Essentials of

Business Statistics

Communicating with Numbers

Jaggia | Kelly

Second Edition

McGraw Hill Education

Essentials of
Business Statistics

The McGraw-Hill Education Series in Operations and Decision Sciences

Supply Chain Management

Benton
Purchasing and Supply Chain Management
Third Edition

Bowersox, Closs, Cooper, and Bowersox
Supply Chain Logistics Management
Fifth Edition

Burt, Petcavage, and Pinkerton
Supply Management
Eighth Edition

Johnson
Purchasing and Supply Management
Sixteenth Edition

Simchi-Levi, Kaminsky, and Simchi-Levi
Designing and Managing the Supply Chain: Concepts, Strategies, Case Studies
Third Edition

Stock and Manrodt
Fundamentals of Supply Chain Management

Project Management

Brown and Hyer
Managing Projects: A Team-Based Approach

Larson and Gray
Project Management: The Managerial Process
Seventh Edition

Service Operations Management

Bordoloi, Fitzsimmons, and Fitzsimmons
Service Management: Operations, Strategy, Information Technology
Ninth Edition

Management Science

Hillier and Hillier
Introduction to Management Science: A Modeling and Case Studies Approach with Spreadsheets
Sixth Edition

Business Research Methods

Cooper and Schindler
Business Research Methods
Thirteenth Edition

Business Forecasting

Keating, Wilson, and John Galt Solutions, Inc.
Business Forecasting
Seventh Edition

Linear Statistics and Regression

Kutner, Nachtsheim, and Neter
Applied Linear Regression Models
Fourth Edition

Business Systems Dynamics

Sterman
Business Dynamics: Systems Thinking and Modeling for a Complex World

Operations Management

Cachon and Terwiesch
Operations Management
Second Edition

Cachon and Terwiesch
Matching Supply with Demand: An Introduction to Operations Management
Fourth Edition

Jacobs and Chase
Operations and Supply Chain Management: The Core
Fifth Edition

Jacobs and Chase
Operations and Supply Chain Management
Fifteenth Edition

Schroeder and Goldstein
Operations Management in the Supply Chain: Decisions and Cases
Seventh Edition

Stevenson
Operations Management
Thirteenth Edition

Swink, Melnyk, Hartley, and Cooper
Managing Operations across the Supply Chain
Fourth Edition

Business Math

Slater and Wittry
Practical Business Math Procedures
Thirteenth Edition

Slater and Wittry
Math for Business and Finance: An Algebraic Approach
Second Edition

Business Statistics

Bowerman, O'Connell, and Murphree
Business Statistics in Practice
Ninth Edition

Doane and Seward
Applied Statistics in Business and Economics
Sixth Edition

Doane and Seward
Essential Statistics in Business and Economics
Third Edition

Jaggia and Kelly
Business Statistics: Communicating with Numbers
Third Edition

Jaggia and Kelly
Essentials of Business Statistics: Communicating with Numbers
Second Edition

Lind, Marchal, and Wathen
Basic Statistics for Business and Economics
Ninth Edition

Lind, Marchal, and Wathen
Statistical Techniques in Business and Economics
Seventeenth Edition

McGuckian
Connect Master: Business Statistics

2e

Essentials of Business Statistics

Communicating with Numbers

SANJIV JAGGIA
California Polytechnic State University

ALISON KELLY
Suffolk University

Mc Graw Hill Education

ESSENTIALS OF BUSINESS STATISTICS: COMMUNICATING WITH NUMBERS, SECOND EDITION

Published by McGraw-Hill Education, 2 Penn Plaza, New York, NY 10121.

This book is printed on acid-free paper.

1 2 3 4 5 6 7 8 9 LWI 21 20 19

ISBN 978-1-260-23951-5
MHID 1-260-23951-9

Portfolio Manager: *Noelle Bathurst*
Product Developers: *Ryan McAndrews*
Marketing Manager: *Harper Christopher*
Content Project Managers: *Pat Frederickson* and *Jamie Koch*
Buyer: *Laura Fuller*
Design: *Egzon Shaqiri*
Content Licensing Specialist: *Ann Marie Jannette*
Cover Design: *Beth Blech*
Compositor: *SPi Global*

All credits appearing on page or at the end of the book are considered to be an extension of the copyright page.

Library of Congress Cataloging-in-Publication Data

Names: Jaggia, Sanjiv, 1960- author. | Hawke, Alison Kelly, author.
Title: Essentials of business statistics : communicating with numbers/Sanjiv Jaggia, California Polytechnic State University, Alison Kelly, Suffolk University.
Description: Second Edition. | Dubuque : McGraw-Hill Education, [2018] | Revised edition of the authors' Essentials of business statistics, c2014.
Identifiers: LCCN 2018023099 | ISBN 9781260239515 (alk. paper)
Subjects: LCSH: Commercial statistics.
Classification: LCC HF1017 .J343 2018 | DDC 519.5-dc23
LC record available at https://lccn.loc.gov/2018023099

Dedicated to Chandrika, Minori, John, Megan, and Matthew

Sanjiv Jaggia

Courtesy of Sanjiv Jaggia

Sanjiv Jaggia is the associate dean of graduate programs and a professor of economics and finance at California Polytechnic State University in San Luis Obispo, California. After earning a Ph.D. from Indiana University, Bloomington, in 1990, Dr. Jaggia spent 17 years at Suffolk University, Boston. In 2003, he became a Chartered Financial Analyst (CFA®). Dr. Jaggia's research interests include empirical finance, statistics, and econometrics. He has published extensively in research journals, including the *Journal of Empirical Finance, Review of Economics and Statistics, Journal of Business and Economic Statistics, Journal of Applied Econometrics,* and *Journal of Econometrics.* Dr. Jaggia's ability to communicate in the classroom has been acknowledged by several teaching awards. In 2007, he traded one coast for the other and now lives in San Luis Obispo, California, with his wife and daughter. In his spare time, he enjoys cooking, hiking, and listening to a wide range of music.

Alison Kelly

Courtesy of Alison Kelly

Alison Kelly is a professor of economics at Suffolk University in Boston, Massachusetts. She received her B.A. degree from the College of the Holy Cross in Worcester, Massachusetts; her M.A. degree from the University of Southern California in Los Angeles; and her Ph.D. from Boston College in Chestnut Hill, Massachusetts. Dr. Kelly has published in journals such as the *American Journal of Agricultural Economics, Journal of Macroeconomics, Review of Income and Wealth, Applied Financial Economics,* and *Contemporary Economic Policy.* She is a Chartered Financial Analyst (CFA®) and teaches review courses in quantitative methods to candidates preparing to take the CFA exam. Dr. Kelly has also served as a consultant for a number of companies; her most recent work focused on how large financial institutions satisfy requirements mandated by the Dodd-Frank Act. She resides in Hamilton, Massachusetts, with her husband, daughter, and son.

A Unique Emphasis on Communicating with Numbers Makes Business Statistics Relevant to Students

We wrote *Essentials of Business Statistics: Communicating with Numbers* because we saw a need for a contemporary, core statistics text that sparked student interest and bridged the gap between how statistics is taught and how practitioners think about and apply statistical methods. Throughout the text, the emphasis is on communicating with numbers rather than on number crunching. In every chapter, students are exposed to statistical information conveyed in written form. By incorporating the perspective of practitioners, it has been our goal to make the subject matter more relevant and the presentation of material more straightforward for students. Although the text is application-oriented and practical, it is also mathematically sound and uses notation that is generally accepted for the topic being covered.

From our years of experience in the classroom, we have found that an effective way to make statistics interesting is to use timely applications. For these reasons, examples in *Essentials of Business Statistics* come from all walks of life, including business, economics, sports, health, housing, the environment, polling, and psychology. By carefully matching examples with statistical methods, students learn to appreciate the relevance of statistics in our world today, and perhaps, end up learning statistics without realizing they are doing so.

*This is probably the **best book** I have seen in terms of explaining concepts.*

Brad McDonald, *Northern Illinois University*

*The book is **well written, more readable and interesting than most stats texts,** and effective in explaining concepts. The examples and cases are particularly good and effective teaching tools.*

Andrew Koch, *James Madison University*

***Clarity and brevity are the most important things I look for**—this text has both in abundance.*

Michael Gordinier, *Washington University, St. Louis*

Continuing Key Features

The second edition of *Essentials of Business Statistics* reinforces and expands six core features that were well-received in the first edition.

Integrated Introductory Cases. Each chapter begins with an interesting and relevant introductory case. The case is threaded throughout the chapter, and once the relevant statistical tools have been covered, a synopsis—a short summary of findings—is provided. The introductory case often serves as the basis of several examples in other chapters.

Writing with Statistics. Interpreting results and conveying information effectively is critical to effective decision making in virtually every field of employment. Students are taught how to take the data, apply it, and convey the information in a meaningful way.

Unique Coverage of Regression Analysis. Relevant and extensive coverage of regression without repetition is an important hallmark of this text.

Written as Taught. Topics are presented the way they are taught in class, beginning with the intuition and explanation and concluding with the application.

Integration of Microsoft Excel®. Students are taught to develop an understanding of the concepts and how to derive the calculation; then Excel is used as a tool to perform the cumbersome calculations. In addition, guidelines for using Minitab, SPSS, JMP, and now R are provided in chapter appendices.

***Connect*®.** *Connect* is an online system that gives students the tools they need to be successful in the course. Through guided examples and LearnSmart adaptive study tools, students receive guidance and practice to help them master the topics.

I really like the case studies and the emphasis on writing. *We are making a big effort to incorporate more business writing in our core courses, so that meshes well.*

Elizabeth Haran, *Salem State University*

For a statistical analyst, your analytical skill is only as good as your communication skill. Writing with statistics reinforces the importance of communication *and provides students with concrete examples to follow.*

Jun Liu, *Georgia Southern University*

Features New to the Second Edition

The second edition of *Essentials of Business Statistics* features a number of improvements suggested by many reviewers and users of the first edition. The following are the major changes.

We focus on the *p*-Value Approach. We have found that students often get confused with the mechanics of implementing a hypothesis test using both the *p*-value approach and the critical value approach. While the critical value approach is attractive when a computer is unavailable and all calculations must be done by hand, most researchers and practitioners favor the *p*-value approach since virtually every statistical software package reports *p*-values. Our decision to focus on the *p*-value approach was further supported by recommendations set forth by the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report 2016* published by the American Statistical Association (http://www.amstat.org/asa/files/pdfs/GAISE/GaiseCollege_Full.pdf). *The GAISE Report* recommends that 'students should be able to interpret and draw conclusions from standard output from statistical software' (page 11) and that instructors should consider shifting away from the use of tables (page 23). Finally, we surveyed users of *Essentials of Business Statistics*, and they unanimously supported our decision to focus on the *p*-value approach. For those instructors interested in covering the critical value approach, it is discussed in the appendix to Chapter 9.

We added dozens of applied exercises with varying levels of difficulty. Many of these exercises include new data sets that encourage the use of the computer; however, just as many exercises retain the flexibility of traditional solving by hand.

We streamlined the Excel instructions. We feel that this modification provides a more seamless reinforcement for the relevant topic. For those instructors who prefer to omit the Excel parts so that they can use a different software, these sections can be easily skipped.

We completely revised Chapter 13 (More on Regression Analysis). Recognizing the importance of regression analysis in applied work, we have made major enhancements to Chapter 13. The chapter now contains the following sections: Dummy Variables, Interaction with Dummy Variables, Nonlinear Relationships, Trend Forecasting Models, and Forecasting with Trend and Seasonality.

In addition to the Minitab, SPSS, and JMP instructions that appear in chapter appendices, we now include instructions for R. The main reason for this addition is that R is an easy-to-use and wildly popular software that merges the convenience of statistical packages with the power of coding.

We reviewed every Connect exercise. Since both of us use Connect in our classes, we have attempted to make the technology component seamless with the text itself. In addition to reviewing every Connect exercise, we have added more conceptual exercises, evaluated rounding rules, and revised tolerance levels. The positive feedback from users of the first edition has been well worth the effort. We have also reviewed every LearnSmart probe. Instructors who teach in an online or hybrid environment will especially appreciate our Connect product.

Here are other noteworthy changes:

- For the sake of simplicity and consistency, we have streamlined or rewritten many Learning Outcomes.
- In Chapter 1 (Statistics and Data), we introduce structured data, unstructured data, and big data; we have also revised the section on online data sources.
- In Chapter 4 (Introduction to Probability), we examine marijuana legalization in the United States in the Writing with Statistics example.
- In Chapter 6 (Continuous Probability Distributions), we cover the normal distribution in one section, rather than two sections.
- In Chapter 7 (Sampling and Sampling Distributions), we added a discussion of the Trump election coupled with social-desirability bias.
- We have moved the section on "Model Assumptions and Common Violations" from Chapter 13 (More on Regression Analysis) to Chapter 12 (Basics of Regression Analysis).

Students Learn Through Real-World Cases and Business Examples . . .

Integrated Introductory Cases

Each chapter opens with a real-life case study that forms the basis for several examples within the chapter. The questions included in the examples create a roadmap for mastering the most important learning outcomes within the chapter. A synopsis of each chapter's introductory case is presented when the last of these examples has been discussed. Instructors of distance learners may find these introductory cases particularly useful.


©Mark Bowden/Getty Images

Introductory Case

Investment Decision

Jacqueline Brennan works as a financial advisor at a large investment firm. She meets with an inexperienced investor who has some questions regarding two approaches to mutual fund investing: growth investing versus value investing. The investor has heard that growth funds invest in companies whose stock prices are expected to grow at a faster rate, relative to the overall stock market, and value funds invest in companies whose stock prices are below their true worth. The investor has also heard that the main component of investment return is through capital appreciation in growth funds and through dividend income in value funds. The investor shows Jacqueline the annual return data for Vanguard's Growth Index mutual fund (henceforth, Growth) and Vanguard's Value Index mutual fund (henceforth, Value). Table 3.1 shows the annual return data for these two mutual funds for the years 2007–2016.

SYNOPSIS OF INTRODUCTORY CASE

Growth and value are two fundamental styles in stock and mutual fund investing. Proponents of growth investing believe that companies that are growing faster than their peers are trendsetters and will be able to maintain their superior growth. By investing in the stocks of these companies, they expect their investment to grow at a rate faster than the overall stock market. By comparison, value investors focus on the stocks of companies that are trading at a discount relative to the overall market or a specific sector. Investors of value stocks believe that these stocks are undervalued and that their price will increase once their true value is recognized by other investors. The debate between growth and value investing is age-old, and which style dominates depends on the sample period used for the analysis.


©Ingram Publishing/Getty Images

An analysis of annual return data for Vanguard's Growth Index mutual fund (Growth) and Vanguard's Value Index mutual fund (Value) for the years 2007 through 2016 provides important information for an investor trying to determine whether to invest in a growth mutual fund, a value mutual fund, or both types of mutual funds. Over this period, the mean return for the Growth fund of 10.09% is greater than the mean return for the Value fund of 7.56%. While the mean return typically represents the reward of investing, it does not incorporate the risk of

In all of these chapters, ***the opening case leads directly into the application questions*** *that students will have regarding the material. Having a strong and related case will certainly provide more benefit to the student, as context leads to improved learning.*

Alan Chow, *University of South Alabama*

This is an excellent approach. *The student gradually gets the idea that he can look at a problem—one which might be fairly complex—and break it down into root components. He learns that a little bit of math could go a long way, and even more math is even more beneficial to evaluating the problem.*

Dane Peterson, *Missouri State University*

and Build Skills to Communicate Results

Writing with Statistics

One of our most important innovations is the inclusion of a sample report within every chapter (except Chapter 1). Our intent is to show students how to convey statistical information in written form to those who may not know detailed statistical methods. For example, such a report may be needed as input for managerial decision making in sales, marketing, or company planning. Several similar writing exercises are provided at the end of each chapter. Each chapter also includes a synopsis that addresses questions raised from the introductory case. This serves as a shorter writing sample for students. Instructors of large sections may find these reports useful for incorporating writing into their statistics courses.

Writing with statistics shows that statistics ***is more than number crunching.***

Greg Cameron, *Brigham Young University*

These technical writing examples provide a ***very useful example of how to make statistics work and turn it into a report*** *that will be useful to an organization. I will strive to have my students learn from these examples.*

Bruce P. Christensen, *Weber State University*

This is an ***excellent approach. . . .*** *The ability to translate numerical information into words that others can understand is critical.*

Scott Bailey, *Troy University*

Excellent. *Students need to become better writers.*

Bob Nauss, *University of Missouri, St. Louis*

WRITING WITH STATISTICS

Professor Lang is a professor of economics at Salem State University. She has been teaching a course in Principles of Economics for over 25 years. Professor Lang has never graded on a curve since she believes that relative grading may unduly penalize (benefit) a good (poor) student in an unusually strong (weak) class. She always uses an absolute scale for making grades, as shown in the two left columns of Table 6.5.

©Image Source, all rights reserved.

TABLE 6.5 Grading Scales with Absolute Grading versus Relative Grading

Absolute Grading		Relative Grading	
Grade	Score	Grade	Probability
A	92 and above	A	0.10
B	78 up to 92	B	0.35
C	64 up to 78	C	0.40
D	58 up to 64	D	0.10
F	Below 58	F	0.05

A colleague of Professor Lang's has convinced her to move to relative grading, since it corrects for unanticipated problems. Professor Lang decides to experiment with grading based on the relative scale as shown in the two right columns of Table 6.5. Using this relative grading scheme, the top 10% of students will get A's, the next 35% B's, and so on. Based on her years of teaching experience, Professor Lang believes that the scores in her course follow a normal distribution with a mean of 78.6 and a standard deviation of 12.4.

Professor Lang wants to use the above information to

1. Calculate probabilities based on the absolute scale. Compare these probabilities to the relative scale.
2. Calculate the range of scores for various grades based on the relative scale. Compare these ranges to the absolute scale.
3. Determine which grading scale makes it harder to get higher grades.

Sample Report— Absolute Grading versus Relative Grading

Many teachers would confess that grading is one of the most difficult tasks of their profession. Two common grading systems used in higher education are relative and absolute. Relative grading systems are norm-referenced or curve-based, in which a grade is based on the student's relative position in class. Absolute grading systems, on the other hand, are criterion-referenced, in which a grade is related to the student's absolute performance in class. In short, with absolute grading, the student's score is compared to a predetermined scale, whereas with relative grading, the score is compared to the scores of other students in the class.

Let X represent a grade in Professor Lang's class, which is normally distributed with a mean of 78.6 and a standard deviation of 12.4. This information is used to derive the grade probabilities based on the absolute scale. For instance, the probability of receiving an A is derived as $P(X \geq 92) = P(Z \geq 1.08) = 0.14$. Other probabilities, derived similarly, are presented in Table 6.A.

TABLE 6.A Probabilities Based on Absolute Scale and Relative Scale

Grade	Probability Based on Absolute Scale	Probability Based on Relative Scale
A	0.14	0.10
B	0.38	0.35
C	0.36	0.40
D	0.07	0.10
F	0.05	0.05

Unique Coverage and Presentation . . .

By comparing this chapter with other books, I think that this is one of the best explanations about regression I have seen.

Cecilia Maldonado, *Georgia Southwestern State University*

Unique Coverage of Regression Analysis

We combine simple and multiple regression in one chapter, which we believe is a seamless grouping and eliminates needless repetition. This grouping allows more coverage of regression analysis than the vast majority of *Essentials* texts. This focus reflects the topic's growing use in practice. However, for those instructors who prefer to cover only simple regression, doing so is still an option.

The authors have put forth a novel and innovative way to present regression which in and of itself should make instructors take a long and hard look at this book. Students should find this book very readable and a good companion for their course.

Harvey A. Singer, *George Mason University*

This is easy for students to follow and I do get the feeling . . . the sections are spoken language.

Zhen Zhu, *University of Central Oklahoma*

Written as Taught

We introduce topics just the way we teach them; that is, the relevant tools follow the opening application. Our roadmap for solving problems is

1. Start with intuition
2. Introduce mathematical rigor, and
3. Produce computer output that confirms results.

We use worked examples throughout the text to illustrate how to apply concepts to solve real-world problems.

that Make the Content More Effective

Integration of Microsoft Excel®

We prefer that students first focus on and absorb the statistical material before replicating their results with a computer. Solving each application manually provides students with a deeper understanding of the relevant concept. However, we recognize that, primarily due to cumbersome calculations or the need for statistical tables, embedding computer output is necessary. Microsoft Excel is the primary software package used in this text. We chose Excel over other statistical packages based on reviewer feedback and the fact that students benefit from the added spreadsheet experience. We provide instructions for using Minitab, SPSS, JMP, and R in chapter appendices.

Using Excel to Obtain Binomial Probabilities

We use Excel's **BINOM.DIST** function to calculate binomial probabilities. In order to find $P(X = x)$, we enter "=BINOM.DIST(x, n, p, 0)" where x is the number of successes, n is the number of trials, and p is the probability of success. If we enter a "1" for the last argument in the function, then Excel returns $P(X \leq x)$.

a. In order to find the probability that exactly 70 American adults are Facebook users, $P(X = 70)$, we enter "=BINOM.DIST(70, 100, 0.68, 0)" and Excel returns 0.0791.

b. In order to find the probability that no more than 70 American adults are Facebook users, $P(X \leq 70)$, we enter "=BINOM.DIST(70, 100, 0.68, 1)" and Excel returns 0.7007.

c. In order to find the probability that at least 70 American adults are Facebook users, $P(X \geq 70) = 1 - P(X \leq 69)$, we enter "=1−BINOM.DIST(69, 100, 0.68, 1)" and Excel returns 0.3784.

*. . . does a solid job of building the intuition behind the concepts and **then adding mathematical rigor to these ideas** before finally verifying the results with Excel.*

Matthew Dean, University of Southern Maine

Real-World Exercises and Case Studies that Reinforce the Material

Mechanical and Applied Exercises

Chapter exercises are a well-balanced blend of mechanical, computational-type problems followed by more ambitious, interpretive-type problems. We have found that simpler drill problems tend to build students' confidence prior to tackling more difficult applied problems. Moreover, we repeatedly use many data sets—including house prices, rents, stock returns, salaries, and debt—in various chapters of the text. For instance, students first use these real data to calculate summary measures, make statistical inferences with confidence intervals and hypothesis tests, and finally, perform regression analysis.

Applied exercises from *The Wall Street Journal, Kiplinger's, Fortune, The New York Times, USA Today;* various websites—Census.gov, Zillow.com, Finance.yahoo.com, ESPN.com; and more.

18. Consider the following hypothesis test:

$$H_0: \mu \leq -5$$
$$H_A: \mu > -5$$

A random sample of 50 observations yields a sample mean ...3. The population standard deviation is 10. Calculate the ...alue. What is the conclusion to the test if $\alpha = 0.05$?

...nsider the following hypothesis test:

$$H_0: \mu \leq 75$$
$$H_A: \mu > 75$$

A random sample of 100 observations yields a sample mean of 80. The population standard deviation is 30. Calculate the p-value. What is the conclusion to the test if $\alpha = 0.10$?

20. Consider the following hypothesis test:

$$H_0: \mu = -100$$
$$H_A: \mu \neq -100$$

A random sample of 36 observations yields a sample mean of −125. The population standard deviation is 42. Conduct the test at $\alpha = 0.01$.

21. Consider the following hypotheses:

$$H_0: \mu = 120$$
$$H_A: \mu \neq 120$$

The population is normally distributed with a population standard deviation of 46.

a. If $\bar{x} = 132$ and $n = 50$, what is the conclusion at the 5% significance level?

b. If $\bar{x} = 108$ and $n = 50$, what is the conclusion at the 10% significance level?

22. FILE ***Excel_1.*** Given the accompanying sample data, use Excel's formula options to determine if the population mean is less than 125 at the 5% significance level. Assume that the population is normally distributed and that the population standard deviation equals 12.

23. FILE ***Excel_2.*** Given the accompanying sample data, use

25. Customers at Costco spend an average of $130 per trip (*The Wall Street Journal,* October 6, 2010). One of Costco's rivals would like to determine whether its customers spend more per trip. A survey of the receipts of 25 customers found that the sample mean was $135.25. Assume that the population standard deviation is $10.50 and that spending follows a normal distribution.

a. Specify the null and alternative hypotheses to test whether average spending at the rival's store is more than $130.

b. Calculate the value of the test statistic and the p-value.

c. At the 5% significance level, what is the conclusion to the test?

26. In May 2008, CNN reported that sports utility vehicles (SUVs) are plunging toward the "endangered" list. Due to the uncertainty of oil prices and environmental concerns, consumers are replacing gas-guzzling vehicles with fuel-efficient smaller cars. As a result, there has been a big drop in the demand for new as well as used SUVs. A sales manager of a used car dealership for SUVs believes that it takes more than 90 days, on average, to sell an SUV. In order to test his claim, he samples 40 recently sold SUVs and finds that it took an average of 95 days to sell an SUV. He believes that the population standard deviation is fairly stable at 20 days.

a. State the null and the alternative hypotheses for the test.

b. What is the p-value?

c. Is the sales manager's claim justified at $\alpha = 0.01$?

27. According to the *Centers for Disease Control and Prevention* (February 18, 2016), 1 in 3 American adults do not get enough sleep. A researcher wants to determine if Americans are sleeping less than the recommended 7 hours of sleep on weekdays. He takes a random sample of 150 Americans and computes the average sleep time of 6.7 hours on weekdays. Assume that the population is normally distributed with a known standard deviation of 2.1 hours. Test the researcher's claim at $\alpha = 0.01$.

I especially like the introductory cases, the ***quality of the end-of-section problems,*** *and the writing examples.*

Dave Leupp, *University of Colorado at Colorado Springs*

Their exercises and problems are excellent!

Erl Sorensen, *Bentley University*

Features that Go Beyond the Typical

Conceptual Review

At the end of each chapter, we present a conceptual review that provides a more holistic approach to reviewing the material. This section revisits the learning outcomes and provides the most important definitions, interpretations, and formulas.

CONCEPTUAL REVIEW

LO 5.1 Describe a discrete random variable and its probability distribution.

A **random variable** summarizes outcomes of an experiment with numerical values. A **discrete random variable** assumes a countable number of distinct values, whereas a **continuous random variable** is characterized by uncountable values in an interval.

The **probability mass function** for a discrete random variable X is a list of the values of X with the associated probabilities; that is, the list of all possible pairs $(x, P(X = x))$. The **cumulative distribution function** of X is defined as $P(X \leq x)$.

LO 5.2 Calculate and interpret summary measures for a discrete random variable.

For a discrete random variable X with values $x_1, x_2, x_3, \ldots$, which occur with probabilities $P(X = x_i)$, the **expected value** of X is calculated as $E(X) = \mu = \Sigma x_i P(X = x_i)$. We interpret the expected value as the long-run average value of the random variable over infinitely many independent repetitions of an experiment. Measures of dispersion indicate whether the values of X are clustered about μ or widely scattered from μ. The variance of X is calculated as $Var(X) = \sigma^2 = \Sigma(x_i - \mu)^2 P(X = x_i)$. The standard deviation of X is $SD(X) = \sigma = \sqrt{\sigma^2}$.

In general, a **risk-averse consumer** expects a reward for taking risk. A risk-averse consumer may decline a risky prospect even if it offers a positive expected gain. A **risk-neutral consumer** completely ignores risk and always accepts a prospect that offers a positive expected gain.

LO 5.3 Calculate and interpret probabilities for a binomial random variable.

A **Bernoulli process** is a series of n independent and identical trials of an experiment such that on each trial there are only two possible outcomes, conventionally labeled "success" and "failure." The probabilities of success and failure, denoted p and $1 - p$, remain the same from trial to trial.

For a **binomial random variable** X, the probability of x successes in n Bernoulli trials is $P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1 - p)^{n-x}$ for $x = 0, 1, 2, \ldots, n$.

The expected value, the variance, and the standard deviation of a binomial random variable are $E(X) = np$, $Var(X) = \sigma^2 = np(1 - p)$, and $SD(X) = \sigma = \sqrt{np(1 - p)}$, respectively.

Most texts basically list what one should have learned but don't add much to that. You do a good job of reminding the reader of what was covered and what was most important about it.

Andrew Koch, *James Madison University*

They have gone beyond the typical [summarizing formulas] and I like the structure. This is a very strong feature of this text.

Virginia M. Miori, *St. Joseph's University*

Students—study more efficiently, retain more and achieve better outcomes. Instructors—focus on what you love—teaching.

SUCCESSFUL SEMESTERS INCLUDE CONNECT

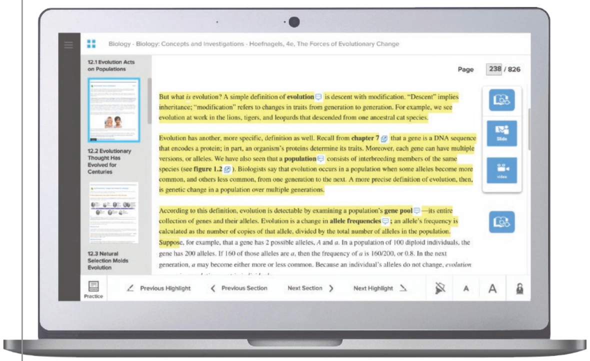
FOR INSTRUCTORS

You're in the driver's seat.

Want to build your own course? No problem. Prefer to use our turnkey, prebuilt course? Easy. Want to make changes throughout the semester? Sure. And you'll save time with Connect's auto-grading too.

65%

Less Time Grading



They'll thank you for it.

Adaptive study resources like SmartBook® help your students be better prepared in less time. You can transform your class time from dull definitions to dynamic debates. Hear from your peers about the benefits of Connect at **www.mheducation.com/highered/connect**

Make it simple, make it affordable.

Connect makes it easy with seamless integration using any of the major Learning Management Systems—Blackboard®, Canvas, and D2L, among others—to let you organize your course in one convenient location. Give your students access to digital materials at a discount with our inclusive access program. Ask your McGraw-Hill representative for more information.

Solutions for your challenges.

A product isn't a solution. Real solutions are affordable, reliable, and come with training and ongoing support when you need it and how you want it. Our Customer Experience Group can also help you troubleshoot tech problems—although Connect's 99% uptime means you might not need to call them. See for yourself at **status.mheducation.com**

For Students

Effective, efficient studying.

Connect helps you be more productive with your study time and get better grades using tools like SmartBook, which highlights key concepts and creates a personalized study plan. Connect sets you up for success, so you walk into class with confidence and walk out with better grades.

©Shutterstock/wavebreakmedia

"I really liked this app—it made it easy to study when you don't have your text-book in front of you."

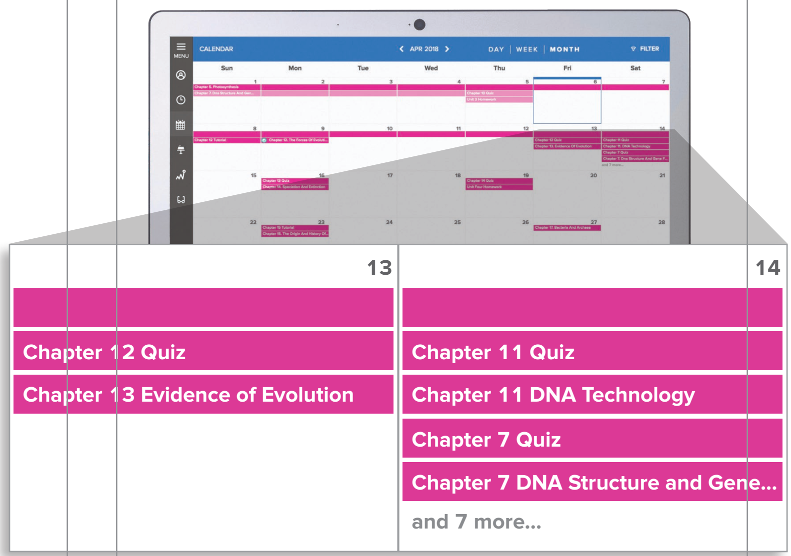
- Jordan Cunningham,
Eastern Washington University

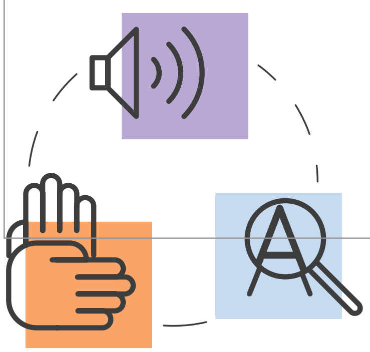
Study anytime, anywhere.

Download the free ReadAnywhere app and access your online eBook when it's convenient, even if you're offline. And since the app automatically syncs with your eBook in Connect, all of your notes are available every time you open it. Find out more at **www.mheducation.com/readanywhere**

No surprises.

The Connect Calendar and Reports tools keep you on track with the work you need to get done and your assignment scores. Life gets busy; Connect tools help you keep learning through it all.





Learning for everyone.

McGraw-Hill works directly with Accessibility Services Departments and faculty to meet the learning needs of all students. Please contact your Accessibility Services office and ask them to email accessibility@mheducation.com, or visit **www.mheducation.com/accessibility** for more information.

What Resources are Available for Instructors?

Instructor Library

The *Connect* Instructor Library is your repository for additional resources to improve student engagement in and out of class. You can select and use any asset that enhances your lecture. The *Connect* Instructor Library includes:

- PowerPoint presentations
- Excel Data Files
- Test Bank
- Instructor's Solutions Manual
- Digital Image Library

Tegrity Campus: Lectures 24/7

Tegrity Campus is integrated in *Connect* to help make your class time available 24/7. With Tegrity, you can capture each one of your lectures in a searchable format for students to review when they study and complete assignments using *Connect.* With a simple one-click start-and-stop process, you can capture everything that is presented to students during your lecture from your computer, including audio. Students can replay any part of any class with easy-to-use browser-based viewing on a PC or Mac.

Educators know that the more students can see, hear, and experience class resources, the better they learn. In fact, studies prove it. With *Tegrity Campus,* students quickly recall key moments by using *Tegrity Campus*'s unique search feature. This search helps students efficiently find what they need, when they need it, across an entire semester of class recordings. Help turn all your students' study time into learning moments immediately supported by your lecture. To learn more about *Tegrity,* watch a two-minute Flash demo at http://tegritycampus.mhhe.com.

ALEKS

ALEKS is an assessment and learning program that provides individualized instruction in Business Statistics, Business Math, and Accounting. Available online in partnership with McGraw-Hill Education, ALEKS interacts with students much like a skilled human tutor, with the ability to assess precisely a student's knowledge and provide instruction on the exact topics the student is most ready to learn. By providing topics to meet individual students' needs, allowing students to move between explanation and practice, correcting and analyzing errors, and defining terms, ALEKS helps students to master course content quickly and easily.

ALEKS also includes an instructor module with powerful, assignment-driven features and extensive content flexibility. ALEKS simplifies course management and allows instructors to spend less time with administrative tasks and more time directing student learning. To learn more about ALEKS, visit www.aleks.com.

MegaStat® for Microsoft Excel®

MegaStat® by J. B. Orris of Butler University is a full-featured Excel add-in that is available online through the *MegaStat* website at **www.mhhe.com/megastat** or through an access card packaged with the text. It works with Excel 2016, 2013, and 2010 (and Excel: Mac 2016). On the website, students have 10 days to successfully download and install *MegaStat* on their local computer. Once installed, *MegaStat* will remain active in Excel with no expiration date or time limitations. The software performs statistical analyses within an Excel workbook. It does basic functions, such as descriptive statistics, frequency distributions, and probability calculations, as well as hypothesis testing, ANOVA, and regression. *MegaStat* output is carefully formatted, and its ease-of-use features include Auto Expand for quick data selection and Auto Label detect. Since *MegaStat* is easy to use, students can focus on learning statistics without being distracted by the software. *MegaStat* is always available from Excel's main menu. Selecting a menu item pops up a dialog box. Screencam tutorials are included that provide a walkthrough of major business statistics topics. Help files are built in, and an introductory user's manual is also included.

What Resources are Available for Students?

Integration of Excel Data Sets. A convenient feature is the inclusion of an Excel data file link in many problems using data files in their calculation. The link allows students to easily launch into Excel, work the problem, and return to *Connect* to key in the answer and receive feedback on their results.

deviation is \$100 (in \$1,000s). What is the value of the test statistic and the p-value?

c. At $\alpha = 0.05$, what is the conclusion to the test? Is the realtor's claim supported by the data?

30. FILE ***Home_Depot.*** The data accompanying this exercise show the weekly stock price for Home Depot. Assume that stock prices are normally distributed with a population standard deviation of \$3.
    a. State the null and the alternative hypotheses in order to test whether or not the average weekly stock price differs from \$30.
    b. Find the value of the test statistic and the p-value.
    c. At $\alpha = 0.05$, can you conclude that the average weekly stock price does not equal \$30?

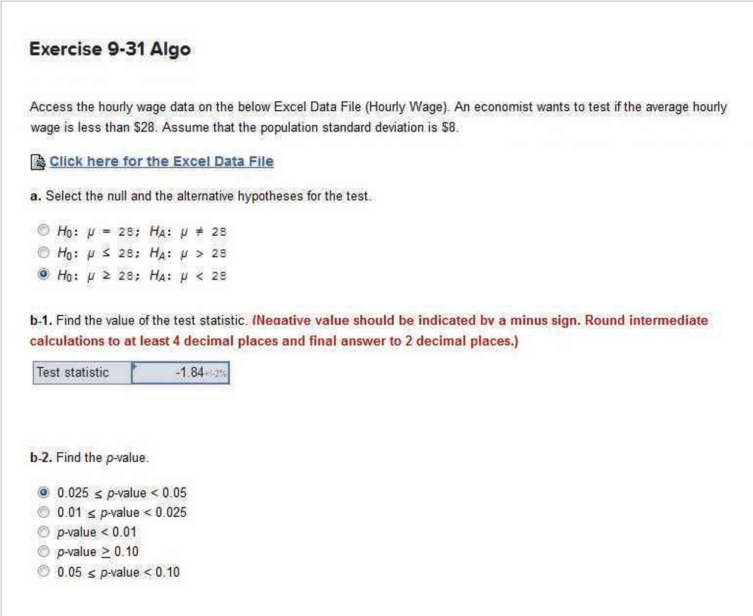
31. FILE ***Hourly_Wage.*** An economist wants to test if the average hourly wage is less than \$22. Assume that the population standard deviation is \$6.
    a. State the null and the alternative hypotheses for the test.
    b. The data accompanying this exercise show hourly wages. Find the value of the test statistic and the p-value.
    c. At $\alpha = 0.05$, what is the conclusion to the test? Is the average hourly wage less than \$22?

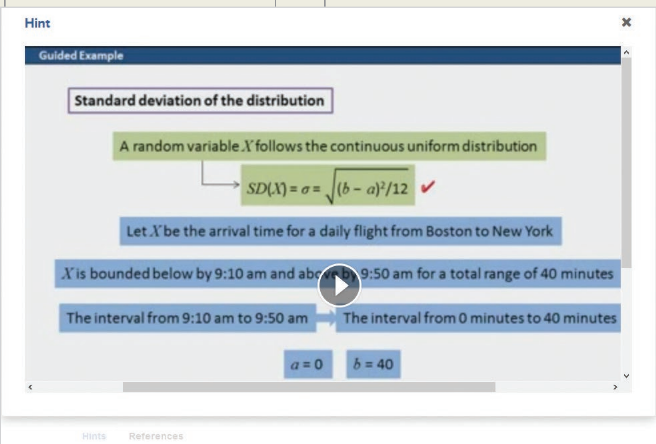
32. FILE ***CT_Undergrad_Debt.*** On average, a college student graduates with \$27,200 in debt (*The Boston Globe*, May 27, 2012). The data accompanying this exercise show the debt for 40 recent undergraduates from Connecticut. Assume that the population standard deviation is \$5,000.
    a. A researcher believes that recent undergraduates from Connecticut have less debt than the national average. Specify the competing hypotheses to test this belief.
    b. Find the value of the test statistic and the p-value.
    c. Do the data support the researcher's claim, at $\alpha = 0.10$?

15. FILE ***CT_Undergrad_Debt.*** A study reports that recent college graduates from New Hampshire face the highest average debt of \$31,048 (*The Boston Globe,* May 27, 2012). A researcher from Connecticut wants to determine how recent undergraduates from that state fare. He collects data on debt from 40 recent undergraduates. A portion of the data is shown in the accompanying table. Assume that the population standard deviation is \$5,000.

Debt
24040
19153
⋮
29329

    a. Construct the 95% confidence interval for the mean debt of all undergraduates from Connecticut.
    b. Use the 95% confidence interval to determine if the debt of Connecticut undergraduates differs from that of New Hampshire undergraduates.





Guided Examples. These narrated video walkthroughs provide students with step-by-step guidelines for solving selected exercises similar to those contained in the text. The student is given personalized instruction on how to solve a problem by applying the concepts presented in the chapter. The video shows the steps to take to work through an exercise. Students can go through each example multiple times if needed.

The *Connect* Student Resource page is the place for students to access additional resources. The Student Resource page offers students quick access to the recommended study tools, data files, and helpful tutorials on statistical programs.

McGraw-Hill Customer Care Contact Information

At McGraw-Hill, we understand that getting the most from new technology can be challenging. That's why our services don't stop after you purchase our products. You can e-mail our product specialists 24 hours a day to get product training online. Or you can search our knowledge bank of frequently asked questions on our support website.

For customer support, call **800-331-5094** or visit **www.mhhe.com/support**. One of our technical support analysts will be able to assist you in a timely fashion.

ACKNOWLEDGMENTS

We would like to acknowledge the following people for providing useful comments and suggestions for past and present editions of all aspects of *Business Statistics.*

John Affisco
Hofstra University
Mehdi Afiat
College of Southern Nevada
Mohammad Ahmadi
University of Tennessee—Chattanooga
Sung Ahn
Washington State University
Mohammad Ahsanullah
Rider University
Imam Alam
University of Northern Iowa
Mostafa Aminzadeh
Towson University
Ardavan Asef-Vaziri
California State University
Antenah Ayanso
Brock University
Scott Bailey
Troy University
Jayanta Bandyopadhyay
Central Michigan University
Samir Barman
University of Oklahoma
Douglas Barrett
University of North Alabama
John Beyers
University of Maryland
Arnab Bisi
Purdue University—West Lafayette
Gary Black
University of Southern Indiana
Randy Boan
Aims Community College
Matthew Bognar
University of Iowa
Juan Cabrera
Ramapo College of New Jersey
Scott Callan
Bentley University
Gregory Cameron
Brigham Young University
Kathleen Campbell
St. Joseph's University
Alan Cannon
University of Texas—Arlington
Michael Cervetti
University of Memphis
Samathy Chandrashekar
Salisbury University
Gary Huaite Chao
University of Pennsylvania—Kutztown
Sangit Chatterjee
Northeastern University
Leida Chen
California Polytechnic State University
Anna Chernobai
Syracuse University
Alan Chesen
Wright State University
Juyan Cho
Colorado State University—Pueblo
Alan Chow
University of South Alabama
Bruce Christensen
Weber State University
Howard Clayton
Auburn University
Robert Collins
Marquette University
M. Halim Dalgin
Kutztown University
Tom Davis
University of Dayton
Matthew Dean
University of Maine
Jason Delaney
University of Arkansas—Little Rock
Ferdinand DiFurio
Tennessee Tech University
Matt Dobra
UMUC
Luca Donno
University of Miami
Joan Donohue
University of South Carolina
David Doorn
University of Minnesota
James Dunne
University of Dayton
Mike Easley
University of New Orleans
Erick Elder
University of Arkansas—Little Rock
Ashraf ElHoubi
Lamar University
Roman Erenshteyn
Goldey-Beacom College
Grace Esimai
University of Texas—Arlington
Soheila Fardanesh
Towson University
Carol Flannery
University of Texas—Dallas
Sydney Fletcher
Mississippi Gulf Coast Community College
Andrew Flight
Portland State University
Samuel Frame
Cal Poly San Luis Obispo
Priya Francisco
Purdue University
Vickie Fry
Westmoreland County Community College
Ed Gallo
Sinclair Community College
Glenn Gilbreath
Virginia Commonwealth University
Robert Gillette
University of Kentucky
Xiaoning Gilliam
Texas Tech University
Mark Gius
Quinnipiac University
Malcolm Gold
Saint Mary's University of Minnesota
Michael Gordinier
Washington University

Deborah Gougeon
University of Scranton
Don Gren
Salt Lake Community College
Thomas G. Groleau
Carthage College
Babita Gupta
CSU Monterey Bay
Robert Hammond
North Carolina State University
Jim Han
Florida Atlantic University
Elizabeth Haran
Salem State University
Jack Harshbarger
Montreat College
Edward Hartono
University of Alabama—Huntsville
Clifford Hawley
West Virginia University
Santhi Heejebu
Cornell College
Paul Hong
University of Toledo
Ping-Hung Hsieh
Oregon State University
Marc Isaacson
Augsburg College
Mohammad Jamal
Northern Virginia Community College
Robin James
Harper College
Molly Jensen
University of Arkansas
Craig Johnson
Brigham Young University—Idaho
Janine Sanders Jones
University of St. Thomas
Vivian Jones
Bethune—Cookman University
Jerzy Kamburowski
University of Toledo
Howard Kaplon
Towson University
Krishna Kasibhatla
North Carolina A&T State University
Mohammad Kazemi
University of North Carolina—Charlotte
Ken Kelley
University of Notre Dame
Lara Khansa
Virginia Tech
Esther C. Klein
St. Francis College
Ronald Klimberg
St. Joseph's University
Andrew Koch
James Madison University
Subhash Kochar
Portland State University
Brandon Koford
Weber University
Randy Kolb
St. Cloud State University
Vadim Kutsyy
San Jose State University
Francis Laatsch
University of Southern Mississippi
David Larson
University of South Alabama
John Lawrence
California State University—Fullerton
Shari Lawrence
Nicholls State University
Radu Lazar
University of Maryland
David Leupp
University of Colorado—Colorado Springs
Carel Ligeon
Auburn University—Montgomery
Carin Lightner
North Carolina A&T State University
Constance Lightner
Fayetteville State University
Scott Lindsey
Dixie State College of Utah
Ken Linna
Auburn University—Montgomery
Andy Litteral
University of Richmond
Jun Liu
Georgia Southern University
Chung-Ping Loh
University of North Florida
Salvador Lopez
University of West Georgia
John Loucks
St. Edward's University
Cecilia Maldonado
Georgia Southwestern State University
Farooq Malik
University of Southern Mississippi
Ken Mayer
University of Nebraska—Omaha
Bradley McDonald
Northern Illinois University
Elaine McGivern
Duquesne University
John McKenzie
Babson University
Norbert Michel
Nicholls State University
John Miller
Sam Houston State University
Virginia Miori
St. Joseph's University
Prakash Mirchandani
University of Pittsburgh
Jason Molitierno
Sacred Heart University
Elizabeth Moliski
University of Texas—Austin
Joseph Mollick
Texas A&M University—Corpus Christi
James Moran
Oregon State University
Khosrow Moshirvaziri
California State University—Long Beach
Tariq Mughal
University of Utah
Patricia Mullins
University of Wisconsin—Madison
Kusum Mundra
Rutgers University—Newark
Anthony Narsing
Macon State College
Robert Nauss
University of Missouri—St. Louis
Satish Nayak
University of Missouri—St. Louis
Thang Nguyen
California State University—Long Beach
Mohammad Oskoorouchi
California State University—San Marcos
Barb Osyk
University of Akron
Bhavik Pathak
Indiana University South Bend
Scott Paulsen
Illinois Central College
James Payne
Calhoun Community College
Norman Pence
Metropolitan State College of Denver

Dane Peterson
Missouri State University
Joseph Petry
University of Illinois—Urbana/Champaign
Courtney Pham
Missouri State University
Martha Pilcher
University of Washington
Cathy Poliak
University of Wisconsin—Milwaukee
Simcha Pollack
St. John's University
Hamid Pourmohammadi
California State University—Dominguez Hills
Tammy Prater
Alabama State University
Zbigniew H. Przasnyski
Loyola Marymount University
Manying Qiu
Virginia State University
Troy Quast
Sam Houston State University
Michael Racer
University of Memphis
Srikant Raghavan
Lawrence Technological University
Bharatendra Rai
University of Massachusetts—Dartmouth
Michael Aaron Ratajczyk
Saint Mary's University of Minnesota
Tony Ratcliffe
James Madison University
David Ravetch
University of California
Bruce Reinig
San Diego State University
Darlene Riedemann
Eastern Illinois University
David Roach
Arkansas Tech University
Carolyn Rochelle
East Tennessee State University
Alfredo Romero
North Carolina A&T State University
Ann Rothermel
University of Akron
Jeff Rummel
Emory University
Deborah Rumsey
The Ohio State University
Stephen Russell
Weber State University
William Rybolt
Babson College
Fati Salimian
Salisbury University
Fatollah Salimian
Perdue School of Business
Samuel Sarri
College of Southern Nevada
Jim Schmidt
University of Nebraska—Lincoln
Patrick Scholten
Bentley University
Bonnie Schroeder
Ohio State University
Pali Sen
University of North Florida
Donald Sexton
Columbia University
Vijay Shah
West Virginia University—Parkersburg
Dmitriy Shaltayev
Christopher Newport University
Soheil Sibdari
University of Massachusetts—Dartmouth
Prodosh Simlai
University of North Dakota
Harvey Singer
George Mason University
Harry Sink
North Carolina A&T State University
Don Skousen
Salt Lake Community College
Robert Smidt
California Polytechnic State University
Gary Smith
Florida State University
Antoinette Somers
Wayne State University
Ryan Songstad
Augustana College
Erland Sorensen
Bentley University
Arun Kumar Srinivasan
Indiana University—Southeast
Scott Stevens
James Madison University
Alicia Strandberg
Temple University
Linda Sturges
Suny Maritime College
Wendi Sun
Rockland Trust
Bedassa Tadesse
University of Minnesota
Pandu Tadikamalta
University of Pittsburgh
Roberto Duncan Tarabay
University of Wisconsin—Madison
Faye Teer
James Madison University
Deborah Tesch
Xavier University
Patrick Thompson
University of Florida
Satish Thosar
University of Redlands
Ricardo Tovar-Silos
Lamar University
Quoc Hung Tran
Bridgewater State University
Elzbieta Trybus
California State University—Northridge
Fan Tseng
University of Alabama—Huntsville
Silvanus Udoka
North Carolina A&T State University
Shawn Ulrick
Georgetown University
Bulent Uyar
University of Northern Iowa
Ahmad Vakil
Tobin College of Business
Tim Vaughan
University of Wisconsin—Eau Claire
Raja Velu
Syracuse University
Holly Verhasselt
University of Houston—Victoria
Zhaowei Wang
Citizens Bank
Rachel Webb
Portland State University
Kyle Wells
Dixie State College
Alan Wheeler
University of Missouri—St. Louis
Mary Whiteside
University of Texas—Arlington
Blake Whitten
University of Iowa
Rick Wing
San Francisco State University

Jan Wolcott
Wichita State University
Rongning Wu
Baruch College
John Yarber
Northeast Mississippi Community College
John C. Yi
St. Joseph's University
Kanghyun Yoon
University of Central Oklahoma
Mark Zaporowski
Canisius College
Ali Zargar
San Jose State University
Dewit Zerom
California State University
Eugene Zhang
Midwestern State University
Ye Zhang
Indiana University—Purdue University—Indianapolis
Yi Zhang
California State University—Fullerton
Yulin Zhang
San Jose State University
Wencang Zhou
Baruch College
Zhen Zhu
University of Central Oklahoma

The editorial staff of McGraw-Hill Education are deserving of our gratitude for their guidance throughout this project, especially Noelle Bathurst, Pat Frederickson, Ryan McAndrews, Harper Christopher, Daryl Horrocks, and Egzon Shaqiri. We would also like to thank Eric Kambestad and Matt Kesselring for their outstanding research assistance.

BRIEF CONTENTS

CONTENTS

Essentials of
Business Statistics

1 Statistics and Data

Learning Objectives

After reading this chapter you should be able to:

LO **1.1** Describe the importance of statistics.

LO **1.2** Differentiate between descriptive statistics and inferential statistics.

LO **1.3** Explain the various data types.

LO **1.4** Describe variables and types of measurement scales.

Every day we are bombarded with data and claims. The analysis of data and the conclusions made from data are part of the field of statistics. A proper understanding of statistics is essential in understanding more of the real world around us, including business, sports, politics, health, social interactions—just about any area of contemporary human activity. In this first chapter, we will differentiate between sound statistical conclusions and questionable conclusions. We will also introduce some important terms that will help us describe different aspects of statistics and their practical importance. You are probably familiar with some of these terms already, from reading or hearing about opinion polls, surveys, and the all-pervasive product ads. Our goal is to place what you already know about these uses of statistics within a framework that we then use for explaining where they came from and what they really mean. A major portion of this chapter is also devoted to a discussion of variables and types of measurement scales. As we will see in later chapters, we need to distinguish between different variables and measurement scales in order to choose the appropriate statistical methods for analyzing data.

©Ian Lishma/Juice Images/Getty Images

Introductory Case

Tween Survey

Luke McCaffrey owns a ski resort two hours outside Boston, Massachusetts, and is in need of a new marketing manager. He is a fairly tough interviewer and believes that the person in this position should have a basic understanding of data fundamentals, including some background with statistical methods. Luke is particularly interested in serving the needs of the "tween" population (children aged 8 to 12 years old). He believes that tween spending power has grown over the past few years, and he wants their skiing experience to be memorable so that they want to return. At the end of last year's ski season, Luke asked 20 tweens four specific questions.

Q1. On your car drive to the resort, which radio station was playing?

Q2. On a scale of 1 to 4, rate the quality of the food at the resort (where 1 is poor, 2 is fair, 3 is good, and 4 is excellent).

Q3. Presently, the main dining area closes at 3:00 pm. What time do you think it should close?

Q4. How much of your *own* money did you spend at the lodge today?

The responses to these questions are shown in Table 1.1

FILE *Tween_Survey*

TABLE 1.1 Tween Responses to Resort Survey

Tween	Q1	Q2	Q3	Q4	Tween	Q1	Q2	Q3	Q4
1	JAMN94.5	4	5:00 pm	20	11	JAMN94.5	3	3:00 pm	0
2	MIX104.1	2	5:00 pm	10	12	JAMN94.5	4	4:00 pm	5
3	KISS108	2	4:30 pm	10	13	KISS108	2	4:30 pm	5
4	JAMN94.5	3	4:00 pm	0	14	KISS108	2	5:00 pm	10
5	KISS108	1	3:30 pm	0	15	KISS108	3	4:00 pm	5
6	JAMN94.5	1	6:00 pm	25	16	JAMN94.5	3	6:00 pm	20
7	KISS108	2	6:00 pm	15	17	KISS108	2	5:00 pm	15
8	KISS108	3	5:00 pm	10	18	MIX104.1	4	6:00 pm	15
9	KISS108	2	4:30 pm	10	19	KISS108	1	5:00 pm	25
10	KISS108	3	4:30 pm	20	20	KISS108	2	4:30 pm	10

Luke asks each job applicant to use the information to

1. Summarize the results of the survey.
2. Provide management with suggestions for improvement.

A synopsis from the job applicant with the best answers is provided at the end of Section 1.3.

LO **1.1**

Describe the importance of statistics.

1.1 THE RELEVANCE OF STATISTICS

In order to make intelligent decisions in a world full of uncertainty, we all have to understand statistics—the language of data. Data are usually compilations of facts, figures, or other contents, both numerical and nonnumerical. Insights from data enable businesses to make better decisions, such as deepening customer engagement, optimizing operations, preventing threats and fraud, and capitalizing on new sources of revenue. We must understand statistics or risk making uninformed decisions and costly mistakes. A knowledge of statistics also provides the necessary tools to differentiate between sound statistical conclusions and questionable conclusions drawn from an insufficient number of data points, "bad" data points, incomplete data points, or just misinformation. Consider the following examples.

Example 1. After Washington, DC, had record amounts of snow in the winter of 2010, the headline of a newspaper asked, "What global warming?"

Problem with conclusion: The existence or nonexistence of climate change cannot be based on one year's worth of data. Instead, we must examine long-term trends and analyze decades' worth of data.

Example 2. A gambler predicts that his next roll of the dice will be a lucky 7 because he did not get that outcome on the last three rolls.

Problem with conclusion: As we will see later in the text when we discuss probability, the probability of rolling a 7 stays constant with each roll of the dice. It does not become more likely if it did not appear on the last roll or, in fact, any number of preceding rolls.

Example 3. On January 10, 2010, nine days prior to a special election to fill the U.S. Senate seat that was vacated due to the death of Ted Kennedy, a *Boston Globe* poll gave the Democratic candidate, Martha Coakley, a 15-point lead over the Republican candidate, Scott Brown. On January 19, 2010, Brown won 52% of the vote, compared to Coakley's 47%, and became a U.S. senator for Massachusetts.

Problem with conclusion: Critics accused the *Globe,* which had endorsed Coakley, of purposely running a bad poll to discourage voters from coming out for Brown. In reality, by the time the *Globe* released the poll, it contained old information from January 2–6, 2010. Even more problematic was that the poll included people who said that they were unlikely to vote!

Example 4. Starbucks Corp., the world's largest coffee-shop operator, reported that sales at stores open at least a year climbed 4% at home and abroad in the quarter ended December 27, 2009. Chief Financial Officer Troy Alstead said that "the U.S. is back in a good track and the international business has similarly picked up. . . . Traffic is really coming back. It's a good sign for what we're going to see for the rest of the year." (www.bloomberg.com, January 20, 2010)

Problem with conclusion: In order to calculate same-store sales growth, which compares how much each store in the chain is selling compared with a year ago, we remove stores that have closed. Given that Starbucks closed more than 800 stores over the past few years to counter large sales declines, it is likely that the sales increases in many of the stores were caused by traffic from nearby, recently closed stores. In this case, same-store sales growth may overstate the overall health of Starbucks.

Example 5. Researchers at the University of Pennsylvania Medical Center found that infants who sleep with a nightlight are much more likely to develop myopia later in life (*Nature,* May 1999).

Problem with conclusion: This example appears to commit the *correlation-to-causation fallacy*. Even if two variables are highly correlated, one does not necessarily cause the other. *Spurious correlation* can make two variables appear closely related when no causal relation exists. Spurious correlation between two variables is not based on any demonstrable relationship, but rather can be explained by confounding factors. For instance, the fact that the cost of a hamburger is correlated with how much people spend on a computer is explained by a confounding factor called inflation, which makes both the hamburger and the computer costs grow over time. In a follow-up study regarding myopia, researchers at The Ohio State University found no link between infants who sleep with a nightlight and the development of myopia (*Nature,* March 2000). They did, however, find strong links between parental myopia and the development of child myopia, and between parental myopia and the parents' use of a nightlight in their children's room. So the confounding factor for both conditions (the use of a nightlight and the development of child myopia) is parental myopia. See www.tylervigen.com/spurious-correlations for some outrageous examples of spurious correlation.

Note the diversity of the sources of these examples—the environment, psychology, polling, business, and health. We could easily include others, from sports, sociology, the physical sciences, and elsewhere. Data and data interpretation show up in virtually every facet of life, sometimes spuriously. All of the preceding examples basically misuse data to add credibility to an argument. A solid understanding of statistics provides you with tools to react intelligently to information that you read or hear.

1.2 WHAT IS STATISTICS?

LO 1.2

Differentiate between descriptive statistics and inferential statistics.

In the broadest sense, we can define the study of statistics as the methodology of extracting useful information from a data set. Three steps are essential for doing good statistics. First, we have to find the right data, which are both complete and lacking any misrepresentation. Second, we must use the appropriate statistical tools, depending on the data at hand. Finally, an important ingredient of a well-executed statistical analysis is to clearly communicate numerical information into written language.

We generally divide the study of statistics into two branches: descriptive statistics and inferential statistics. **Descriptive statistics** refers to the summary of important aspects of a data set. This includes collecting data, organizing the data, and then presenting the data in the form of charts and tables. In addition, we often calculate numerical measures that summarize, for instance, the data's typical value and the data's variability. Today, the techniques encountered in descriptive statistics account for the most visible application of statistics—the abundance of quantitative information that is collected and published in our society every day. The unemployment rate, the president's approval rating, the Dow Jones Industrial Average, batting averages, the crime rate, and the divorce rate are but a few of the many "statistics" that can be found in a reputable newspaper on a frequent, if not daily, basis. Yet, despite the familiarity of descriptive statistics, these methods represent only a minor portion of the body of statistical applications.

The phenomenal growth in statistics is mainly in the field called inferential statistics. Generally, **inferential statistics** refers to drawing conclusions about a large set of data—called a **population**—based on a smaller set of **sample** data. A population is defined as all members of a specified group (not necessarily people), whereas a sample is a subset of that particular population. The individual values contained in a population or a sample are often referred to as **observations**. In most statistical applications, we must rely on sample data in order to make inferences about various characteristics of the population. For example, a 2016 Gallup survey found that only 50% of Millennials plan to be with their current job for more than a year. Researchers use this sample result, called a

sample statistic, in an attempt to estimate the corresponding unknown **population parameter**. In this case, the parameter of interest is the percentage of *all* Millennials who plan to be with their current job for more than a year. It is generally not feasible to obtain population data and calculate the relevant parameter directly, due to prohibitive costs and/or practicality, as discussed next.

POPULATION VERSUS SAMPLE

A population consists of all items of interest in a statistical problem. A sample is a subset of the population. We analyze sample data and calculate a sample statistic to make inferences about the unknown population parameter.

The Need for Sampling

A major portion of inferential statistics is concerned with the problem of estimating population parameters or testing hypotheses about such parameters. If we have access to data that encompass the entire population, then we would know the values of the parameters. Generally, however, we are unable to use population data for two main reasons.

- **Obtaining information on the entire population is expensive**. Consider how the monthly unemployment rate in the United States is calculated by the Bureau of Labor Statistics (BLS). Is it reasonable to assume that the BLS counts every unemployed person each month? The answer is a resounding NO! In order to do this, every home in the country would have to be contacted. Given that there are approximately 160 million individuals in the labor force, not only would this process cost too much, it would take an inordinate amount of time. Instead, the BLS conducts a monthly sample survey of about 60,000 households to measure the extent of unemployment in the United States.
- **It is impossible to examine every member of the population**. Suppose we are interested in the average length of life of a Duracell AAA battery. If we tested the duration of each Duracell AAA battery, then in the end, all batteries would be dead and the answer to the original question would be useless.

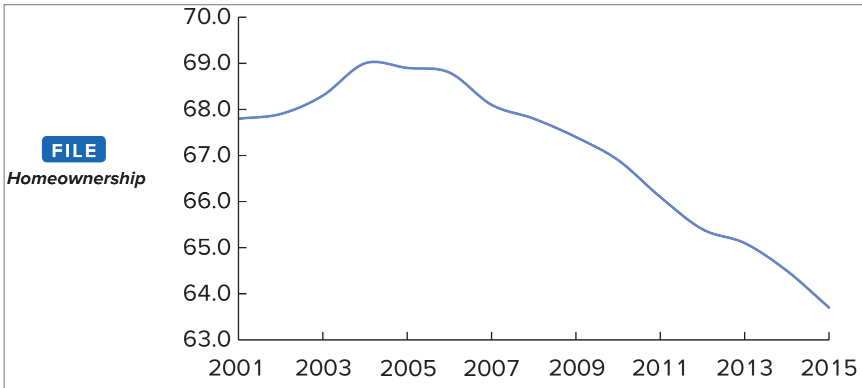
LO **1.3**

Explain the various data types.

Cross-Sectional and Time Series Data

Sample data are generally collected in one of two ways. **Cross-sectional data** refer to data collected by recording a characteristic of many subjects at the same point in time, or without regard to differences in time. Subjects might include individuals, households, firms, industries, regions, and countries. The tween data set presented in Table 1.1 in the introductory case is an example of cross-sectional data because it contains tween responses to four questions at the end of the ski season. It is unlikely that all 20 tweens took the questionnaire at exactly the same time, but the differences in time are of no relevance in this example. Other examples of cross-sectional data include the recorded scores of students in a class, the sale prices of single-family homes sold last month, the current price of gasoline in different states in the United States, and the starting salaries of recent business graduates from The Ohio State University.

Time series data refer to data collected over several time periods focusing on certain groups of people, specific events, or objects. Time series can include hourly, daily, weekly, monthly, quarterly, or annual observations. Examples of time series data include the hourly body temperature of a patient in a hospital's intensive care unit, the daily price of General Electric stock in the first quarter of 2015, the weekly exchange rate between the U.S. dollar and the euro over the past six months, the monthly sales of cars at a dealership in 2016, and the annual growth rate of India in the last decade.



Source: Federal Reserve Bank of St. Louis

FIGURE 1.1 Homeownership Rate (%) in the United States from 2001 through 2015

Figure 1.1 shows a plot of the national homeownership rate in the United States from 2001 through 2015. According to the U.S. Census Bureau, the national homeownership rate in the first quarter of 2016 plummeted to 63.6% from a high of 69.4% in 2004. An obvious explanation for the decline in homeownership is the stricter lending practices caused by the housing market crash in 2007 that precipitated a banking crisis and the Great Recession. This decline can also be attributed to home prices outpacing wages in the sample period.

CROSS-SECTIONAL DATA AND TIME SERIES DATA

Cross-sectional data contain values of a characteristic of many subjects at the same point or approximately the same point in time. Time series data contain values of a characteristic of a subject over time.

Structured and Unstructured Data

As mentioned earlier, consumers and businesses are increasingly turning to data to make decisions. When you hear the word "data," you probably imagine lots of numbers and perhaps some charts and graphs as well. In reality, data can come in multiple forms. For example, information exchange in social networking services such as Facebook, LinkedIn, Twitter, YouTube, and blogs also constitutes data. In order to better understand the various forms of data, we make a distinction between structured data and unstructured data.

The term **structured data** generally refers to data that has a well-defined length and format. Structured data reside in a predefined row-column format. Examples of structured data include numbers, dates, and groups of words and numbers called strings. Structured data generally consist of numerical information that is objective. In other words, structured data are not open to interpretation. The data set that appears in Table 1.1 from the introductory case is an example of structured data.

Unlike structured data, **unstructured data** (or unmodeled data) do not conform to a predefined row-column format. They tend to be textual (e.g., written reports, e-mail messages, doctor's notes, or open-ended survey responses) or have multimedia contents (e.g., photographs, videos, and audio data). Even though these data may have some implied structure (e.g., a report title, an e-mail's subject line, or a time stamp on a photograph), they are still considered unstructured because they do not conform to a row-column model required in most database systems. Social media data, such as those that appear on Facebook, LinkedIn, Twitter, YouTube, and blogs, are examples of unstructured data.

Big Data

Nowadays, businesses and organizations generate and gather more and more data at an increasing pace. The term **big data** is a catchphrase, meaning a massive volume of both structured and unstructured data that are extremely difficult to manage, process, and analyze using traditional data processing tools. Despite the challenges, big data present great opportunities to glean intelligence from data that affects company revenues, margins, and organizational efficiency.

Big data, however, do not necessarily imply complete (population) data. Take, for example, the analysis of all Facebook users. It certainly involves big data, but if we consider all Internet users in the world, Facebook data are only a very large sample. There are many Internet users who do not use Facebook, so the data on Facebook do not represent the population. Even if we define the population as pertaining to those who use online social media, Facebook is still one of many social media services that consumers use. Therefore, Facebook data would still just be considered a large sample.

In addition, we may choose not to use a big data set in its entirety even when it is available. Sometimes it is just inconvenient to analyze a very large data set because it is computationally burdensome, even with a modern, high-capacity computer system. Other times, the additional benefits of working with a big data set may not justify its associated additional resource costs. In sum, we often choose to work with a small data set, which, in a sense, is a sample drawn from big data.

STRUCTURED DATA, UNSTRUCTURED DATA, AND BIG DATA

Structured data reside in a predefined row-column format, while unstructured data do not conform to a predefined row-column format. The term big data is used to describe a massive volume of both structured and unstructured data that are extremely difficult to manage, process, and analyze using traditional data processing tools. The availability of big data, however, does not necessarily imply complete (population) data.

In this textbook, we will not cover specialized tools to manage, process, and analyze big data. Instead, we will focus on structured data. Text analytics and other sophisticated tools to analyze unstructured data are beyond the scope of this textbook.

Data on the Web

©Comstock Images/Jupiter Images

At every moment, data are being generated at an increasing velocity from countless sources in an overwhelming volume. Many experts believe that 90% of the data in the world today were created in the last two years alone. Not surprisingly, businesses continue to grapple with how to best ingest, understand, and operationalize large volumes of data. We access much of the data in this text by simply using a search engine like Google. These search engines direct us to data-providing websites. For instance, searching for economic data may lead you to the Bureau of Economic Analysis (www.bea.gov), the Bureau of Labor Statistics (www.bls.gov/data), the Federal Reserve Economic Data (research.stlouisfed.org), and the U.S. Census Bureau (www.census.gov/data.html). These websites provide data on inflation, unemployment, GDP, and much more, including useful international data. The National Climatic Data Center (www.ncdc.noaa.gov/data-access) provides a large collection of environmental, meteorological, and climate data. Similarly, transportation data can be found at www.its-rde.net. The University of Michigan has compiled sentiment data found at www.sca.isr.umich.edu. Several cities in the United States have publicly available data in categories such as finance, community and economic development, education, and crime. For example, the Chicago data portal data.cityofchicago.org provides a large volume of city-specific data. Excellent world development indicator data are available at data.worldbank.org. The happiness index data for most countries are available at www.happyplanetindex.org/data.

Private corporations also make data available on their websites. For example, Yahoo Finance (www.finance.yahoo.com) and Google Finance (www.google.com/finance) list data such as stock prices, mutual fund performance, and international market data. Zillow (www.zillow.com/) supplies data for recent home sales, monthly rent, mortgage rates, and so forth. Similarly, www.espn.go.com offers comprehensive sports data on both professional and college teams. Finally, *The Wall Street Journal, The New York Times, USA Today, The Economist, Business Week, Forbes,* and *Fortune* are all reputable publications that provide all sorts of data. We would like to point out that all of the above data sources represent only a fraction of publicly available data.

EXERCISES 1.2

1. It came as a big surprise when Apple's touch screen iPhone 4, considered by many to be the best smartphone ever, was found to have a problem (*The New York Times,* June 24, 2010). Users complained of weak reception, and sometimes even dropped calls, when they cradled the phone in their hands in a particular way. A quick survey at a local store found that 2% of iPhone 4 users experienced this reception problem.
   a. Describe the relevant population.
   b. Does 2% denote the population parameter or the sample statistic?
2. Many people regard video games as an obsession for youngsters, but, in fact, the average age of a video game player is 35 years (Telegraph.co.uk, July 4, 2013). Is the value 35 likely the actual or the estimated average age of the population? Explain.
3. An accounting professor wants to know the average GPA of the students enrolled in her class. She looks up information on Blackboard about the students enrolled in her class and computes the average GPA as 3.29.
   a. Describe the relevant population.
   b. Does the value 3.29 represent the population parameter or the sample statistic?
4. Business graduates in the United States with a marketing concentration earn high salaries. According to the Bureau of Labor Statistics, the average annual salary for marketing managers was $140,660 in 2015.
   a. What is the relevant population?
   b. Do you think the average salary of $140,660 was computed from the population? Explain.
5. Research suggests that depression significantly increases the risk of developing dementia later in life (*BBC News,* July 6, 2010). In a study involving 949 elderly persons, it was reported that 22% of those who had depression went on to develop dementia, compared to only 17% of those who did not have depression.
   a. Describe the relevant population and the sample.
   b. Do the numbers 22% and 17% represent population parameters or sample statistics?
6. Go to www.finance.yahoo.com/ to get a current stock quote for General Electric, Co. (ticker symbol = GE). Then, click on historical prices to record the monthly adjusted close price of General Electric stock in 2016. Create a table that uses this information. What type of data do these numbers represent? Comment on the data.
7. Ask 20 of your friends whether they live in a dormitory, a rental unit, or other form of accommodation. Also find out their approximate monthly lodging expenses. Create a table that uses this information. What type of data do these numbers represent? Comment on the data.
8. Go to www.zillow.com/ and find the sale price data of 20 single-family homes sold in Las Vegas, Nevada, in the last 30 days. In the data set, include the sale price, the number of bedrooms, the square footage, and the age of the house. What type of data do these numbers represent? Comment on the data.
9. The Federal Reserve Bank of St. Louis is a good source for downloading economic data. Go to research.stlouisfed.org/fred2/ to extract quarterly data on gross private saving (GPSAVE) from 2012 to 2015 (16 observations). Create a table that uses this information. Plot the data over time and comment on the savings trend in the United States.
10. Go to the U.S. Census Bureau website at www.census.gov/ and extract the most recent median household income for Alabama, Arizona, California, Florida, Georgia, Indiana, Iowa, Maine, Massachusetts, Minnesota, Mississippi, New Mexico, North Dakota, and Washington. What type of data do these numbers represent? Comment on the regional differences in income.
11. Go to *The New York Times* website at www.nytimes.com/ and review the front page. Would you consider the data on the page to be structured or unstructured? Explain.

12. Conduct an online search to compare price and fuel economy of small hybrid vehicles such as Toyota Prius, Ford Fusion, and Chevrolet Volt. Would the resulting data be structured or unstructured? Explain.
13. Ask your peers about their online social media usage. In particular collect information on (a) whether they use Facebook, Instagram, and Snapchat, (b) how often they use each social media service, and (c) their overall satisfaction with each of these services using a 5-point numerical scale where 1 represents totally unsatisfied and 5 represents totally satisfied. Are the resulting data structured or unstructured? Are the data cross-sectional or time series?
14. Conduct an online search for a weekly car rental in Seattle, Washington, and Portland, Oregon, for different car types and rental car companies for the year 2017. Are the data structured or unstructured? Are the data cross-sectional or time series?

LO **1.4**

Describe variables and types of measurement scales.

1.3 VARIABLES AND SCALES OF MEASUREMENT

When we conduct a statistical investigation, we invariably focus on people, objects, or events with particular characteristics. When a characteristic of interest differs in kind or degree among various observations, then the characteristic can be termed a **variable**. We further categorize a variable as either qualitative or quantitative. For a **qualitative variable**, we use labels or names to identify the distinguishing characteristic of each observation. For instance, the 2010 Census asked each respondent to indicate gender on the form. Each respondent chose either male or female. Gender is a qualitative variable. Other examples of qualitative variables include race, profession, type of business, the manufacturer of a car, and so on.

A variable that assumes meaningful numerical values is called a **quantitative variable**. Quantitative variables, in turn, are either discrete or continuous. A **discrete variable** assumes a countable number of values. Consider the number of children in a family or the number of points scored in a basketball game. We may observe values such as 3 children in a family or 90 points being scored in a basketball game, but we will not observe 1.3 children or 92.5 scored points. The values that a discrete variable assumes need not be whole numbers. For example, the price of a stock for a particular firm is a discrete variable. The stock price may take on a value of $20.37 or $20.38, but it cannot take on a value between these two points. Finally, a discrete variable may assume an infinite number of values, but these values are countable; that is, they can be presented as a sequence x_1, x_2, x_3, and so on. The number of cars that cross the Golden Gate Bridge on a Saturday is a discrete variable. Theoretically, this variable assumes the values 0, 1, 2, . . .

A **continuous variable** is characterized by uncountable values within an interval. Weight, height, time, and investment return are all examples of continuous variables. For example, an unlimited number of values occur between the weights of 100 and 101 pounds, such as 100.3, 100.625, 100.8342, and so on. In practice, however, continuous variables may be measured in discrete values. We may report a newborn's weight (a continuous variable) in discrete terms as 6 pounds 10 ounces and another newborn's weight in similar discrete terms as 6 pounds 11 ounces.

QUALITATIVE VARIABLES VERSUS QUANTITATIVE VARIABLES

A variable is a general characteristic being observed on a set of people, objects, or events, where each observation varies in kind or degree. Labels or names are used to categorize the distinguishing characteristics of a qualitative variable; eventually, these attributes may be coded into numbers for purposes of data processing. A quantitative variable assumes meaningful numerical values, and can be further categorized as either discrete or continuous. A discrete variable assumes a countable number of values, whereas a continuous variable is characterized by uncountable values within an interval.

In order to choose the appropriate statistical methods for summarizing and analyzing data, we need to distinguish between different measurement scales. All data measurements can be classified into one of four major categories: nominal, ordinal, interval, and ratio. Nominal and ordinal scales are used for qualitative variables, whereas interval and ratio scales are used for quantitative variables. We discuss these scales in ascending order of sophistication.

The Nominal Scale

The **nominal scale** represents the least sophisticated level of measurement. If we are presented with nominal data, all we can do is categorize or group the data. The values in the data set differ merely by name or label. Consider the following example.

©ymgerman/Shutterstock

Each company listed in Table 1.2 is a member of the Dow Jones Industrial Average (DJIA). The DJIA is a stock market index that shows how 30 large, publicly owned companies based in the United States have traded during a standard trading session in the stock market. Table 1.2 also shows where stocks of these companies are traded: on either the National Association of Securities Dealers Automated Quotations (Nasdaq) or the New York Stock Exchange (NYSE). These data are classified as nominal scale since we are simply able to group or categorize them. Specifically, only four stocks are traded on Nasdaq, whereas the remaining 26 are traded on the NYSE.

TABLE 1.2 Companies of the DJIA and Exchange Where Stock Is Traded

Company	Exchange	Company	Exchange
3M (MMM)	NYSE	Intel (INTC)	Nasdaq
American Express (AXP)	NYSE	Johnson & Johnson (JNJ)	NYSE
Apple (AAPL)	Nasdaq	JPMorgan Chase (JPM)	NYSE
Boeing (BA)	NYSE	McDonald's (MCD)	NYSE
Caterpillar (CAT)	NYSE	Merck (MRK)	NYSE
Chevron (CVX)	NYSE	Microsoft (MFST)	Nasdaq
Cisco (CSCO)	Nasdaq	Nike (NKE)	NYSE
Coca-Cola (KO)	NYSE	Pfizer (PFE)	NYSE
Disney (DIS)	NYSE	Procter & Gamble (PG)	NYSE
Dupont (DD)	NYSE	Travelers (TRV)	NYSE
ExxonMobil (XOM)	NYSE	United Health (UNH)	NYSE
General electric (GE)	NYSE	United Tech. Corp. (UTX)	NYSE
Goldman Sachs (GS)	NYSE	Verizon (VZ)	NYSE
Home Depot (HD)	NYSE	Visa (V)	NYSE
IBM (IBM)	NYSE	Walmart (WMT)	NYSE

Source: Money.CNN.com information retrieved March 21, 2015.

Often we substitute *numbers* for the particular qualitative characteristic or trait that we are grouping. One reason why we do this is for ease of exposition; always referring to the National Association of Securities Dealers Automated Quotations, or even Nasdaq, becomes awkward and unwieldy. In addition, as we will see later in the text, statistical analysis is greatly facilitated by using numbers instead of names. For example, we might use the number 0 to show that a company's stock is traded on Nasdaq and the number 1 to show that a company's stock is traded on the NYSE, or in tabular form:

Exchange	Number of Companies Trading on Exchange
0	4
1	26

The Ordinal Scale

Compared to the nominal scale, the **ordinal scale** reflects a stronger level of measurement. With ordinal data we are able to both *categorize* and *rank* the data with respect to some characteristic or trait. The weakness with ordinal data is that we cannot interpret the difference between the ranked values because the actual numbers used are arbitrary. For example, suppose you are asked to classify the service at a particular hotel as excellent, good, fair, or poor. A standard way to record the ratings is

Category	Rating
Excellent	4
Good	3
Fair	2
Poor	1

Here the value attached to excellent (4) is higher than the value attached to good (3), indicating that the response of excellent is preferred to good. However, another representation of the ratings might be

Category	Rating
Excellent	100
Good	80
Fair	70
Poor	40

Excellent still receives a higher value than good, but now the difference between the two categories is 20 (100 − 80), as compared to a difference of 1 (4 − 3) when we use the first classification. In other words, *differences between categories are meaningless with ordinal data.* (We also should note that we could reverse the ordering so that, for instance, excellent equals 40 and poor equals 100; this renumbering would not change the nature of the data.)

EXAMPLE 1.1

FILE
Tween_Survey

In the introductory case, four questions were posed to tweens. The first question (Q1) asked tweens to name the radio station that was playing on the ride to the resort, and the second question (Q2) asked tweens to rate the food quality at the resort on a scale of 1 to 4. The tweens' responses to these questions are shown in Table 1.1 in the introductory case.

a. What is the scale of measurement of the radio station data?

b. How are the data based on the ratings of the food quality similar to the radio station data? How are the data different?

c. Summarize the tweens' responses to Q1 and Q2 in tabular form. How can the resort use the information from these responses?

SOLUTION:

a. When asked which radio station played on the car ride to the resort, tweens responded with one of the following answers: JAMN94.5, MIX104.1, or KISS108. These are nominal data—the values in the data differ merely in name or label.

b. Since we can both categorize and rank the food quality data, we classify these responses as ordinal data. Ordinal data are similar to nominal data in the sense that we can categorize the data. The main difference between ordinal and nominal data is that the categories of ordinal data are ranked. A rating of 4 is better than a rating of 3. With the radio station data, we cannot say that KISS108 is ranked higher than MIX104.1; some tweens may argue otherwise, but we simply categorize nominal data without ranking.

c. With respect to the radio station data (Q1), we can assign 1 to JAMN94.5, 2 to MIX104.1, and 3 to KISS108. Counting the responses that fall into each category, we find that six tweens listened to 1, two listened to 2, and 12 listened to 3, or in tabular form:

Radio Station	Number of Tweens
1	6
2	2
3	12

Twelve of the 20 tweens, or 60%, listened to KISS108. This information could prove useful to the management of the resort as they make decisions as to where to allocate their advertising dollars. If the resort could only choose to advertise at one radio station, it would appear that KISS108 would be the wise choice.

Given the food quality responses (Q2), we find that three of the tweens rated food quality with a 4, six tweens rated food quality with a 3, eight tweens rated food quality with a 2, and three tweens rated food quality with a 1, or in tabular form:

Rating	Number of Tweens
4	3
3	6
2	8
1	3

The food quality results may be of concern to management. Just as many tweens rated the food quality as excellent as compared to poor. Moreover, the majority $[(8 + 3)/20 = 55\%]$ felt that the food was, at best, fair. Perhaps a more extensive survey that focuses solely on food quality would reveal the reason for their apparent dissatisfaction.

As mentioned earlier, nominal and ordinal scales are used for *qualitative variables*. Values corresponding to a qualitative variable are typically expressed in words but are coded into numbers for purposes of data processing. When summarizing the results of a qualitative variable, we typically count the number or calculate the percentage of persons or objects that fall into each possible category. With a qualitative variable, we are unable to perform meaningful arithmetic operations such as adding and subtracting.

The Interval Scale

With data that are measured on an **interval scale**, not only can we categorize and rank the data, we are also assured that the differences between scale values are meaningful. Thus, the arithmetic operations of addition and subtraction are meaningful. The Fahrenheit scale for temperatures is an example of an interval scale. Not only is 60 degrees Fahrenheit hotter than 50 degrees Fahrenheit, the same difference of 10 degrees also exists between 90 and 80 degrees Fahrenheit.

The main drawback of data on an interval scale is that the value of zero is arbitrarily chosen; the zero point of an interval scale does not reflect a complete absence of what is being measured. No specific meaning is attached to zero degrees Fahrenheit other than to say it is 10 degrees colder than 10 degrees Fahrenheit. With an arbitrary zero point, meaningful ratios cannot be constructed. For instance, it is senseless to say that 80 degrees is twice as hot as 40 degrees; in other words, the ratio 80/40 has no meaning.

The Ratio Scale

The **ratio scale** represents the strongest level of measurement. Ratio data have all the characteristics of interval data as well as a *true zero* point, which allows us to interpret the ratios of values. A ratio scale is used to measure many types of data in business analysis. Variables such as sales, profits, and inventory levels are expressed as ratio data. A meaningful zero allows us to state, for example, that profits for firm A are double those of firm B. Measurements such as weight, time, and distance are also measured on a ratio scale since zero is meaningful.

Unlike qualitative data, arithmetic operations are valid on interval- and ratio-scaled values. In later chapters, we will calculate summary measures for the typical value and the variability of quantitative variables; we cannot calculate these measures if the variable is qualitative in nature.

EXAMPLE 1.2

FILE
Tween_Survey

In the last two questions from the introductory case's survey (Q3 and Q4), the 20 tweens were asked: "What time should the main dining area close?" and "How much of your *own* money did you spend at the lodge today?" Their responses appear in Table 1.1 in the introductory case.

a. How are the time data classified? In what ways do the time data differ from ordinal data? What is a potential weakness of this measurement scale?

b. What is the measurement scale of the money data? Why is it considered the strongest form of data?

c. In what ways is the information from Q3 and Q4 useful for the resort?

SOLUTION:

a. Clock time responses, such as 3:00 pm and 3:30 pm, or 5:30 pm and 6:00 pm, are on an interval scale. Interval data are a stronger measurement scale than ordinal data because differences between interval-scaled values are meaningful. In this particular example, we can say that 3:30 pm is 30 minutes later than 3:00 pm and 6:00 pm is 30 minutes later than 5:30 pm. The weakness with interval data is that the value of zero is arbitrary. Here, with the clock time responses, we have no apparent zero point; however, we could always arbitrarily define a zero point, say, at 12:00 am. Thus, although differences are comparable with interval data, ratios are meaningless due to the arbitrariness of the zero point. In other words, it is senseless to form the ratio 6:00 pm/3:00 pm and conclude that 6:00 pm is twice as long a time period as 3:00 pm.

b. Since the tweens' responses are in dollar amounts, this is ratio data. The ratio scale is the strongest form of data because we can categorize and rank values as well as calculate meaningful differences. Moreover, since there is a natural zero point, valid ratios can also be calculated. For example, the data show that three tweens spent \$20. These tweens spent four times as much as the three tweens that spent \$5 (\$20/\$5 = 4).

c. A review of the clock time responses (Q3) in Table 1.1 shows that the vast majority of the tweens would like the dining area to remain open later. In fact, only one tween feels that the dining area should close at 3:00 pm. An inspection of the money responses (Q4) in Table 1.1 indicates that only three of the 20 tweens did not spend any of his/her own money. This is very important information. It does appear that the discretionary spending of this age group is significant. The resort would be wise to cater to some of their preferences.

SYNOPSIS OF INTRODUCTORY CASE

A preliminary survey of tween preferences conducted by the management of a ski resort two hours outside Boston, Massachusetts, revealed some interesting information.

- Tweens were first asked to name the radio station that they listened to on the way to the resort. The responses show that 60% of the tweens listened to KISS108. If the resort wishes to contact tweens using this medium, it may want to direct its advertising dollars to this station.
- Next, the tweens were asked to rate the food quality at the resort on a scale of 1 to 4 (where 1 is poor, 2 is fair, 3 is good, and 4 is excellent). The survey results with respect to food quality are disturbing. The majority of the tweens, 55% (11/20), felt that the food was, at best, fair. A more extensive study focusing on food quality appears necessary.
- Tweens were then asked what time the main dining area should close, given that it presently closes at 3:00 pm. The data suggest that the vast majority of the tweens (19 out of 20) would like the dining area to remain open later.

©Dennis Welsh/UpperCut Images/Getty Images

- Finally, the tweens were asked to report the amount of their *own* money that they spent at the lodge. The resort is likely pleased with the responses to this question because 17 of the 20 tweens spent their own money at the lodge. This finding appears consistent with the belief that tween spending is growing.

EXERCISES 1.3

15. Which of the following variables are qualitative and which are quantitative? If the variable is quantitative, then specify whether the variable is discrete or continuous.
    a. Points scored in a football game.
    b. Racial composition of a high school classroom.
    c. Heights of 15-year-olds.
16. Which of the following variables are qualitative and which are quantitative? If the variable is quantitative, then specify whether the variable is discrete or continuous.
    a. Colors of cars in a mall parking lot.
    b. Time it takes each student to complete a final exam.
    c. The number of patrons who frequent a restaurant.
17. In each of the following scenarios, define the type of measurement scale.
    a. A kindergarten teacher marks whether each student is a boy or a girl.
    b. A ski resort records the daily temperature during the month of January.
    c. A restaurant surveys its customers about the quality of its waiting staff on a scale of 1 to 4, where 1 is poor and 4 is excellent.
18. In each of the following scenarios, define the type of measurement scale.
    a. An investor collects data on the weekly closing price of gold throughout a year.

b. An analyst assigns a sample of bond issues to one of the following credit ratings, given in descending order of credit quality (increasing probability of default): AAA, AA, BBB, BB, CC, D.

c. The dean of the business school at a local university categorizes students by major (i.e., accounting, finance, marketing, etc.) to help in determining class offerings in the future.

19. In each of the following scenarios, define the type of measurement scale.

a. A meteorologist records the amount of monthly rainfall over the past year.

b. A sociologist notes the birth year of 50 individuals.

c. An investor monitors the daily stock price of BP following the 2010 oil disaster in the Gulf of Mexico.

20. A professor records the majors of her 30 students as follows:

Accounting	Economics	Undecided	Finance	Management
Management	Finance	Marketing	Economics	Management
Marketing	Finance	Marketing	Accounting	Finance
Finance	Undecided	Management	Undecided	Economics
Economics	Accounting	Management	Undecided	Economics
Accounting	Economics	Management	Accounting	Economics

a. What is the measurement scale of these data?

b. Summarize the results in tabular form.

c. What information can be extracted from the data?

21. **FILE** ***DOW_Characteristics.*** The accompanying table shows a portion of the 30 companies that comprise the Dow Jones Industrial Average (DJIA). The second column shows the year that the company joined the DJIA (Year). The third column shows each company's Morningstar rating (Rating). (Five stars is the best rating that a company can receive, indicating that the company's stock price is undervalued and thus a very good buy. One star is the worst rating a company can be given, implying that the stock price is overvalued and a bad buy.) Finally, the fourth column shows each company's stock price as of March 17, 2017 (Price in $).

Company	Year	Rating	Price
3M (MMM)	1976	**	192.36
American Express (AMX)	1982	***	79.25
⋮	⋮	⋮	⋮
Wal-Mart (WMT)	1991	****	69.89

Source: Morningstar ratings retrieved from www.morningstar.com on March 17, 2017; stock prices retrieved from www.finance.yahoo.com.

a. What is the measurement scale of the Year data? What are the strengths of this type of data? What are the weaknesses?

b. What is the measurement scale of Morningstar's star-based rating system? Summarize Morningstar's star-based rating system for the companies in tabular form. Let 5 denote *****, 4 denote ****, and so on. What information can be extracted from these data?

c. What is the measurement scale of the Stock Price data? What are its strengths?

CONCEPTUAL REVIEW

LO 1.1 Describe the importance of statistics.

A proper understanding of statistical ideas and concepts helps us understand more of the real world around us, including issues in business, sports, politics, health, and social interactions. We must understand statistics or risk making bad decisions and costly mistakes. A knowledge of statistics also provides the necessary tools to differentiate between sound statistical conclusions and questionable conclusions drawn from an insufficient number of data points, "bad" data points, incomplete data points, or just misinformation.

LO 1.2 Differentiate between descriptive statistics and inferential statistics.

The study of statistics is generally divided into two branches: descriptive statistics and inferential statistics. **Descriptive statistics** refers to the summary of a data set in the form of tables, graphs, and/or the calculation of numerical measures. **Inferential statistics** refers to extracting useful information from a sample to draw conclusions about a population. A **population** consists of all items of interest in a statistical problem; a **sample** is a subset of that population.

In general, we use sample data rather than population data for two main reasons: (1) obtaining information on the entire population is expensive and/or (2) it is impossible to examine every item of the population.

LO 1.3 Explain the various data types.

Cross-sectional data contain values of a characteristic of many subjects at the same point in time or without regard to differences in time. **Time series data** contain values of a characteristic of a subject over time.

Structured data conform but **unstructured data** do not conform to a predefined row-column format.

Big data are a massive volume of both structured and unstructured data that are extremely difficult to manage, process, and analyze using traditional data processing tools. Big data, however, do not necessarily imply complete (population) data.

LO 1.4 Describe variables and types of measurement scales.

A variable is categorized as either qualitative or quantitative. For a **qualitative variable**, we use labels or names to identify the distinguishing characteristic of each observation. A **quantitative variable** assumes meaningful numerical values and can be further categorized as either **discrete** or **continuous**. A discrete variable assumes a countable number of values, whereas a continuous variable is characterized by uncountable values within an interval.

All data measurements can be classified into one of four major categories.

- The values on a **nominal scale** differ merely by name or label. These values are then simply categorized or grouped by name.
- The values on an **ordinal scale** can be categorized *and* ranked; however, differences between the ranked values are meaningless.
- Values on the **interval scale** can be categorized and ranked, and differences between values are meaningful. The main drawback of the interval scale is that the value of zero is arbitrarily chosen; this implies that ratios constructed from interval-scaled values bear no significance.
- Ratio data have all the characteristics of interval data as well as a true zero point; thus, as its name implies, meaningful ratios can be calculated with values on the ratio scale.

Nominal and ordinal scales are used for qualitative variables. When summarizing the results of qualitative data, we typically count the number or calculate the percentage of persons or objects that fall into each possible category. Interval and ratio scales are used for quantitative variables. Unlike qualitative variables, arithmetic operations are valid on quantitative variables.

2 Tabular and Graphical Methods

Learning Objectives

After reading this chapter you should be able to:

LO **2.1** Summarize qualitative data by constructing a frequency distribution.

LO **2.2** Construct and interpret a pie chart and a bar chart.

LO **2.3** Summarize quantitative data by constructing a frequency distribution.

LO **2.4** Construct and interpret a histogram, a polygon, and an ogive.

LO **2.5** Construct and interpret a stem-and-leaf diagram.

LO **2.6** Construct and interpret a scatterplot.

People often have difficulty processing information provided by data in its raw form. A useful way of interpreting data effectively is through data visualization. In this chapter, we present several tabular and graphical tools that can help us organize and present data. We first make a table referred to as a frequency distribution using qualitative data. For visual representations of qualitative data, we construct a pie chart or a bar chart. For quantitative data, we again make a frequency distribution. In addition to giving us an overall picture of where the data tend to cluster, a frequency distribution using quantitative data also shows us how the data are spread out from the lowest value to the highest value. For visual representations of quantitative data, we construct a histogram, a polygon, an ogive, and a stem-and-leaf diagram. Finally, we show how to construct a scatterplot, which graphically depicts the relationship between two quantitative variables. We will find that a scatterplot is a very useful tool when conducting correlation and regression analysis, topics discussed in depth later in the text.


©Mitch Diamond/Photodisc/Getty Images

Introductory Case

House Prices in Southern California

Mission Viejo, a city located in Southern California, was named the safest city in California and the third-safest city in the nation (CQPress.com, November 23, 2009). Matthew Edwards, a relocation specialist for a real estate firm in Mission Viejo, often relays this piece of information to clients unfamiliar with the many benefits that the city offers. Recently, a client from Seattle, Washington, asked Matthew for a summary of recent sales. The client is particularly interested in the availability of houses in the $500,000 range. Table 2.1 shows the sale price for 36 single-family houses in Mission Viejo during June 2010.

TABLE 2.1 Recent Sale Price of Houses in Mission Viejo, CA, for June 2010 (data in $1,000s)

430	670	530	521	669	445
520	417	525	350	660	412
460	533	430	399	702	735
475	525	330	560	540	537
670	538	575	440	460	630
521	370	555	425	588	430

Source: www.zillow.com.

FILE
MV_Houses

Matthew wants to use the sample information to

1. Make summary statements concerning the range of house prices.
2. Comment on where house prices tend to cluster.
3. Calculate appropriate percentages in order to compare house prices in Mission Viejo, California, to those in Seattle, Washington.

A synopsis of this case is provided in Section 2.2.