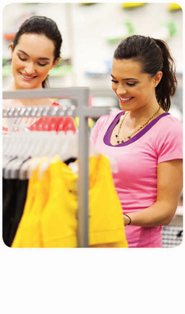
Business Statistics

Second Edition

COMMUNICATING WITH NUMBERS



Jaggia / Kelly

Mc Graw Hill Education

BUSINESS STATISTICS

Second Edition

BUSINESS STATISTICS

Communicating with Numbers

Sanjiv Jaggia
California Polytechnic State University

Alison Kelly
Suffolk University

McGraw Hill Education

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 0 DOW/DOW 1 0 9 8 7 6 5

ISBN 978-0-07-802055-1
MHID 0-07-802055-7

Senior Vice President, Products & Markets: *Kurt L. Strand*
Vice President, General Manager, Products & Markets: *Marty Lange*
Vice President, Content Design & Delivery: *Kimberly Meriwether David*
Managing Director: *Jame Heine*
Marketing Director: *Lynn Breithaupt*
Brand Manager: *Dolly Womack*
Director, Product Development: *Rose Koos*
Product Developer: *Christina Holt*
Director of Digital Content: *Doug Ruby*
Digital Product Analyst: *Kevin Shanahan*
Director, Content Design & Delivery: *Linda Avenarius*
Program Manager: *Mark Christianson*
Content Project Managers: *Harvey Yep / Bruce Gin*
Buyer: *Jennifer Pickel*
Design: *Srdjan Savanovic*
Content Licensing Specialists: *Keri Johnson / John Leland / Rita Hingtgen*
Cover Image: © *Comstock/Stockbyte/Getty Images/RF;* © *Mitch Diamond/Photodisc/Getty Images/RF;* © *Mark Bowden/iStock/Getty Images Plus/Getty Images:* © *Rob Tringali//Getty Images;* © *Image Source, all rights reserved/RF;* © *Honqi Zhang/iStock/Getty Images Plus/Getty Images/RF;* © *imageBROKER/Alamy /RF;* © *Lim Hyeonsu/TongRo Images/Corbis/RF;* © *Yellow Dog Productions/Digital Vision/Getty Images/RF*
Compositor: *MPS Limited, A Macmillan Company*
Printer: *R. R. Donnelley*

Library of Congress Cataloging-in-Publication Data

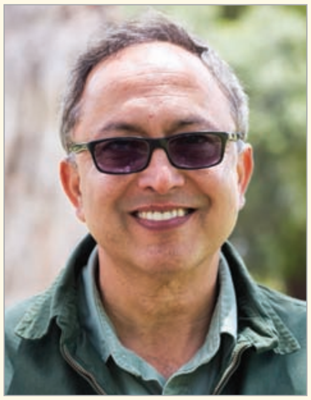
Jaggia, Sanjiv, 1960-
Business statistics: communicating with numbers / Sanjiv Jaggia, California Polytechnic State University, Alison Kelly, Suffolk University.
Second Edition.
pages cm.—(Business statistics)
ISBN 978-0-07-802055-1 (hardback)
1. Commercial statistics. I. Hawke, Alison Kelly. II. Title.
HF1017.J34 2015
519.5—dc23 2015023383

www.mhhe.com

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

ABOUT THE AUTHORS

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*, 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

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 regularly 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 focuses on how large financial institutions satisfy requirements mandated by the Dodd-Frank Act. She resides in Hamilton, Massachusetts, with her husband and two children.

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

Statistics can be a fun and enlightening course for both students and teachers. From our years of experience in the classroom, we have found that an effective way to make statistics interesting is to use timely business applications to which students can relate. If interest can be sparked at the outset, students may end up learning statistics without realizing they are doing so. By carefully matching timely applications with statistical methods, students learn to appreciate the relevance of business statistics in our world today. We wrote *Business Statistics: Communicating with Numbers* because we saw a need for a contemporary, core statistics textbook 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 professional users, it has been our goal to make the subject matter more relevant and the presentation of material more straightforward for students.

In *Business Statistics*, we have incorporated fundamental topics that are applicable for students with various backgrounds and interests. The text is intellectually stimulating, practical, and visually attractive, from which students can learn and instructors can teach. Although it is application oriented, it is also mathematically sound and uses notation that is generally accepted for the topic being covered.

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 *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 it 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 a business environment. Students are taught how to take the data, apply it, and convey the information in a meaningful way.

Unique Coverage of Regression Analysis. Relevant 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, and JMP are provided in chapter appendices.

Connect® Business Statistics. *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 *Business Statistics* features a number of improvements suggested by numerous reviewers and users of the first edition.

First, every section of every chapter has been scrutinized, and if a change would enhance readability, then that change was made. In addition, Excel instructions have been streamlined in every chapter. We feel that this modification provides a more seamless reinforcement for the relevant topic. For those instructors who prefer to omit the Excel parts, these sections can be easily skipped. Moreover, most chapters now include an appendix that provides brief instructions for Minitab, SPSS, and JMP. More detailed instructions for Minitab, SPSS, and JMP can be found in *Connect*.

Dozens of applied exercises of varying levels of difficulty have been added to just about every section of every chapter. 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.

Both of us use *Connect* in our classes. In an attempt to make the technology component seamless with the text itself, we have reviewed every *Connect* exercise. In addition, we have painstakingly revised tolerance levels and added rounding rules. The positive feedback from users due to these adjustments has been well worth the effort. In addition, we have included numerous new exercises in *Connect*. We have also reviewed every probe from LearnSmart. Instructors who teach in an online or hybrid environment will especially appreciate these modifications.

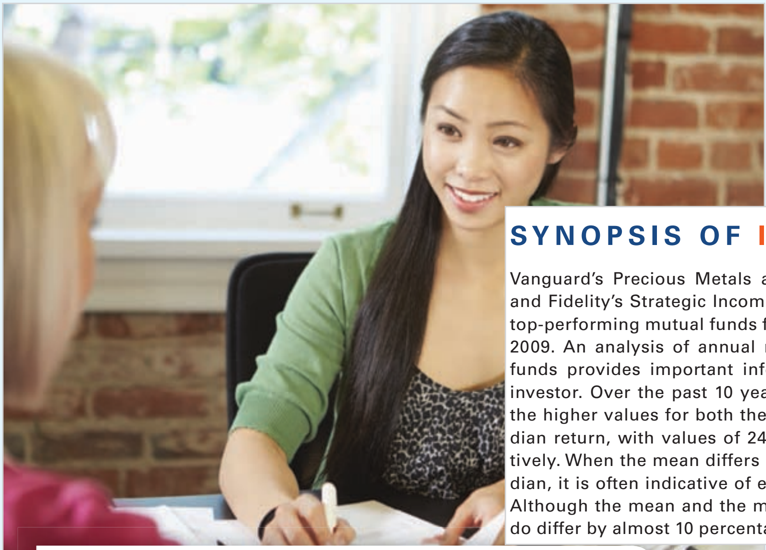
Here are some of the more noteworthy, specific changes:

- Some of the Learning Outcomes have been rewritten for the sake of consistency.
- In Chapter 3 (Numerical Descriptive Measures), the discussion of the weighted mean occurs in Section 3.1 (Measures of Central Location) instead of Section 3.7 (Summarizing Grouped Data). Section 3.6 has been renamed from "Chebyshev's Theorem and the Empirical Rule" to "Analysis of Relative Location"; in addition, we have added a discussion of z-scores in this section.
- In Chapter 4 (Introduction to Probability), the term *a priori* has been replaced by *classical*.
- In Chapter 5 (Discrete Probability Distributions), the use of graphs now complements the discussion of the binomial and Poisson distributions.
- In Chapter 7 (Sampling and Sampling Distributions), the standard error of a statistic is now denoted as "*se*" instead of "*SD*." For instance, the standard error of the sample mean is now denoted as $se(\bar{X})$ instead of $SD(\bar{X})$.
- The discussion of the properties of estimators has been moved from Section 8.1 to an appendix in Chapter 7.
- In Section 16.1 (Polynomial Models), the discussion of the marginal effects of x on y has been expanded.
- In Section 17.1 (Dummy Variables), there is now an example of how to conduct a hypothesis test when the original reference group must be changed.
- In Chapter 18 (Time Series Forecasting), the data used for the "Writing with Statistics" example has been revised.

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.



SYNOPSIS OF INTRODUCTORY CASE

Vanguard's Precious Metals and Mining fund (Metals) and Fidelity's Strategic Income fund (Income) were two top-performing mutual funds for the years 2000 through 2009. An analysis of annual return data for these two funds provides important information for any type of investor. Over the past 10 years, the Metals fund posts the higher values for both the mean return and the median return, with values of 24.65% and 33.83%, respectively. When the mean differs dramatically from the median, it is often indicative of extreme values or outliers. Although the mean and the median for the Metals fund do differ by almost 10 percentage points, a boxplot analysis reveals no outliers. The mean return and

INTRODUCTORY CASE

Investment Decision

Rebecca Johnson works as an investment counselor at a large bank. Recently, an inexperienced investor asked Johnson about clarifying some differences between two top-performing mutual funds from the last decade: Vanguard's Precious Metals and Mining fund (henceforth, Metals) and Fidelity's Strategic Income fund (henceforth, Income). The investor shows Johnson the return data that he has accessed over the Internet, but the investor has trouble interpreting the data. Table 3.1 shows the return data for these two mutual funds for the years 2000–2009.

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 take 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

The Associated Press reports that income inequality is at record levels in the United States (September 28, 2010). Over the years, the rich have become richer while working-class wages have stagnated. A local Latino politician has been vocal regarding his concern about the welfare of Latinos, especially given the recent downturn of the U.S. economy. In various speeches, he has stated that the mean salary of Latino households in his county has fallen below the 2008 mean of $49,000. He has also stated that the proportion of Latino households making less than $30,000 has risen above the 2008 level of 20%. Both of his statements are based on income data for 36 Latino households in the county, as shown in Table 9.5.

TABLE 9.5 Representative Sample of Latino Household Incomes in 2010

FILE *Latino_Income*

22	36	78	103	38	43
62	53	26	28	25	31
62	44	51	38	77	37
29	38	46	52	61	57
20	72	41	73	16	32
52	28	69	27	53	46

Incomes are measured in $1,000s and have been adjusted for inflation.

Trevor Jones is a newspaper reporter who is interested in verifying the concerns of the local politician.

Trevor wants to use the sample information to:

1. Determine if the mean income of Latino households has fallen below the 2008 level of $49,000.
2. Determine if the percentage of Latino households making less than $30,000 has risen above 20%.

Sample Report— Assessing Whether Data Follow the Normal Distribution

As part of a broader report concerning the mutual fund industry in general, three-year return data for the 50 largest mutual funds were collected with the objective of determining whether or not the data follow a normal distribution. Information of this sort is particularly useful because much statistical inference is based on the assumption of normality. If the assumption of normality is not supported by the data, it may be more appropriate to use nonparametric techniques to make valid inferences. Table 12.A shows relevant summary statistics for three-year returns for the 50 largest mutual funds.

TABLE 12.A Three-Year Return Summary Measures for the 50 Largest Mutual Funds, August 2008

Mean	Median	Standard Deviation	Skewness	Kurtosis
5.96%	4.65%	3.39%	1.37	2.59

The average three-year return for the 50 largest mutual funds is 5.96%, with a median of 4.65%. When the mean is significantly greater than the median, it is often an indication of a positively skewed distribution. The skewness coefficient of 1.37 seems to support this claim. Moreover, the kurtosis coefficient of 2.59 suggests a distribution that is more peaked than the normal distribution. A formal test will determine whether the conclusion from the sample can be deemed real or due to chance.

The goodness-of-fit test is first applied to check for normality. The raw data is converted into a frequency distribution with five intervals ($k = 5$). Expected frequencies are

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

Our coverage of regression analysis is more extensive than that of the vast majority of texts. This focus reflects the topic's growing use in practice. We combine simple and multiple regression in one chapter, which we believe is a seamless grouping and eliminates needless repetition. 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. Three more in-depth chapters cover statistical inference, nonlinear relationships, dummy variables, and binary choice models.

Chapter 14: Regression Analysis
Chapter 15: Inference with Regression Models
Chapter 16: Regression Models for Nonlinear Relationships
Chapter 17: Regression Models with Dummy Variables

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*

The inclusion of material used on a regular basis by investment professionals adds real-world credibility to the text and course and better prepares students for the real world.
Bob Gillette, *University of Kentucky*

Inclusion of Important Topics

In our teaching outside the classroom, we have found that several fundamental topics important to business are not covered by the majority of traditional texts. For example, most books do not integrate the geometric mean, mean-variance analysis, and the Sharpe ratio with descriptive statistics. Similarly, the discussion of probability concepts generally does not include odds ratios, risk aversion, and the analysis of portfolio returns. We cover these important topics throughout the text. Overall, our text contains material that practitioners use on a regular basis.

THE SHARPE RATIO

The **Sharpe ratio** measures the extra reward per unit of risk. The Sharpe ratio for an investment I is computed as:

$$\frac{\bar{x}_I - \bar{R}_f}{s_I}$$

where $\bar{x}_I$ is the mean return for the investment, $\bar{R}_f$ is the mean return for a risk-free asset such as a Treasury bill (T-bill), and s_I is the standard deviation for the investment.

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®

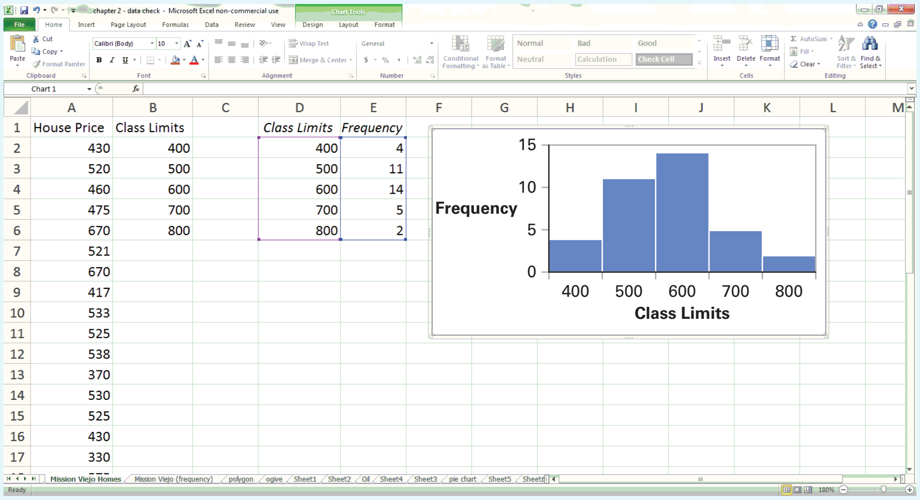
We prefer that students first focus on and absorb the statistical material before replicating their results with a computer. We feel that solving each application manually provides students with a deeper understanding of the relevant concept. However, we recognize that, primarily due to cumbersome calculations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

Using Excel to Construct a Histogram

A. FILE Open ***MV_Houses*** (Table 2.1).

B. In a column next to the data, enter the values of the upper limits of each class, or in this example, 400, 500, 600, 700, and 800; label this column "Class Limits." The reason for these entries is explained in step D. The house-price data and the class limits (as well as the resulting frequency distribution and histogram) are shown in Figure 2.8.

FIGURE 2.8 Constructing a histogram from raw data with Excel



> . . . 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 the text. For instance, students first use these real data to calculate summary measures and then continue on to make statistical inferences with confidence intervals and hypothesis tests and 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.

Applications

...he Department of Transportation (DOT) fields thousands of ...mplaints about airlines each year. The DOT categorizes ... tallies complaints, and then periodically publishes ...gs of airline performance. The following table ...ents the 2006 results for the 10 largest U.S. airlines.

	Complaints*	Airline	Complaints*
...west ...s	1.82	Northwest Airlines	8.84
...Blue ...ways	3.98	Delta Airlines	10.35
Alaska Airlines	5.24	American Airlines	10.87
AirTran Airways	6.24	US Airways	13.59
Continental Airlines	8.83	United Airlines	13.60

Source: Department of Transportation; *per million passengers.

a. Which airline fielded the least amount of complaints? Which airline fielded the most? Calculate the range.
b. Calculate the mean and the median number of complaints for this sample.
c. Calculate the variance and the standard deviation.

44. The monthly closing stock prices (rounded to the nearest dollar) for Starbucks Corp. and Panera Bread Co. for the first six months of 2010 are reported in the following table.

to promise good returns (*The Wall Street Journal*, September 24, 2010). Marcela Treisman works for an investment firm in Michigan. Her assignment is to analyze the rental market in Ann Arbor, which is home to the University of Michigan. She gathers data on monthly rent for 2011 along with the square footage of 40 homes. A portion of the data is shown in the accompanying table.

Monthly Rent	Square Footage
645	500
675	648
⋮	⋮
2400	2700

Source: http://www.zillow.com.

a. Calculate the mean and the standard deviation for monthly rent.
b. Calculate the mean and the standard deviation for square footage.
c. Which sample data exhibit greater relative dispersion?

46. FILE ***Largest_Corporations.*** Access the data accompanying this exercise. It shows the Fortune 500 rankings of America's largest corporations for 2010. Next to each corporation are its market capitalization (in billions of dollars as of March 26, 2010) and its total return to investors for the year 2009.
a. Calculate the coefficient of variation for market

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 Distinguish between discrete and continuous random variables.

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

LO 5.2 Describe the probability distribution for a discrete random variable.

The **probability distribution 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 5.3 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 $SD(X) = \sigma = \sqrt{\sigma^2}$.

LO 5.4 Distinguish between risk-neutral, risk-averse, and risk-loving consumers.

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. Finally, a **risk-loving consumer** may accept a risky prospect even if the expected gain is negative.

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*

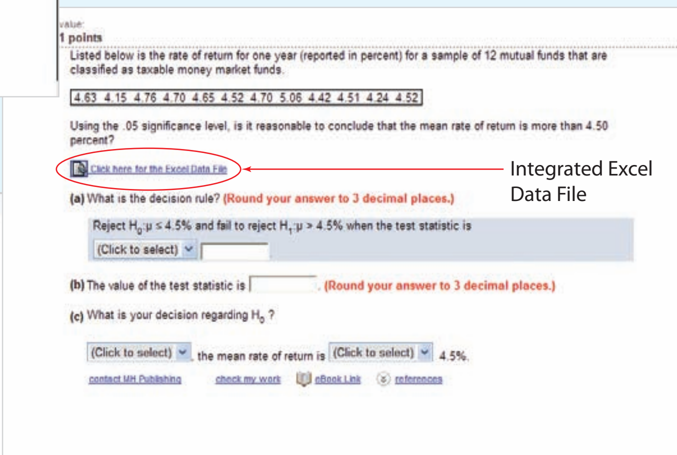
What Technology Connects Students . . .

McGraw-Hill *Connect*® *Business Statistics*

McGraw-Hill *Connect Business Statistics* is an online assignment and assessment solution that connects students with the tools and resources they'll need to achieve success through faster learning, higher retention, and more efficient studying. It provides instructors with tools to quickly select content for assignments according to the topics and learning objectives they want to emphasize.

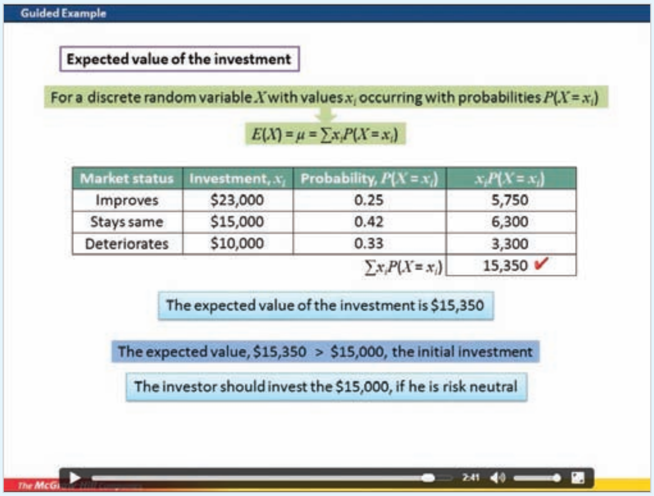
Online Assignments. *Connect Business Statistics* helps students learn more efficiently by providing practice material and feedback when they are needed. *Connect* grades homework automatically and provides instant feedback on any problems that students are challenged to solve.

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.

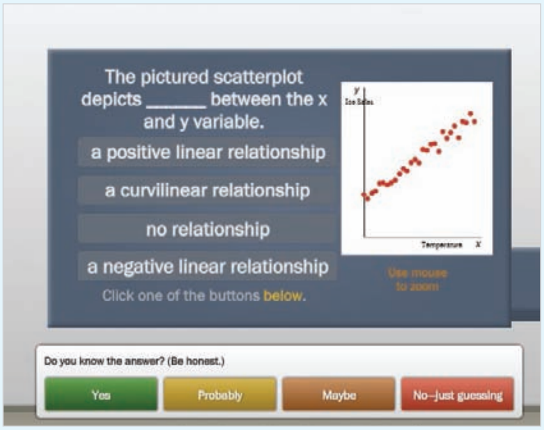


to Success in Business Statistics?

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.



LearnSmart. LearnSmart adaptive self-study technology in *Connect Business Statistics* helps students make the best use of their study time. LearnSmart provides a seamless combination of practice, assessment, and remediation for every concept in the textbook. LearnSmart's intelligent software adapts to students by supplying questions on a new concept when students are ready to learn it. With LearnSmart, students will spend less time on topics they understand and instead focus on the topics they need to master.



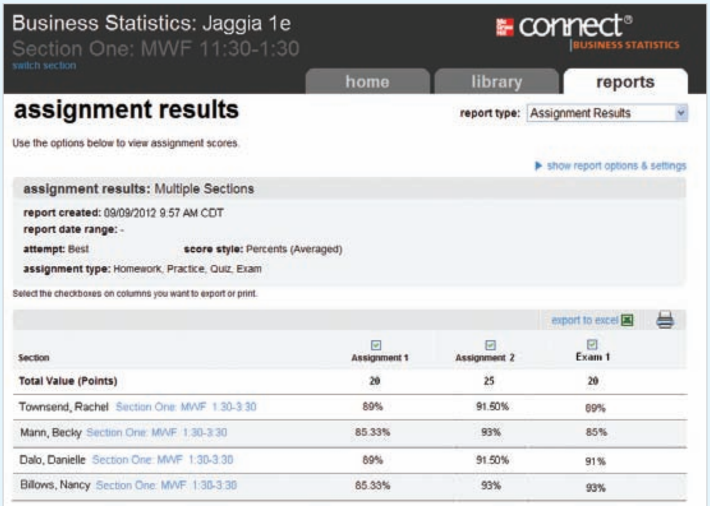
SmartBook®, which is powered by LearnSmart, is the first and only adaptive reading experience designed to change the way students read and learn. It creates a personalized reading experience by highlighting the most relevant concepts a student needs to learn at that moment in time. As a student engages with SmartBook, the reading experience continuously adapts by highlighting content based on what the student knows and doesn't know. This ensures that the focus is on the content he or she needs to learn, while simultaneously promoting long-term retention of material. Use SmartBook's real-time reports to quickly identify the concepts that require more attention from individual students or the entire class. The end result? Students are more engaged with course content, can better prioritize their time, and come to class ready to participate.

What Technology Connects Students . . .

Simple Assignment Management and Smart Grading. When it comes to studying, time is precious. *Connect Business Statistics* helps students learn more efficiently by providing feedback and practice material when they need it, where they need it. When it comes to teaching, your time also is precious. The grading function enables you to

- Have assignments scored automatically, giving students immediate feedback on their work and the ability to compare their work with correct answers.
- Access and review each response; manually change grades or leave comments for students to review.

Student Reporting. *Connect Business Statistics* keeps instructors informed about how each student, section, and class is performing, allowing for more productive use of lecture and office hours. The progress-tracking function enables you to



- View scored work immediately and track individual or group performance with assignment and grade reports.
- Access an instant view of student or class performance relative to topic and learning objectives.
- Collect data and generate reports required by many accreditation organizations, such as AACSB.

Instructor Library. The *Connect Business Statistics* 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 Business Statistics* Instructor Library includes:

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

to Success in Business Statistics?

***Connect* Insight.** *Connect* Insight is *Connect*'s new one-of-a-kind visual analytics dashboard—now available for both instructors and students—that provides at-a-glance information regarding student performance, which is immediately actionable. By presenting assignment, assessment, and topical performance results together with a time metric that is easily visible for aggregate or individual results, *Connect* Insight gives the user the ability to take a just-in-time approach to teaching and learning, which was never before available. *Connect* Insight presents data that empowers students and helps instructors efficiently and effectively improve class performance.

Mobile. Students and instructors can now enjoy convenient anywhere, anytime access to *Connect* with a new mobile interface that's been designed for optimal use of tablet functionality. More than just a new way to access *Connect*, users can complete assignments, check progress, study, and read material, with full use of LearnSmart, SmartBook, and *Connect* Insight—*Connect*'s new at-a-glance visual analytics dashboard.

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**.

What Software Is Available with This Text?

MegaStat® for Microsoft Excel® 2003, 2007, and 2010 (and Excel: Mac 2011)

Access Card ISBN: 0077426274 *Note: Best option for both Windows and Mac users.*

MegaStat® by J. B. Orris of Butler University is a full-featured Excel add-in that is available through the access card packaged with the text or on the *MegaStat* website at www.mhhe.com/megastat. It works with Excel 2003, 2007, and 2010 (and Excel: Mac 2011). 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 Instructors?

Online Course Management

The **Best** of **Both Worlds**

McGraw-Hill Higher Education and Blackboard have teamed up. What does this mean for you?

1. **Single sign-on.** Now you and your students can access McGraw-Hill's *Connect*® and Create™ right from within your Blackboard course—all with one single sign-on.
2. **Deep integration of content and tools.** You get a single sign-on with *Connect* and Create, and you also get integration of McGraw-Hill content and content engines right into Blackboard. Whether you're choosing a book for your course or building *Connect* assignments, all the tools you need are right where you want them—inside of Blackboard.
3. **One grade book.** Keeping several grade books and manually synchronizing grades into Blackboard is no longer necessary. When a student completes an integrated *Connect* assignment, the grade for that assignment automatically (and instantly) feeds your Blackboard grade center.
4. **A solution for everyone.** Whether your institution is already using Blackboard or you just want to try Blackboard on your own, we have a solution for you. McGraw-Hill and Blackboard can now offer you easy access to industry-leading technology and content, whether your campus hosts it or we do. Be sure to ask your local McGraw-Hill representative for details.

What Resources Are Available for Students?

CourseSmart

ISBN: 1259335062

CourseSmart is a convenient way to find and buy eTextbooks. CourseSmart has the largest selection of eTextbooks available anywhere, offering thousands of the most commonly adopted textbooks from a wide variety of higher-education publishers. CourseSmart eTextbooks are available in one standard online reader with full text search, notes and highlighting, and e-mail tools for sharing notes between classmates. Visit **www.CourseSmart.com** for more information on ordering.

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/lrwin, 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**.

ACKNOWLEDGMENTS

We would like to acknowledge the following people for their help in the development of the first and second editions of *Business Statistics*, as well as the ancillaries and digital content.

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
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
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
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
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
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
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
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
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
Raja Velu
Syracuse University
Holly Verhasselt
University of Houston—Victoria
Zhaowei Wang
Citizen's 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
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/Irwin are deserving of our gratitude for their guidance throughout this project, especially Christina Holt, Dolly Womack, Doug Ruby, Harvey Yep, Bruce Gin, and Srdjan Savanovic.

BRIEF CONTENTS

CONTENTS

PART THREE
Probability and Probability Distributions

PART FOUR
Basic Inference

CHAPTER 12

CHI-SQUARE TESTS 402

PART FIVE

Advanced Inference

CHAPTER 13

ANALYSIS OF VARIANCE 432

CHAPTER 14

REGRESSION ANALYSIS 476

CHAPTER 15

INFERENCE WITH REGRESSION MODELS 514

CHAPTER 20

NONPARAMETRIC TESTS *686*

APPENDIXES

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 need for sampling and discuss various data types.**
- LO **1.4** **Describe variables and various 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, which are referenced throughout the text, 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 various 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.

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

TABLE 1.1 Tween Responses to Skylark Valley Resort Survey

FILE *Tween_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

1.1 THE RELEVANCE OF STATISTICS

Describe the importance of statistics.

In order to make intelligent decisions in a world full of uncertainty, we all have to understand statistics—the language of data. Unfortunately, many people avoid learning statistics because they believe (incorrectly!) that statistics simply deals with incomprehensible formulas and tedious calculations, and that it has no use in real life. This type of thinking is far from the truth because we encounter statistics *every day* in real life. We must understand statistics or risk making uninformed decisions and costly mistakes. While it is true that statistics incorporates formulas and calculations, it is logical reasoning that dictates how the data are collected, the calculations implemented, and the results communicated. 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 stated, "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 on a relation that arises in the data solely because each of those variables is related to some third variable. In a follow-up study, 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 cause of both conditions (the use of a nightlight and the development of child myopia) is parental myopia.

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. In most statistical applications, we must rely on sample data in order to make inferences about various characteristics of the population. For example, a 2010 survey of 1,208 registered voters by a USA TODAY/Gallup Poll found that President Obama's job performance was viewed favorably by only 41% of those polled, his lowest rating in a USA TODAY/Gallup Poll since he took office in January 2009 (*USA TODAY*, August 3, 2010). 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* registered voters that view the president's job performance favorably. 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**.

LO 1.3

Explain the need for sampling and discuss various data types.

The Need for Sampling

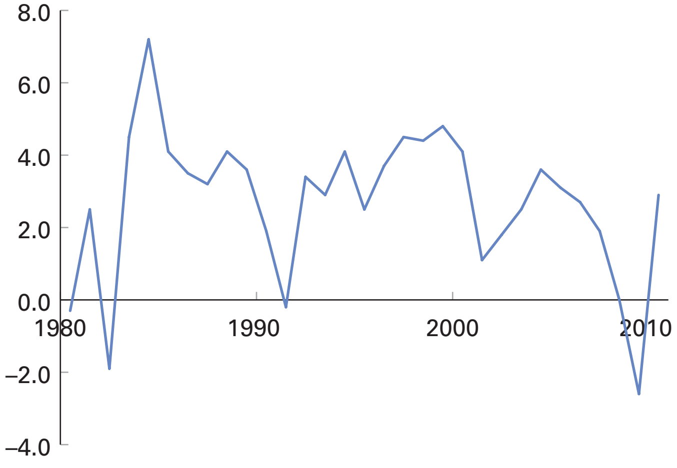
A major portion of inferential statistics is concerned with the problem of estimating population parameters or testing hypotheses about such parameters. If we have access to data that encompass the entire population, then we would know the values of the parameters. Generally, however, we are unable to 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 over 150 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.

Types of Data

Sample data are generally collected in one of two ways. **Cross-sectional data** refers 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 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 refers to data collected by recording a characteristic of a subject over several time periods. 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 IBM stock in the first quarter of 2015, the weekly exchange rate between the U.S. dollar and the euro, the monthly sales of cars at a dealership in 2014, and the annual growth rate of India in the last decade. Figure 1.1 shows a plot of the real (inflation-adjusted) GDP growth rate of the United States from 1980 through 2010. The average growth rate for this period is 2.7%, yet the plot indicates a great deal of variability in the series. It exhibits a wavelike movement, spiking downward in 2008 due to the economic recession before rebounding in 2010.



SOURCE: Bureau of Economic Analysis.

FIGURE 1.1 Real GDP growth rate from 1980 through 2010

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.

Getting Started on the Web

As you can imagine, there is an abundance of data on the Internet. We accessed much of the data in this text by simply using a search engine like Google. These search engines often directed us to the same data-providing sites. For instance, the U.S. federal government publishes a great deal of economic and business data. The Bureau of Economic Analysis (BEA), the Bureau of Labor Statistics (BLS), the Federal Reserve Economic Data (FRED), and the U.S. Census Bureau provide data on inflation, unemployment, gross domestic product (GDP), and much more. Zillow.com is a real estate site that supplies data such as recent home sales, monthly rent, and mortgage rates. Finance.yahoo.com is a financial site that lists data such as stock prices, mutual fund performance, and international market data. *The Wall Street Journal*, *The New York Times*, *USA Today*, *The Economist*, and *Fortune* are all reputable publications that provide all sorts of data. Finally, espn.com offers comprehensive sports data on both professional and college teams. We list these sites in Table 1.2 and summarize *some* of the data that are available.

TABLE 1.2 Select Internet Data Sites

Internet Site	Select Data Availability
Bureau of Economic Analysis (BEA)	National and regional data on gross domestic product (GDP) and personal income, international data on trade in goods and services.
Bureau of Labor Statistics (BLS)	Inflation rates, unemployment rates, employment, pay and benefits, spending and time use, productivity.
Federal Reserve Economic Data (FRED)	Banking, business/fiscal data, exchange rates, reserves, monetary base.
U.S. Census Bureau	Economic indicators, foreign trade, health insurance, housing, sector-specific data.
zillow.com	Recent home sales, home characteristics, monthly rent, mortgage rates.
finance.yahoo.com	Historical stock prices, mutual fund performance, international market data.
The Wall Street Journal, The New York Times, USA Today, The Economist, and *Fortune*	Poverty, crime, obesity, and plenty of business-related data.
espn.com	Professional and college teams' scores, rankings, standings, individual player statistics.

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 (Reuters.com, August 21, 2009). 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 $104,400 in 2007.
   a. What is the relevant population?
   b. Do you think the average salary of $104,400 was computed from the population? Explain.
5. Recent 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 Google, Inc. (ticker symbol = GOOG). Then, click on historical prices to record the monthly adjusted close price of Google stock in 2010. 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 2008 to 2011 (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. Another good source of data is the U.S. Census Bureau. Go to 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.

LO 1.4

1.3 VARIABLES AND SCALES OF MEASUREMENT

Describe variables and various types of measurement scales.

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 the 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.

Each company listed in Table 1.3 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 1.3 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.

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.