, predictions, or other generalizations about a larger set of data.

Although we'll discuss both descriptive and inferential statistics in the following chapters, the primary theme of the text is **inference.**

Let's begin by examining some business studies that illustrate applications of statistics.

Study 1.1 "Best-Selling Girl Scout Cookies" (Source: www.girlscouts.org): Since 1917, the Girl Scouts of America have been selling boxes of cookies. In 2017, there were 12 varieties for sale: Thin Mints, Samoas, Lemonades, Tagalongs, Do-si-dos, Trefoils, Savannah Smiles, Thanks-A-Lot, Dulce de Leche, Cranberry Citrus Crisps, Chocolate Chip, and Thank U Berry Much. Each of the approximately 150 million boxes of Girl Scout cookies sold in 2017 was classified by variety. The results are summarized in Figure 1.2. From the graph, you can clearly see that the best-selling variety is Thin Mints (25%), followed by Samoas (19%) and Tagalongs (13%). Since Figure 1.1 *describes* the variety of categories of the boxes of Girl Scout cookies sold, the graphic is an example of *descriptive statistics.*

Study 1.2 "Executive Compensation vs. Typical Worker Pay" (*Source*: 24/7 Wall Street, *USA Today*, October 15, 2018): How big is the gap between what a firm pays its CEO and what it pays its typical worker? To answer this question, 24/7 Wall Street reviewed the ratio between annual CEO base pay and typical worker salary at 168 large US companies, using data from benefits and compensation information provided by the website Payscale. This information was used to compute the ratio of CEO pay to the typical worker salary

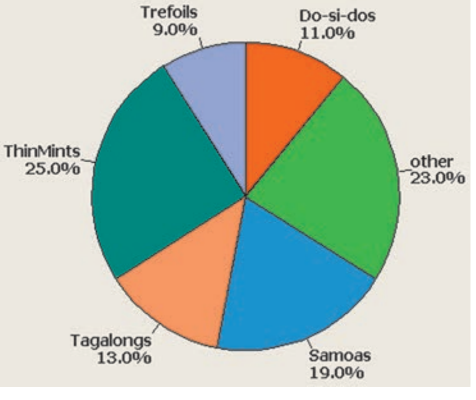

Figure 1.1
Best-selling Girl Scout cookies
Source: "Best-Selling Girl Scout Cookies," based on www.girlscouts.org.

at each company.* The data for the 10 companies with the highest ratio in the sample of 168 companies in the study are shown in Table 1.1. An analysis of the data for all 168 firms revealed that the "average" ratio of CEO pay to typical worker pay was 205.† In other words, on average, CEOs in the sample earn around 205 times what their firm's typical worker earns. Armed with this sample information, an economist might *infer* that the average ratio of CEO pay to typical worker pay for *all* US firms is 205. Thus, this study is an example of *inferential statistics*.

Table 1.1 Ratio of CEO Compensation to Typical Worker Pay

	Company	CEO	CEO Base Pay	CEO Total Compensation	Typical Worker Pay	Ratio
1	**CVS Health**	Larry J. Merlo	$12,105,481	$22,855,374	$27,900	434
2	**CBS Corp.**	Leslie Moonves	$23,652,883	$56,352,801	$59,900	395
3	**Walt Disney**	Robert A. Iger	$26,208,003	$43,490,567	$71,400	367
4	**TGX Comp.**	Carol Meyrowitz	$7,330,584	$17,962,232	$22,400	327
5	**21st Century Fox**	K. Rupert Murdoch	$17,047,636	$22,192,923	$54,800	311
6	**Comcast**	Brian L. Roberts	$16,819,942	$27,520,744	$55,800	301
7	**L Brands**	Leslie H. Wexner	$9,665,925	$26,669,306	$33,900	285
8	**Honeywell Int.**	David M. Cote	$22,767,851	$33,105,851	$81,600	279
9	**PepsiCo**	Indra K. Nooyi	15,937,828	$22,189,307	$61,500	259
10	**Wynn Resorts**	Stephen A. Wynn	$11,930,391	$20,680,391	$50,100	238

Source: 24/7 Wall Street, *USA Today*, Oct. 15, 2018.

Study 1.3 "Does rudeness really matter in the workplace?" (*Academy of Management Journal*, October 2007): Many studies have established that rudeness in the workplace can lead to retaliatory and counterproductive behavior. However, there has been little research on how rude behaviors influence a victim's task performance. In one study, college students enrolled in a management course were randomly assigned to one of two experimental conditions: rudeness condition (45 students) and control group (53 students). Each student was asked to write down as many uses for a brick as possible in 5 minutes; this value (total number of uses) was used as a performance measure for each student. For those students in the rudeness condition, the facilitator displayed rudeness by berating the students in general for being irresponsible and unprofessional (due to a late-arriving associate of the researchers). No comments were made about the late-arriving associate of the researchers to students in the control group. As you might expect, the researchers discovered that the performance levels for students in the rudeness condition were generally lower than the performance levels for students in the control group; thus, they concluded that rudeness in the workplace negatively affects job performance. As in Study 1.2, this study is an example of the use of inferential statistics. The researchers used data collected on 98 college students in a simulated work environment to make an inference about the performance levels of all workers exposed to rudeness on the job.

These studies provide three real-life examples of the uses of statistics in business, economics, and management. Notice that each involves an analysis of data, either for the purpose of describing the data set (Study 1.1) or for making inferences about a data set (Studies 1.2 and 1.3).

*The ratio was calculated using the *median* worker salary at each firm. A formal definition of median is given in Chapter 2. For now, think of the median as the *typical* salary for a worker, i.e., one that falls in the middle of all worker salaries.

†A formal definition of *average* is also given in Chapter 2. Like the median, think of the average as another way to express the *middle* salary.

1.3 Fundamental Elements of Statistics

Statistical methods are particularly useful for studying, analyzing, and learning about *populations* of *experimental units*.

An **experimental (or observational) unit** is an object (e.g., person, thing, transaction, or event) upon which we collect data.

A **population** is a set of units (usually people, objects, transactions, or events) that we are interested in studying.

For example, populations may include (1) *all* employed workers in the United States; (2) *all* registered voters in California; (3) *everyone* who has purchased a particular brand of cell phone; (4) *all* the cars produced last year by a particular assembly line; (5) the *entire* stock of spare parts at United Airlines' maintenance facility; (6) *all* sales made at the drive-through window of a McDonald's restaurant during a given year; and (7) the set of *all* accidents occurring on a particular stretch of interstate during a holiday period. Notice that the first three population examples (1–3) are sets (groups) of people, the next two (4–5) are sets of objects, the next (6) is a set of transactions, and the last (7) is a set of events. Also notice that *each set includes all the experimental units in the population* of interest.

In studying a population, we focus on one or more characteristics or properties of the experimental units in the population. We call such characteristics *variables*. For example, we may be interested in the variables age, gender, income, and/or the number of years of education of the people currently unemployed in the United States.

A **variable** is a characteristic or property of an individual experimental (or observational) unit.

The name *variable* is derived from the fact that any particular characteristic may vary among the experimental units in a population.

In studying a particular variable, it is helpful to be able to obtain a numerical representation for it. Often, however, numerical representations are not readily available, so the process of measurement plays an important supporting role in statistical studies. **Measurement** is the process we use to assign numbers to variables of individual population units. We might, for instance, measure the preference for a food product by asking a consumer to rate the product's taste on a scale from 1 to 10. Or we might measure workforce age by simply asking each worker, "How old are you?" In other cases, measurement involves the use of instruments such as timers, scales, and calipers.

If the population we wish to study is small, it is possible to measure a variable for every unit in the population. For example, if you are measuring the starting salary for all University of Michigan MBA graduates last year, it is at least feasible to obtain every salary. When we measure a variable for every experimental unit of a population, the result is called a **census** of the population. Typically, however, the populations of interest in most applications are much larger, involving perhaps many thousands or even an infinite number of units. Examples of large populations include the seven listed above, as well as all invoices produced in the last year by a *Fortune* 500 company, all potential buyers of a new iPad, and all stockholders of a firm listed on the New York

Stock Exchange. For such populations, conducting a census would be prohibitively time-consuming and/or costly. A reasonable alternative would be to select and study a *subset* (or portion) of the units in the population.

A **sample** is a subset of the units of a population.

For example, suppose a company is being audited for invoice errors. Instead of examining all 15,472 invoices produced by the company during a given year, an auditor may select and examine a sample of just 100 invoices (see Figure 1.2). If he is interested in the variable "invoice error status," he would record (measure) the status (error or no error) of each sampled invoice.

After the variable(s) of interest for every experimental unit in the sample (or population) is (are) measured, the data are analyzed, either by descriptive or by inferential statistical methods. The auditor, for example, may be interested only in *describing* the error rate in the sample of 100 invoices. More likely, however, he will want to use the information in the sample to make *inferences* about the population of all 15,472 invoices.

A **statistical inference** is an estimate or prediction or some other generalization about a population based on information contained in a sample.

*That is, we use the information contained in the sample to learn about the larger population.** Thus, from the sample of 100 invoices, the auditor may estimate the total number of invoices containing errors in the population of 15,472 invoices. The auditor's inference about the quality of the firm's invoices can be used in deciding whether to modify the firm's billing operations.

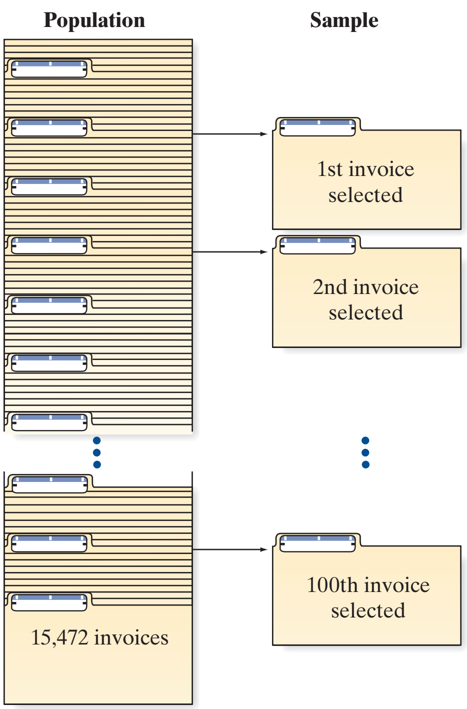

Figure 1.2
A sample of all company invoices

*The terms *population* and *sample* are often used to refer to the sets of measurements themselves, as well as to the units on which the measurements are made. When a single variable of interest is being measured, this usage causes little confusion. But when the terminology is ambiguous, we'll refer to the measurements as *population data sets* and *sample data sets*, respectively.

EXAMPLE 1.1

Key Elements of a Statistical Problem—Ages of Cable TV News Viewers

Problem According to the most recent Nielsen survey of cable TV news viewers, the average age of CNN viewers is 60 years. Suppose a rival network (e.g., FOX) executive hypothesizes that the average age of FOX viewers is greater than 60. To test her hypothesis, she samples 200 FOX viewers and determines the age of each.

a. Describe the population.
b. Describe the variable of interest.
c. Describe the sample.
d. Describe the inference.

Solution

a. The population is the set of units of interest to the TV executive, which is the set of all FOX viewers.
b. The age (in years) of each viewer is the variable of interest.
c. The sample must be a subset of the population. In this case, it is the 200 FOX viewers selected by the executive.
d. The inference of interest involves the *generalization* of the information contained in the sample of 200 viewers to the population of all FOX viewers. In particular, the executive wants to *estimate* the average age of the viewers in order to determine whether it exceeds 60 years. She might accomplish this by calculating the average age in the sample and using the sample average to estimate the population average.

Look Back A key to diagnosing a statistical problem is to identify the data set collected (in this example, the ages of the 200 FOX TV viewers) as a population or sample.

EXAMPLE 1.2

Key Elements of a Statistical Problem—Pepsi vs. Coca-Cola

Problem *Cola wars* is the popular term for the intense competition between Coca-Cola and Pepsi displayed in their marketing campaigns. Their campaigns have featured claims of consumer preference based on taste tests. For example, the *Huffington Post* (November 11, 2013) conducted a blind taste test of 9 cola brands that included Coca-Cola and Pepsi. (Pepsi finished 1st and Coke finished 5th.) Suppose, as part of a Pepsi marketing campaign, 1,000 cola consumers are given a blind taste test (i.e., a taste test in which the two brand names are disguised). Each consumer is asked to state a preference for brand A or brand B.

a. Describe the population.
b. Describe the variable of interest.
c. Describe the sample.
d. Describe the inference.

Solution

a. Because we are interested in the responses of cola consumers in a taste test, a cola consumer is the experimental unit. Thus, the population of interest is the collection or set of all cola consumers.
b. The characteristic that Pepsi wants to measure is the consumer's cola preference as revealed under the conditions of a blind taste test, so cola preference is the variable of interest.
c. The sample is the 1,000 cola consumers selected from the population of all cola consumers.
d. The inference of interest is the *generalization* of the cola preferences of the 1,000 sampled consumers to the population of all cola consumers. In particular, the

preferences of the consumers in the sample can be used to *estimate* the percentage of all cola consumers who prefer each brand.

Look Back In determining whether the statistical application is inferential or descriptive, we assess whether Pepsi is interested in the responses of only the 1,000 sampled customers (descriptive statistics) or in the responses for the entire population of consumers (inferential statistics).

Now Work Exercise 1.16

The preceding definitions and examples identify four of the five elements of an inferential statistical problem: a population, one or more variables of interest in a sample, and an inference. But making the inference is only part of the story. We also need to know its **reliability**–that is, how good the inference is. The only way we can be certain that an inference about a population is correct is to include the entire population in our sample. However, because of *resource constraints* (e.g., insufficient time and/or money), we usually can't work with whole populations, so we base our inferences on just a portion of the population (a sample). Consequently, whenever possible, it is important to determine and report the reliability of each inference made. Reliability, then, is the fifth element of inferential statistical problems.

The measure of reliability that accompanies an inference separates the science of statistics from the art of fortune-telling. A palm reader, like a statistician, may examine a sample (your hand) and make inferences about the population (your life). However, unlike statistical inferences, the palm reader's inferences include no measure of reliability.

Suppose, like the TV executive in Example 1.1, we are interested in the *error of estimation* (i.e., the difference between the average age of the population of TV viewers and the average age of a sample of TV viewers). Using statistical methods, we can determine a *bound on the estimation error*. This bound is simply a number that our estimation error (the difference between the average age of the sample and the average age of the population) is not likely to exceed. We'll see in later chapters that bound is a measure of the uncertainty of our inference. The reliability of statistical inferences is discussed throughout this text. For now, we simply want you to realize that an inference is incomplete without a measure of its reliability.

A **measure of reliability** is a statement (usually quantified) about the degree of uncertainty associated with a statistical inference.

Let's conclude this section with a summary of the elements of both descriptive and inferential statistical problems and an example to illustrate a measure of reliability.

Four Elements of Descriptive Statistical Problems

1. The population or sample of interest
2. One or more variables (characteristics of the population or experimental units) that are to be investigated
3. Tables, graphs, or numerical summary tools
4. Identification of patterns in the data

Five Elements of Inferential Statistical Problems

1. The population of interest
2. One or more variables (characteristics of the population or experimental units) that are to be investigated
3. The sample of population units
4. The inference about the population based on information contained in the sample
5. A measure of reliability for the inference

EXAMPLE 1.3

Reliability of an Inference—Pepsi vs. Coca-Cola

Problem Refer to Example 1.2, in which the cola preferences of 1,000 consumers were indicated in a taste test. Describe how the reliability of an inference concerning the preferences of all cola consumers in the Pepsi bottler's marketing region could be measured.

Solution When the preferences of 1,000 consumers are used to estimate the preferences of all consumers in the region, the estimate will not exactly mirror the preferences of the population. For example, if the taste test shows that 56% of the 1,000 consumers chose Pepsi, it does not follow (nor is it likely) that exactly 56% of all cola drinkers in the region prefer Pepsi. Nevertheless, we can use sound statistical reasoning (which is presented later in the text) to ensure that our sampling procedure will generate estimates that are almost certainly within a specified limit of the true percentage of all consumers who prefer Pepsi. For example, such reasoning might assure us that the estimate of the preference for Pepsi from the sample is almost certainly within 5% of the actual population preference. The implication is that the actual preference for Pepsi is between 51% [i.e., $(56 - 5)\%$] and 61% [i.e., $(56 + 5)\%$]—that is, $(56 \pm 5)\%$. This interval represents a measure of reliability for the inference.

Look Back The interval 56 ± 5 is called a *confidence interval*, because we are "confident" that the true percentage of customers who prefer Pepsi in a taste test falls into the range (51, 61). In Chapter 6, we learn how to assess the degree of confidence (e.g., 90% or 95% confidence) in the interval.

STATISTICS IN ACTION REVISITED

Identifying the Population, Sample, and Inference

Consider the study on the link between a CEO's golf handicap and the company's stock performance, reported in the *New York Times*. The newspaper gathered information on golf handicaps of corporate executives obtained from a *Golf Digest* survey sent to CEOs of the 300 largest US companies. (A golf handicap is a numerical "index" that allows golfers to compare skills; the lower the handicap, the better the golfer.) For the 51 CEOs who reported their handicaps, the *New York Times* then determined each CEO's company stock market performance over a 3-year period (measured as a rate-of-return index, from a low value of 0 to a high value of 100). Thus, the experimental unit for the study is a corporate executive, and the two variables measured are golf handicap and stock performance index. Also, the data for the 51 CEOs represent a sample selected from the much larger population of all corporate executives in the United States. (These data are available in the **GLFCEO** file.)

The *New York Times* discovered a "statistical correlation" (a method discussed in Chapter 11) between golf handicap and stock performance. Thus, the newspaper inferred that the better the CEO is at golf, the better the company's stock performance.

Data Set: GLFCEO

1.4 Processes (Optional)

Sections 1.2 and 1.3 focused on the use of statistical methods to analyze and learn about populations, which are sets of *existing* units. Statistical methods are equally useful for analyzing and making inferences about *processes*.

A **process** is a series of actions or operations that transforms inputs to outputs. A process produces or generates output over time.

The most obvious processes of interest to businesses are those of production or manufacturing. A manufacturing process uses a series of operations performed by

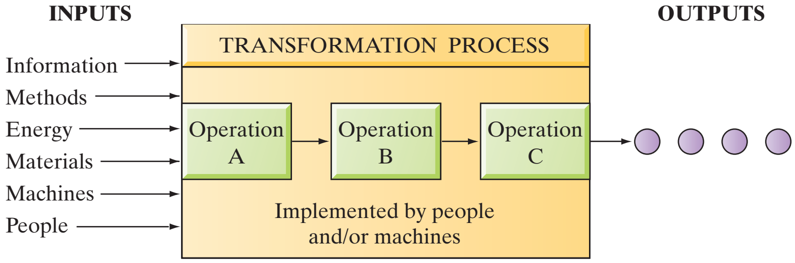

Figure 1.3
Graphical depiction of a manufacturing process

people and machines to convert inputs, such as raw materials and parts, to finished products (the outputs). Examples include the process used to produce the paper on which these words are printed, automobile assembly lines, and oil refineries.

Figure 1.3 presents a general description of a process and its inputs and outputs. In the context of manufacturing, the process in the figure (i.e., the transformation process) could be a depiction of the overall production process or it could be a depiction of one of the many processes (sometimes called *subprocesses*) that exist within an overall production process. Thus, the output shown could be finished goods that will be shipped to an external customer or merely the output of one of the steps or subprocesses of the overall process. In the latter case, the output becomes input for the next subprocess. For example, Figure 1.4 could represent the overall automobile assembly process, with its output being fully assembled cars ready for shipment to dealers. Or, it could depict the windshield assembly subprocess, with its output of partially assembled cars with windshields ready for "shipment" to the next subprocess in the assembly line.

Besides physical products and services, businesses and other organizations generate streams of numerical data over time that are used to evaluate the performance of the organization. Examples include weekly sales figures, quarterly earnings, and yearly profits. The US economy (a complex organization) can be thought of as generating streams of data that include the gross domestic product (GDP), stock prices, and the Consumer Price Index. Statisticians and other analysts conceptualize these data streams as being generated by processes. Typically, however, the series of operations or actions that cause particular data to be realized are either unknown or so complex (or both) that the processes are treated as *black boxes*.

A process whose operations or actions are unknown or unspecified is called a **black box.**

Frequently, when a process is treated as a black box, its inputs are not specified either. The entire focus is on the output of the process. A black box process is illustrated in Figure 1.4.

In studying a process, we generally focus on one or more characteristics, or properties, of the output. For example, we may be interested in the weight or the length of the units produced or even the time it takes to produce each unit. As with characteristics of population units, we call these characteristics *variables*. In studying processes whose output is already in numerical form (i.e., a stream of numbers), the characteristic, or property, represented by the numbers (e.g., sales, GDP, or stock prices) is typically the variable of interest. If the output is not numeric, we use *measurement processes* to assign

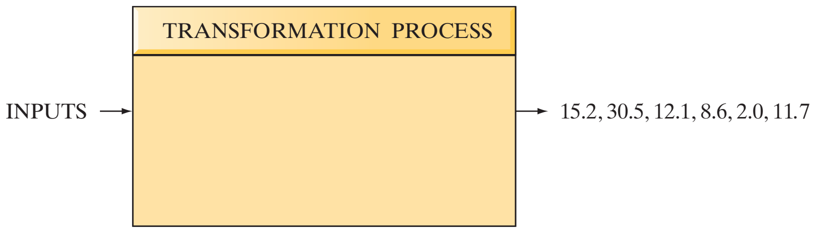

Figure 1.4
A black box process with numerical output