This International Student Edition is for use outside of the U.S.

Business Statistics

COMMUNICATING WITH NUMBERS

Regression
Nonparametrics Nonlinear
Education Competing Hypothesis Logistic
Mean Bell-Shaped Control Charts
Stock Returns p-Value Advertising
Case Studies Happiness
Risk
Transportation Distributions Sales R
Assumptions
Forecasting Insights Big Data
Housing
Health Reports Probability
Descriptive Analytics Polling
Explanatory Variables
Sports Excel Salaries
Scatterplots
Discrimination
Inference Proportion
ANOVA
Binomial

Jaggia | Kelly

Fourth Edition

Mc Graw Hill

BUSINESS STATISTICS

Mc
Graw
Hill

The McGraw Hill Series in Operations and Decision Sciences

SUPPLY CHAIN MANAGEMENT

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

Johnson
Purchasing and Supply Management
Sixteenth Edition

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

Stock and Manrodt
Fundamentals of Supply Chain Management

PROJECT MANAGEMENT

Larson and Gray
Project Management: The Managerial Process
Eighth 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

Schindler
Business Research Methods
Fourteenth Edition

BUSINESS FORECASTING

Keating and Wilson
Forecasting and Predictive Analytics
Seventh 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
Sixteenth Edition

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

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

Stevenson
Operations Management
Fourteenth Edition

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

BUSINESS STATISTICS

Bowerman, Drougas, Duckworth, Froelich, Hummel, Moninger, and Schur
Business Statistics and Analytics in Practice
Ninth Edition

Doane and Seward
Applied Statistics in Business and Economics
Seventh Edition

Doane and Seward
Essential Statistics in Business and Economics
Third Edition

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

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

Jaggia and Kelly
Business Statistics: Communicating with Numbers
Fourth Edition

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

BUSINESS ANALYTICS

Jaggia, Kelly, Lertwachara, and Chen
Business Analytics: Communicating with Numbers

BUSINESS MATH

Slater and Wittry
Practical Business Math Procedures
Thirteenth Edition

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

Fourth Edition

BUSINESS STATISTICS

Communicating with Numbers

Sanjiv Jaggia
California Polytechnic State University

Alison Kelly
Suffolk University

Mc Graw Hill

BUSINESS STATISTICS

Published by McGraw Hill LLC, 1325 Avenue of the Americas, New York, NY 10121.

Some ancillaries, including electronic and print components, may not be available to customers outside the United States.

This book is printed on acid-free paper.

1 2 3 4 5 6 7 8 9 LWI 24 23 22 21

ISBN 978-1-260-59756-1
MHID 1-260-59756-3

Cover Image: *McGraw Hill*

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

The Internet addresses listed in the text were accurate at the time of publication. The inclusion of a website does not indicate an endorsement by the authors or McGraw Hill LLC, and McGraw Hill LLC does not guarantee the accuracy of the information presented at these sites.

mheducation.com/highered

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

ABOUT THE AUTHORS

Sanjiv Jaggia

Courtesy Sanjiv Jaggia

Sanjiv Jaggia is a professor of economics and finance at California Polytechnic State University in San Luis Obispo. Dr. Jaggia holds a Ph.D. from Indiana University and is a Chartered Financial Analyst (CFA®). He enjoys research in statistics and data analytics applied to a wide range of business disciplines. Dr. Jaggia has published numerous papers in leading academic journals and has co-authored three successful textbooks, two in business statistics and one in business analytics. His ability to communicate in the classroom has been acknowledged by several teaching awards. Dr. Jaggia resides in San Luis Obispo 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. Dr. Kelly holds a Ph.D. from Boston College and is a Chartered Financial Analyst (CFA®). Dr. Kelly has published in a wide variety of academic journals and has co-authored three successful textbooks, two in business statistics and one in business analytics. Her courses in applied statistics and econometrics are popular with students as well as working professionals. She 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. Dr. Kelly resides in Hamilton, Massachusetts, with her husband, daughter, and son. In her spare time, she enjoys exercising and gardening.

Courtesy Alison Kelly

Business Statistics: Communicating with Numbers

Reviewer Quotes

"[Jaggia and Kelly's text is] an introductory statistics textbook which is rigorous in statistics with modern technology embedded."

-Qiang Zhen, University of North Florida

"This introductory statistics book is relevant and approachable. The book and its materials support teaching in various modalities. It offers an applied orientation with a reasonable and appropriate theoretical foundation."

-Kathryn Ernstberger, Indiana University Southeast

"The authors . . . do an excellent job of introducing the concepts. Illustrations, modern and relevant examples and applications, and exercises. . . This is a well-rounded and excellent textbook in introductory statistics and data analysis."

-Mohammad A. Kazemi, University of North Carolina, Charlotte

"This book and its accompanying online resources is ideal for an introductory in statistics with an emphasis in business. There is a thoughtful balance between concepts and applications."

-Ted Galanthay, Ithaca College

"Excellent coverage. . . It is a great book."

-Ricardo S. Tovar-Silos, Lamar University

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

We wrote *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 *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.

Continuing Key Features

The fourth edition of *Business Statistics* reinforces and expands six core features that were well-received in earlier editions.

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 Data. 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® and R. Students are taught to develop an understanding of the concepts and how to derive the calculation; then Excel and R are used as a tool to perform the cumbersome calculations.

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.

Features New to the Fourth Edition

In the fourth edition of *Business Statistics,* we have made substantial revisions that address the current needs of the market. These revisions are based on the feedback of countless reviewers and users of our earlier editions.

The emphasis in this edition has been to strengthen the connection between business statistics and data analytics. More than ever, colleges and universities across the United States and abroad are incorporating business analytics into their curricula, and businesses are scrambling to find qualified professionals who can translate statistical analysis into decisions that improve performance. We believe that the fourth edition will not only introduce students to data analytics, but will also excite them to further explore the field. There are four major innovations in this edition.

Descriptive: More emphasis on data preparation and visualization

- New sections devoted to data preparation in Chapter 1
- New sections devoted to data visualization methods in Chapter 2
- Discussion of subsetted means in Chapter 3
- Discussion of pivot tables used to analyze empirical probabilities in Chapter 4

Predictive: Significant rewrite of regression and forecasting

- Streamlined discussion of goodness-of-fit measures in Chapter 14
- Revised section on model assumptions and common violations in Chapter 15
- Improved visualizations to explore nonlinear models in Chapters 16, 17, and 18
- New subsection on accuracy rates for binary choice models in Chapter 17
- Separate sections devoted to linear and nonlinear forecasting models in Chapter 18

Technology: More reliance on statistical software and Connect

- Greater use of Excel and R in solving problems
- Expanded R instructions in all regression and forecasting chapters
- Excel and/or R instructions included in most exercises in Connect
- Improved Connect product to facilitate teaching in an online environment

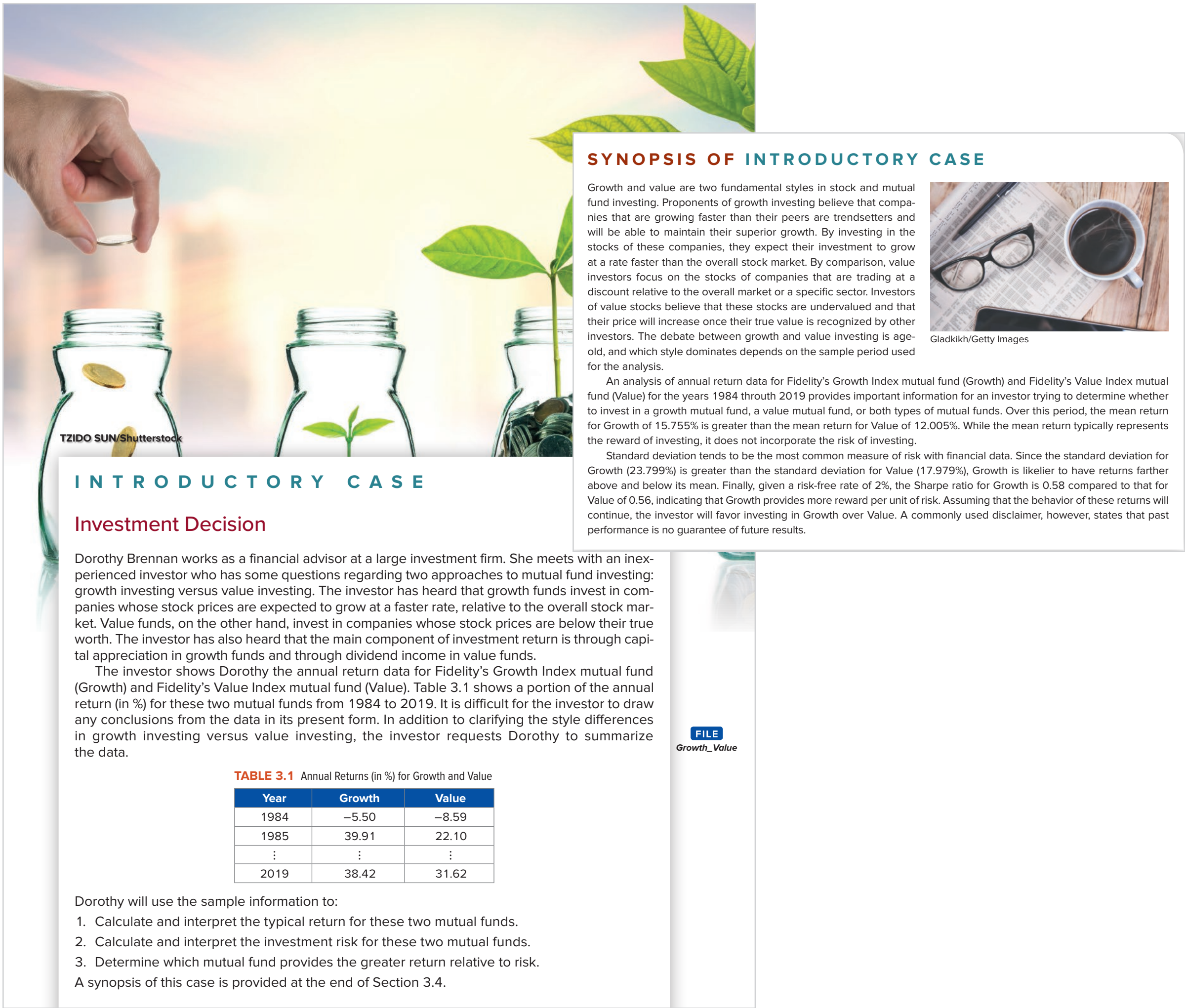
Storytelling: More relevant discussion

- Numerous new examples, exercises, and case studies
- Updated data to make the applications more current
- Big data used in the writing sections for Chapters 1, 2, 3, 15, 16, and 17
- Big data used for suggested case studies in several chapters

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.



TZIDO SUN/Shutterstock

INTRODUCTORY CASE

Investment Decision

Dorothy 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. Value funds, on the other hand, 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 Dorothy the annual return data for Fidelity's Growth Index mutual fund (Growth) and Fidelity's Value Index mutual fund (Value). Table 3.1 shows a portion of the annual return (in %) for these two mutual funds from 1984 to 2019. It is difficult for the investor to draw any conclusions from the data in its present form. In addition to clarifying the style differences in growth investing versus value investing, the investor requests Dorothy to summarize the data.

FILE *Growth_Value*

TABLE 3.1 Annual Returns (in %) for Growth and Value

Year	Growth	Value
1984	−5.50	−8.59
1985	39.91	22.10
⋮	⋮	⋮
2019	38.42	31.62

Dorothy will use the sample information to:

1. Calculate and interpret the typical return for these two mutual funds.
2. Calculate and interpret the investment risk for these two mutual funds.
3. Determine which mutual fund provides the greater return relative to risk.

A synopsis of this case is provided at the end of Section 3.4.

SYNOPSIS OF INTRODUCTORY CASE

Gladkikh/Getty Images

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.

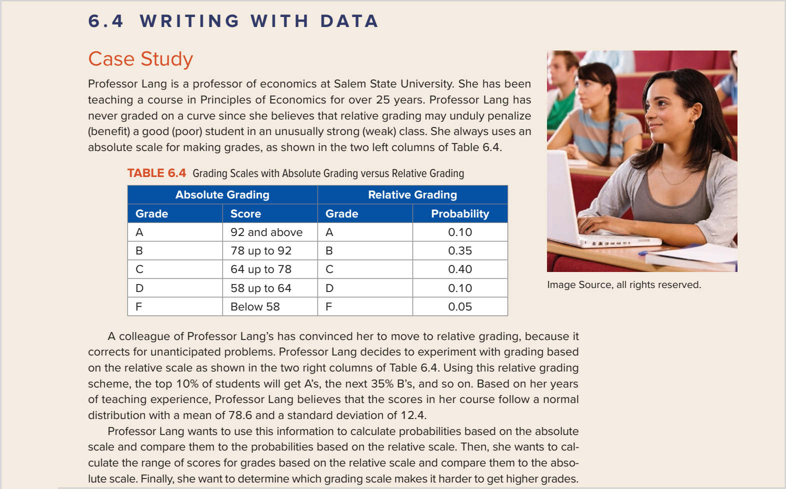
An analysis of annual return data for Fidelity's Growth Index mutual fund (Growth) and Fidelity's Value Index mutual fund (Value) for the years 1984 throuth 2019 provides important information for an investor trying to determine whether to invest in a growth mutual fund, a value mutual fund, or both types of mutual funds. Over this period, the mean return for Growth of 15.755% is greater than the mean return for Value of 12.005%. While the mean return typically represents the reward of investing, it does not incorporate the risk of investing.

Standard deviation tends to be the most common measure of risk with financial data. Since the standard deviation for Growth (23.799%) is greater than the standard deviation for Value (17.979%), Growth is likelier to have returns farther above and below its mean. Finally, given a risk-free rate of 2%, the Sharpe ratio for Growth is 0.58 compared to that for Value of 0.56, indicating that Growth provides more reward per unit of risk. Assuming that the behavior of these returns will continue, the investor will favor investing in Growth over Value. A commonly used disclaimer, however, states that past performance is no guarantee of future results.

and Build Skills to Communicate Results

Writing with Data

One of our most important innovations is the inclusion of a sample report within every chapter. 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 every Writing with Data section. 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.



6.4 WRITING WITH DATA

Case Study

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 6.4.

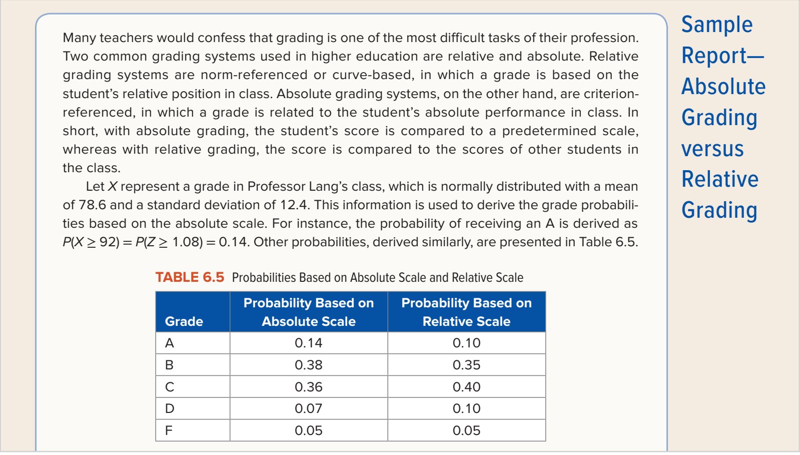
Image Source, all rights reserved.

TABLE 6.4 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, because 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 6.4. 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 this information to calculate probabilities based on the absolute scale and compare them to the probabilities based on the relative scale. Then, she wants to calculate the range of scores for grades based on the relative scale and compare them to the absolute scale. Finally, she want to 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 6.5.

TABLE 6.5 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 . . .

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 importance in the emerging field of data analytics. We combine simple and multiple regression in one chapter, which we believe is a seamless grouping and eliminates needless repetition. The detailed Excel and R instructions eliminate the need for tedious manual calculations. Three more in-depth chapters cover statistical inference, nonlinear relationships, dummy variables, and the linear probability and the logistic regression models. The emphasis in all chapters is on conceptualization, model selection, and interpretation of the results with reference to professionally created figures and tables.

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***

Inclusion of Important Topics

We have incorporated several important topics, often ignored in traditional textbooks, including data preparation, pivot tables, geometric mean return, Sharpe ratio, accuracy rates, etc. From our experience from working outside the classroom, we have found that professionals use these topics 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

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

where $\bar{x}_I$ is the mean return for the investment, $\overline{R}_f$ 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.

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. Use Excel or R to estimate the appropriate model,
3. Communicate the 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 and R

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 embedding computer output is often necessary in order to avoid cumbersome calculations or the need for statistical tables. Microsoft Excel and R are the primary software packages used in this text. We chose Excel and R over other statistical packages based on their widespread use and reviewer feedback. For instructors who prefer to focus only on Excel, the R instructions sections are easily skipped. We provide brief guidelines for using Minitab, SPSS, and JMP in chapter appendices.

Using Excel

We use Excel's **BINOM.DIST** function to calculate binomial probabilities. We enter =BINOM.DIST(*x, n, p,* TRUE or FALSE) where *x* is the number of successes, *n* is the number of trials, and *p* is the probability of success. For the last argument, we enter TRUE if we want to find the cumulative probability function $P(X \leq x)$ or FALSE if we want to find the probability mass function $P(X = x)$.

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, FALSE) 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, TRUE) 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, TRUE) and Excel returns 0.3784.

Using R

We use R's **dbinom** and **pbinom** functions to calculate binomial probabilities. In order to calculate the probability mass function $P(X = x)$, we enter dbinom(*x, n, p*) where *x* is the number of successes, *n* is the number of trials, and *p* is the probability of success. In order to calculate the cumulative probability function $P(X \leq x)$, we enter pbinom(*x, n, p*).

a. In order to find $P(X = 70)$, we enter:

```
> dbinom(70, 100, 0.68)
```

And R returns: 0.07907911.

b. In order to find $P(X \leq 70)$, we enter:

```
> pbinom(70, 100, 0.68)
```

And R returns: 0.7006736.

c. In order to find $P(X \geq 70) = 1 - P(X \leq 69)$, we enter:

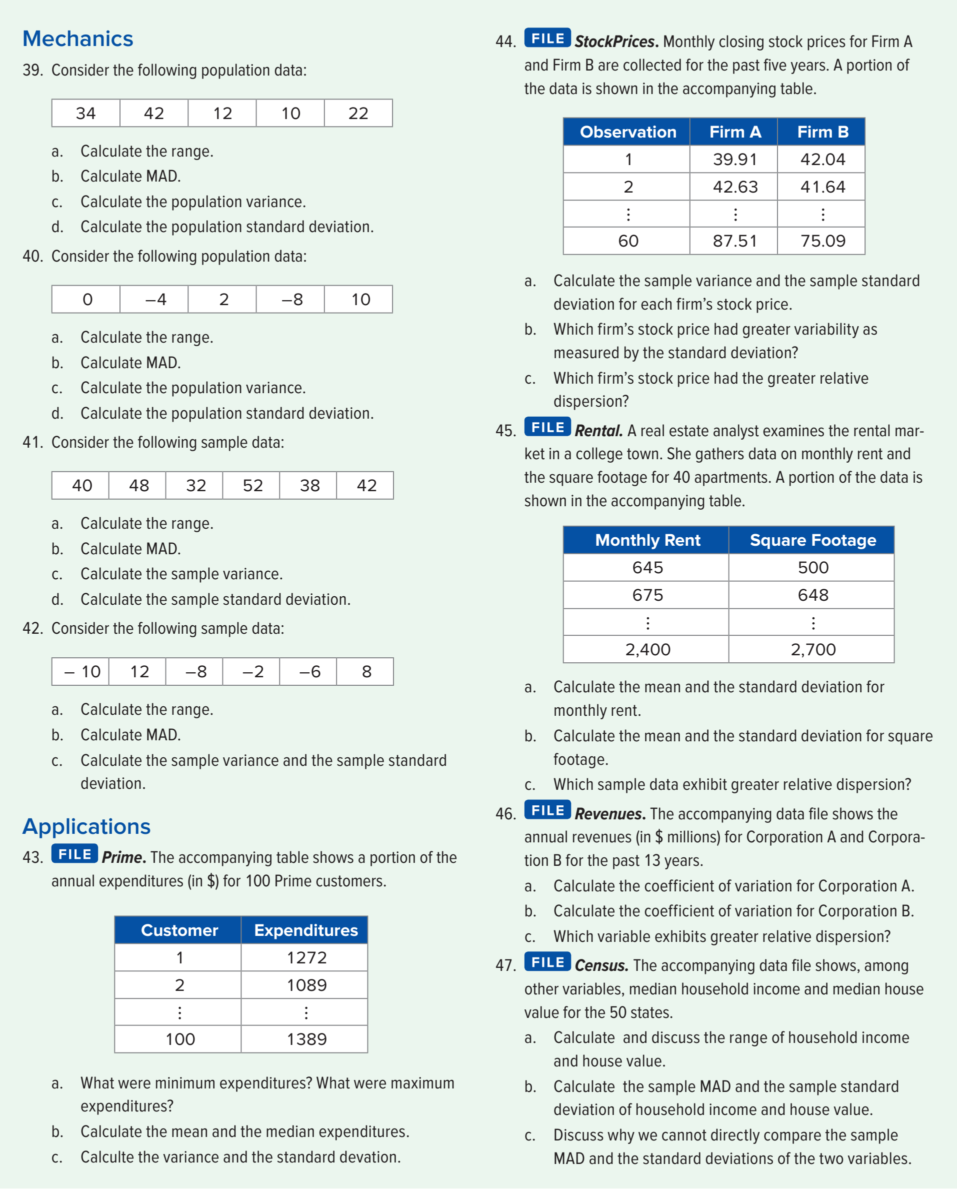
```
> 1 – pbinom(69, 100, 0.68)
```

And R returns: 0.3784055.

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, sales, personality types, health measures, expenditures, 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.



Mechanics

39. Consider the following population data:

34	42	12	10	22

a. Calculate the range.
b. Calculate MAD.
c. Calculate the population variance.
d. Calculate the population standard deviation.

40. Consider the following population data:

0	−4	2	−8	10

a. Calculate the range.
b. Calculate MAD.
c. Calculate the population variance.
d. Calculate the population standard deviation.

41. Consider the following sample data:

40	48	32	52	38	42

a. Calculate the range.
b. Calculate MAD.
c. Calculate the sample variance.
d. Calculate the sample standard deviation.

42. Consider the following sample data:

– 10	12	−8	−2	−6	8

a. Calculate the range.
b. Calculate MAD.
c. Calculate the sample variance and the sample standard deviation.

Applications

43. FILE ***Prime.*** The accompanying table shows a portion of the annual expenditures (in $) for 100 Prime customers.

Customer	Expenditures
1	1272
2	1089
⋮	⋮
100	1389

a. What were minimum expenditures? What were maximum expenditures?
b. Calculate the mean and the median expenditures.
c. Calculte the variance and the standard devation.

44. FILE ***StockPrices.*** Monthly closing stock prices for Firm A and Firm B are collected for the past five years. A portion of the data is shown in the accompanying table.

Observation	Firm A	Firm B
1	39.91	42.04
2	42.63	41.64
⋮	⋮	⋮
60	87.51	75.09

a. Calculate the sample variance and the sample standard deviation for each firm's stock price.
b. Which firm's stock price had greater variability as measured by the standard deviation?
c. Which firm's stock price had the greater relative dispersion?

45. FILE ***Rental.*** A real estate analyst examines the rental market in a college town. She gathers data on monthly rent and the square footage for 40 apartments. A portion of the data is shown in the accompanying table.

Monthly Rent	Square Footage
645	500
675	648
⋮	⋮
2,400	2,700

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?

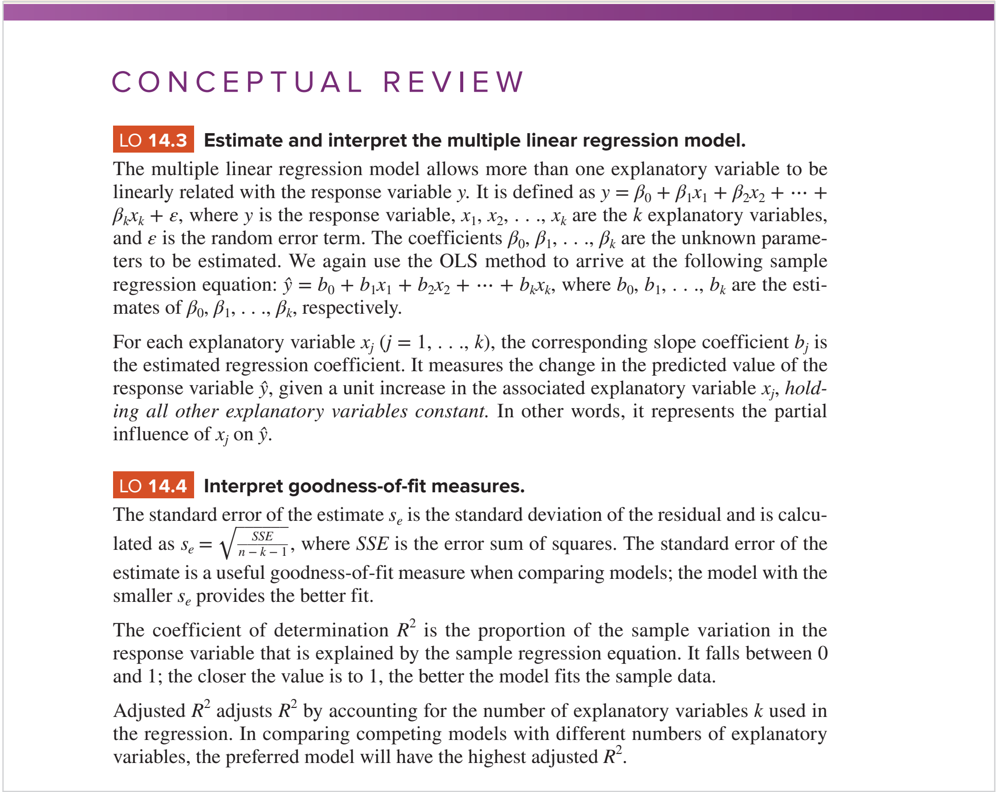
46. FILE ***Revenues.*** The accompanying data file shows the annual revenues (in $ millions) for Corporation A and Corporation B for the past 13 years.
a. Calculate the coefficient of variation for Corporation A.
b. Calculate the coefficient of variation for Corporation B.
c. Which variable exhibits greater relative dispersion?

47. FILE ***Census.*** The accompanying data file shows, among other variables, median household income and median house value for the 50 states.
a. Calculate and discuss the range of household income and house value.
b. Calculate the sample MAD and the sample standard deviation of household income and house value.
c. Discuss why we cannot directly compare the sample MAD and the standard deviations of the two variables.

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 14.3 Estimate and interpret the multiple linear regression model.

The multiple linear regression model allows more than one explanatory variable to be linearly related with the response variable y. It is defined as $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$, where y is the response variable, $x_1, x_2, \ldots, x_k$ are the k explanatory variables, and ε is the random error term. The coefficients $\beta_0, \beta_1, \ldots, \beta_k$ are the unknown parameters to be estimated. We again use the OLS method to arrive at the following sample regression equation: $\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k$, where $b_0, b_1, \ldots, b_k$ are the estimates of $\beta_0, \beta_1, \ldots, \beta_k$, respectively.

For each explanatory variable x_j ($j = 1, \ldots, k$), the corresponding slope coefficient b_j is the estimated regression coefficient. It measures the change in the predicted value of the response variable $\hat{y}$, given a unit increase in the associated explanatory variable x_j, *holding all other explanatory variables constant.* In other words, it represents the partial influence of x_j on $\hat{y}$.

LO 14.4 Interpret goodness-of-fit measures.

The standard error of the estimate s_e is the standard deviation of the residual and is calculated as $s_e = \sqrt{\frac{SSE}{n-k-1}}$, where SSE is the error sum of squares. The standard error of the estimate is a useful goodness-of-fit measure when comparing models; the model with the smaller s_e provides the better fit.

The coefficient of determination R^2 is the proportion of the sample variation in the response variable that is explained by the sample regression equation. It falls between 0 and 1; the closer the value is to 1, the better the model fits the sample data.

Adjusted R^2 adjusts R^2 by accounting for the number of explanatory variables k used in the regression. In comparing competing models with different numbers of explanatory variables, the preferred model will have the highest adjusted R^2.

McGraw Hill connect®

Instructors: Student Success Starts with You

Tools to enhance your unique voice

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

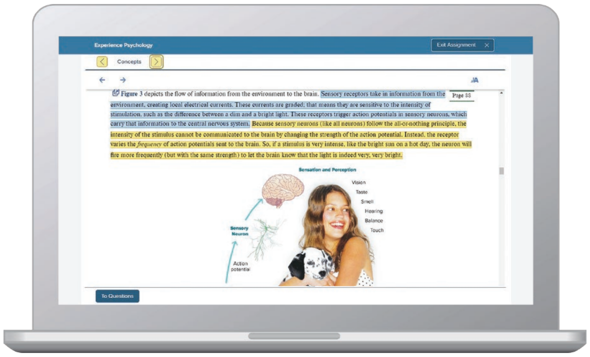


Figure 3 depicts the flow of information from the environment to the brain. Sensory receptors take in information from the environment, creating local electrical currents. These currents are graded; that means they are sensitive to the intensity of stimulation, such as the difference between a dim and a bright light. These receptors trigger action potentials in sensory neurons, which carry that information to the central nervous system. Because sensory neurons (like all neurons) follow the all-or-nothing principle, the intensity of the stimulus cannot be communicated to the brain by changing the strength of the action potential. Instead, the receptor varies the *frequency* of action potentials sent to the brain. So, if a stimulus is very intense, like the bright sun on a hot day, the neuron will fire more frequently (but with the same strength) to let the brain know that the light is indeed very, very bright.

Laptop: McGraw Hill; Woman/dog: George Doyle/Getty Images

Study made personal

Incorporate adaptive study resources like SmartBook® 2.0 into your course and help your students be better prepared in less time. Learn more about the powerful personalized learning experience available in SmartBook 2.0 at **www.mheducation.com/highered/connect/smartbook**

Affordable solutions, added value

Make technology work for you with LMS integration for single sign-on access, mobile access to the digital textbook, and reports to quickly show you how each of your students is doing. And with our Inclusive Access program you can provide all these tools at a discount to your students. Ask your McGraw Hill representative for more information.

Padlock: Jobalou/Getty Images

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. Visit **www.supportateverystep.com** for videos and resources both you and your students can use throughout the semester.

Checkmark: Jobalou/Getty Images

SUPPORT AT every step

Students: Get Learning that Fits You

Effective tools for efficient studying

Connect is designed to make you more productive with simple, flexible, intuitive tools that maximize your study time and meet your individual learning needs. Get learning that works for you with Connect.

Study anytime, anywhere

Download the free ReadAnywhere app and access your online eBook or SmartBook 2.0 assignments when it's convenient, even if you're offline. And since the app automatically syncs with your eBook and SmartBook 2.0 assignments in Connect, all of your work is available every time you open it. Find out more at **www.mheducation.com/readanywhere**

> ***"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

Calendar: owattaphotos/Getty Images

Everything you need in one place

Your Connect course has everything you need—whether reading on your digital eBook or completing assignments for class, Connect makes it easy to get your work done.

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/about/accessibility** for more information.

Top: Jenner Images/Getty Images, Left: Hero Images/Getty Images, Right: Hero Images/Getty Images

Remote Proctoring & Browser-Locking Capabilities

New remote proctoring and browser-locking capabilities, hosted by Proctorio within Connect, provide control of the assessment environment by enabling security options and verifying the identity of the student.

Seamlessly integrated within Connect, these services allow instructors to control students' assessment experience by restricting browser activity, recording students' activity, and verifying students are doing their own work.

Instant and detailed reporting gives instructors an at-a-glance view of potential academic integrity concerns, thereby avoiding personal bias and supporting evidence-based claims.

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, visit 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, 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.

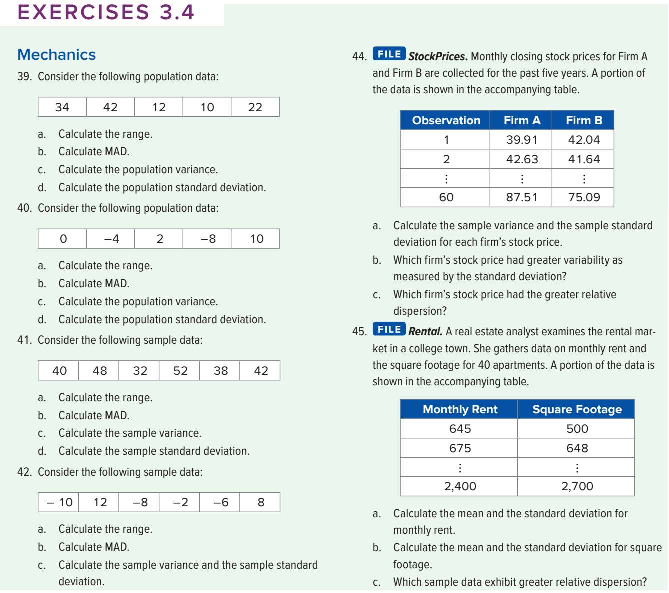
ALEKS also includes an instructor module with powerful, assignment-driven features and extensive content flexibility. ALEKS simplifies course management and allows instructors to spend less time with administrative tasks and more time directing student learning. To learn more about ALEKS, visit www.aleks.com.

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.



EXERCISES 3.4

Mechanics

39. Consider the following population data:

34	42	12	10	22

a. Calculate the range.
b. Calculate MAD.
c. Calculate the population variance.
d. Calculate the population standard deviation.

40. Consider the following population data:

0	−4	2	−8	10

a. Calculate the range.
b. Calculate MAD.
c. Calculate the population variance.
d. Calculate the population standard deviation.

41. Consider the following sample data:

40	48	32	52	38	42

a. Calculate the range.
b. Calculate MAD.
c. Calculate the sample variance.
d. Calculate the sample standard deviation.

42. Consider the following sample data:

− 10	12	−8	−2	−6	8

a. Calculate the range.
b. Calculate MAD.
c. Calculate the sample variance and the sample standard deviation.

44. FILE ***StockPrices.*** Monthly closing stock prices for Firm A and Firm B are collected for the past five years. A portion of the data is shown in the accompanying table.

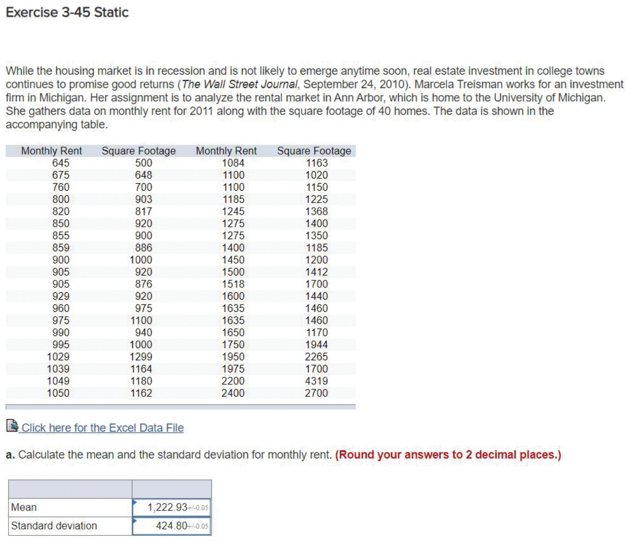
Observation	Firm A	Firm B
1	39.91	42.04
2	42.63	41.64
⋮	⋮	⋮
60	87.51	75.09

a. Calculate the sample variance and the sample standard deviation for each firm's stock price.
b. Which firm's stock price had greater variability as measured by the standard deviation?
c. Which firm's stock price had the greater relative dispersion?

45. FILE ***Rental.*** A real estate analyst examines the rental market in a college town. She gathers data on monthly rent and the square footage for 40 apartments. A portion of the data is shown in the accompanying table.

Monthly Rent	Square Footage
645	500
675	648
⋮	⋮
2,400	2,700

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?



Exercise 3-45 Static

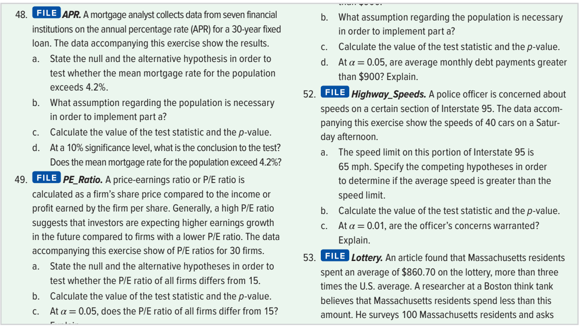
While the housing market is in recession and is not likely to emerge anytime soon, real estate investment in college towns continues to promise good returns (*The Wall Street Journal*, September 24, 2010). Marcela Treisman works for an investment firm in Michigan. Her assignment is to analyze the rental 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. The data is shown in the accompanying table.

Monthly Rent	Square Footage	Monthly Rent	Square Footage
645	500	1084	1163
675	648	1100	1020
760	700	1100	1150
800	903	1185	1225
820	817	1245	1368
850	920	1275	1400
855	900	1275	1350
859	886	1400	1185
900	1000	1450	1200
905	920	1500	1412
905	876	1518	1700
929	920	1600	1440
960	975	1635	1460
975	1100	1635	1460
990	940	1650	1170
995	1000	1750	1944
1029	1299	1950	2265
1039	1164	1975	1700
1049	1180	2200	4319
1050	1162	2400	2700

Click here for the Excel Data File

a. Calculate the mean and the standard deviation for monthly rent. **(Round your answers to 2 decimal places.)**

Mean	1,222.93 +/-0.05
Standard deviation	424.80 +/-0.05



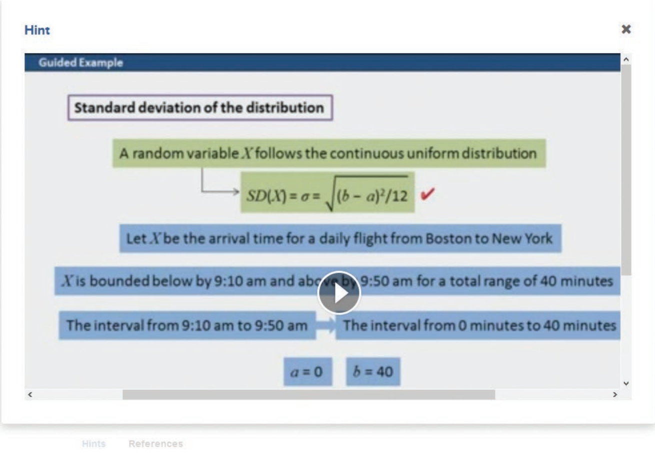
48. FILE ***APR.*** A mortgage analyst collects data from seven financial institutions on the annual percentage rate (APR) for a 30-year fixed loan. The data accompanying this exercise show the results.
a. State the null and the alternative hypothesis in order to test whether the mean mortgage rate for the population exceeds 4.2%.
b. What assumption regarding the population is necessary in order to implement part a?
c. Calculate the value of the test statistic and the *p*-value.
d. At a 10% significance level, what is the conclusion to the test? Does the mean mortgage rate for the population exceed 4.2%?

49. FILE ***PE_Ratio.*** A price-earnings ratio or P/E ratio is calculated as a firm's share price compared to the income or profit earned by the firm per share. Generally, a high P/E ratio suggests that investors are expecting higher earnings growth in the future compared to firms with a lower P/E ratio. The data accompanying this exercise show of P/E ratios for 30 firms.
a. State the null and the alternative hypotheses in order to test whether the P/E ratio of all firms differs from 15.
b. Calculate the value of the test statistic and the *p*-value.
c. At $\alpha = 0.05$, does the P/E ratio of all firms differ from 15?

b. What assumption regarding the population is necessary in order to implement part a?
c. Calculate the value of the test statistic and the *p*-value.
d. At $\alpha = 0.05$, are average monthly debt payments greater than $900? Explain.

52. FILE ***Highway_Speeds.*** A police officer is concerned about speeds on a certain section of Interstate 95. The data accompanying this exercise show the speeds of 40 cars on a Saturday afternoon.
a. The speed limit on this portion of Interstate 95 is 65 mph. Specify the competing hypotheses in order to determine if the average speed is greater than the speed limit.
b. Calculate the value of the test statistic and the *p*-value.
c. At $\alpha = 0.01$, are the officer's concerns warranted? Explain.

53. FILE ***Lottery.*** An article found that Massachusetts residents spent an average of $860.70 on the lottery, more than three times the U.S. average. A researcher at a Boston think tank believes that Massachusetts residents spend less than this amount. He surveys 100 Massachusetts residents and asks



Hint

Guided Example

Standard deviation of the distribution

A random variable X follows the continuous uniform distribution

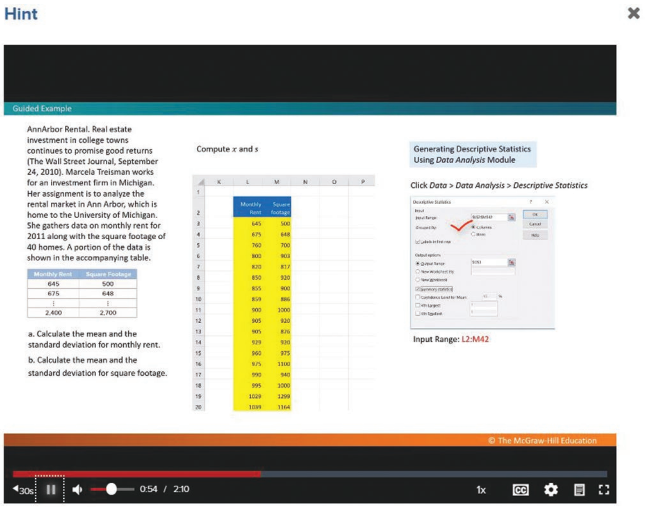
$SD(X) = \sigma = \sqrt{(b-a)^2/12}$ ✓

Let X be the arrival time for a daily flight from Boston to New York

X is bounded below by 9:10 am and above by 9:50 am for a total range of 40 minutes

The interval from 9:10 am to 9:50 am → The interval from 0 minutes to 40 minutes

$a = 0$ $b = 40$



Hint

Guided Example

AnnArbor Rental. Real estate investment in college towns continues 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
⋮	⋮
2,400	2,700

a. Calculate the mean and the standard deviation for monthly rent.

b. Calculate the mean and the standard deviation for square footage.

Compute x and s

	K	L	M	N	O	P
1						
2		Monthly Rent	Square footage			
3		645	500			
4		675	648			
5		760	700			
6		800	903			
7		820	817			
8		850	920			
9		855	900			
10		859	886			
11		900	1000			
12		905	920			
13		905	876			
14		929	920			
15		960	975			
16		975	1100			
17		990	940			
18		995	1000			
19		1029	1299			
20		1039	1164			

Generating Descriptive Statistics Using *Data Analysis* Module

Click *Data* > *Data Analysis* > *Descriptive Statistics*

Input Range: L2:M42

© The McGraw-Hill Education

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*
Kathryn Ernstberger *Indiana University*
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*
Sheila Diann Hammon *Athens 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*
Yogesh Joshi *Kingsborough Community College*
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*
Brent Marinan *University of Arizona*
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*
Melissa Patterson *Chabot College*
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*
Sue Schou *Boise 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*
Anne-Louise Statt *University of Michigan–Dearborn*
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*
Qiang Zhen *University of North Florida*
Wencang Zhou *Baruch College*
Zhen Zhu *University of Central Oklahoma*

The editorial staff of McGraw Hill are deserving of our gratitude for their guidance throughout this project, especially Noelle Bathurst, Pat Frederickson, Ryan McAndrews, Harper Christopher, Jamie Koch, and Matt Diamond.

We would also like to thank Jenna Eisenman, Matthew Hawke, and Megan Hawke for their outstanding research assistance.

BRIEF CONTENTS

CONTENTS

PART FIVE
Advanced Inference

PART SIX

Supplementary Topics

BUSINESS STATISTICS

1 Data and Data Preparation

LEARNING OBJECTIVES

After reading this chapter, you should be able to:

LO **1.1** **Explain the various data types.**

LO **1.2** **Describe variables and types of measurement scales.**

LO **1.3** **Inspect and explore data.**

LO **1.4** **Apply data subsetting.**

In just about any contemporary human activity, we use statistics to analyze large amounts of data for making better decisions. Managers, consumers, sports enthusiasts, politicians, and medical professionals are increasingly turning to data to boost a company's revenue, deepen customer engagement, find better options on consumer products, prevent threats and fraud, succeed in sports and elections, provide better diagnoses and cures for diseases, and so on. In this chapter, we will describe various types of data and measurement scales of variables that are used in statistics.

It is important to note that after obtaining relevant data, we often spend a considerable amount of time on inspecting and preparing the data for subsequent analysis. In this chapter, we will discuss a few important data preparation tasks. We will use counting and sorting of relevant variables to inspect and explore data. Finally, we will discuss a commonly used technique called subsetting, where only a portion (subset) of the data is used for the analysis.

Yuliia Mazurkevych/Shutterstoick

INTRODUCTORY CASE

Gaining Insights into Retail Customer Data

Organic Food Superstore is an online grocery store that specializes in providing organic food products to health-conscious consumers. The company offers a membership-based service that ships fresh ingredients for a wide range of chef-designed meals to its members' homes. Catherine Hill is a marketing manager at Organic Food Superstore. She has been assigned to market the company's new line of Asian-inspired meals. Research has shown that the most likely customers for healthy ethnic cuisines are college-educated millennials (born between 1982 and 2000).

In order to spend the company's marketing dollars efficiently, Catherine wants to focus on this target demographic when designing the marketing campaign. With the help of the information technology (IT) group, Catherine has acquired a representative sample that includes each customer's identification number (CustID), sex (Sex), race (Race), birthdate (BirthDate), whether the customer has a college degree (College), household size (HHSize), annual income (Income), total spending (Spending), total number of orders during the past 24 months (Orders), and the channel through which the customer was originally acquired (Channel). Table 1.1 shows a portion of the data set.

TABLE 1.1 A Sample of Organic Food Superstore Customers

CustID	Sex	Race	BirthDate	. . .	Channel
1530016	Female	Black	12/16/1986	. . .	SM
1531136	Male	White	5/9/1993	. . .	TV
⋮	⋮	⋮	⋮	⋮	⋮
1579979	Male	White	7/5/1999	. . .	SM

FILE
Customers

Catherine wants to use the Customers data set to:

1. Identify Organic Food Superstore's college-educated millennial customers.
2. Compare the profiles of female and male college-educated millennial customers.

A synopsis of this case is provided at the end of Section 1.3.

LO 1.1

1.1 TYPES OF DATA

Explain the various types of data.

In general, data are compilations of facts, figures, or other contents, both numerical and nonnumerical. Data of all types and formats are generated from multiple sources. Insights from all of these data improves a company's bottom-line and enhances consumer experience. At the core, business statistics benefits companies by developing better marketing strategies, deepening customer engagement, enhancing efficiency in procurement, uncovering ways to reduce expenses, identifying emerging market trends, mitigating risk and fraud, etc. We often find a large amount of data at our disposal. However, we also derive insights from relatively small data sets, such as from consumer focus groups, marketing surveys, or reports from government agencies.

Every day, consumers and businesses use data from various sources to help make decisions. In order to make intelligent decisions in a world full of uncertainty, we have to understand statistics—the language of data. In the broadest sense, statistics is the science of extracting useful information from data. Three steps are essential for doing good statistics. An important first step for making decisions is to find the right data, which are both complete and lacking any misrepresentation, and prepare it for the analysis. 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 information into verbal and written language. It is important to note that numerical results are not very useful unless they are accompanied with clearly stated actionable business insights.

DATA AND STATISTICS

Data are compilations of facts, figures, or other contents, both numerical and nonnumerical. Statistics is the science that deals with the collection, preparation, analysis, interpretation, and presentation of data.

In the introductory case, Catherine wants to target college-educated millennials when designing the marketing campaign so that she spends the company's marketing dollars efficiently. Before we analyze the information that Catherine has gathered, it is important to understand different types of data and measurement scales of variables. In this section, we focus on data types.

LO 1.1

Sample and Population Data

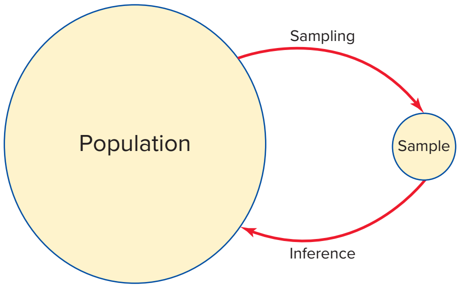
Explain the various types of data.

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 the data by providing, for example, the typical value and the variability of the variable of interest. 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.

Figure 1.1 depicts the flow of information between a population and a sample. Consider, for example, a 2016 Gallop survey that found that only 50% of millennials plan to stay at 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.

FIGURE 1.1 Population versus Sample



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.

It is generally not feasible to obtain population data due to prohibitive costs and/or practicality. We rely on sampling because 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.

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.

Table 1.2 is an example of a cross-sectional data set. It lists the team standings for the National Basketball Association's Eastern Conference at the end of the 2018–2019 season. The eight teams may not have ended the season precisely on the same day and time, but the differences in time are of no relevance in this example. Other examples of cross-sectional data include the recorded grades of students in a class, the sale prices of single-family homes sold last month, the current price of gasoline in different cities in the United States, and the starting salaries of recent business graduates from the University of Connecticut.

TABLE 1.2 2018–2019 NBA Eastern Conference

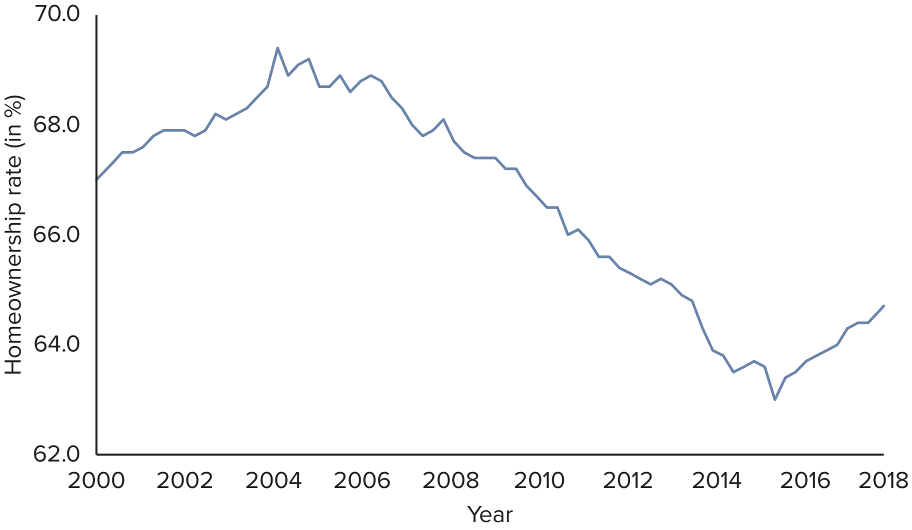
Team name	Wins	Losses	Winning percentage
Milwaukee Bucks	60	22	0.732
Toronto Raptors*	58	24	0.707
Philadephia 76ers	51	31	0.622
Boston Celtics	49	33	0.598
Indiana Pacers	48	34	0.585
Brooklyn Nets	42	40	0.512
Orlando Magic	42	40	0.512
Detroit Pistons	41	41	0.500

*The Toronto Raptors won their first NBA title during the 2018–2019 season.

Time series data refer to data collected over several time periods focusing on certain groups of people, specific events, or objects. Time series data 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 2020, 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 2020, and the annual population growth rate of India in the last decade. In these examples, temporal ordering is relevant and meaningful.

Figure 1.2 shows a plot of the national homeownership rate in the U.S. from 2000 to 2018. 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 explanation for the decline in the homeownership rate is the stricter lending practices caused by the housing market crash in 2007 that precipitated a banking crisis and deep recession. This decline can also be attributed to home prices outpacing wages in the sample period.

FIGURE 1.2
Homeownership Rate (in %) in the U.S. from 2000 through 2018



Structured and Unstructured Data

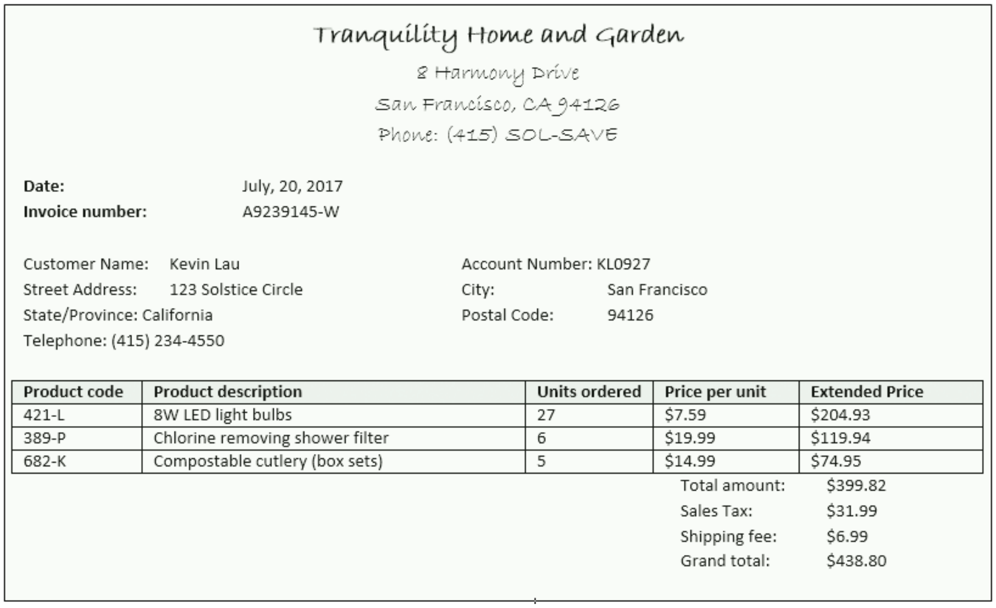
When you think of data, the first image that probably pops in your head is lots of numbers and perhaps some charts and graphs. In reality, data can come in multiple forms. For example, information exchange in social networking websites such as Facebook, LinkedIn, and Twitter also constitute data. In order to better understand the various forms of data, we make a distinction between structured and unstructured data.

Generally, **structured data** reside in a predefined, row-column format. We use spreadsheet or database applications to enter, store, query, and analyze structured data.

Examples of structured data include numbers, dates, and groups of words and numbers, typically stored in a tabular format. Structured data often consist of numerical information that is objective and is not open to interpretation.

Point-of-sale and financial data are examples of structured data and are usually designed to capture a business process or transaction. Examples include the sale of retail products, money transfer between bank accounts, and the student enrollment in a university course. When individual consumers buy products from a retail store, each transaction is captured into a record of structured data.

Consider the sales invoice shown in Figure 1.3. Whenever a customer places an order like this, there is a predefined set of data to be collected, such as the transaction date, shipping address, and the units of product being purchased. Even though a receipt or an invoice may not always be presented in rows and columns, the predefined structure allows businesses and organizations to translate the data on the document into a row-column format.



Tranquility Home and Garden
8 Harmony Drive
San Francisco, CA 94126
Phone: (415) SOL-SAVE

Date: July, 20, 2017
Invoice number: A9239145-W

Customer Name: Kevin Lau
Street Address: 123 Solstice Circle
State/Province: California
Telephone: (415) 234-4550

Account Number: KL0927
City: San Francisco
Postal Code: 94126

Product code	Product description	Units ordered	Price per unit	Extended Price
421-L	8W LED light bulbs	27	$7.59	$204.93
389-P	Chlorine removing shower filter	6	$19.99	$119.94
682-K	Compostable cutlery (box sets)	5	$14.99	$74.95

Total amount:	$399.82
Sales Tax:	$31.99
Shipping fee:	$6.99
Grand total:	$438.80

FIGURE 1.3 A sample invoice from a retail transaction

For decades, companies and organizations relied mostly on structured data to run their businesses and operations. Today, with the advent of the digital age, most experts agree that only about 20% of all data used in business decisions are structured data. The remaining 80% are unstructured.

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, e-mail's subject line, or a time stamp on a photograph), they are still considered unstructured as they do not conform to a row-column model required in most database systems. Social media data such as Twitter, YouTube, Facebook, 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 catch-phrase, meaning a massive amount 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 gain knowledge and business intelligence with potential game-changing impacts on company revenues, competitive advantage, and organizational efficiency.

More formally, a widely accepted definition of big data is “high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation” (www.gartner.com). The three characteristics (the three Vs) of big data are:

- **Volume:** An immense amount of data is compiled from a single source or a wide range of sources, including business transactions, household and personal devices, manufacturing equipment, social media, and other online portals.
- **Velocity:** In addition to volume, data from a variety of sources get generated at a rapid speed. Managing these data streams can become a critical issue for many organizations.
- **Variety:** Data also come in all types, forms, and granularity, both structured and unstructured. These data may include numbers, text, and figures as well as audio, video, e-mails, and other multimedia elements.

In addition to the three defining characteristics of big data, we also need to pay close attention to the veracity of the data and the business value that they can generate. **Veracity** refers to the credibility and quality of data. One must verify the reliability and accuracy of the data content prior to relying on the data to make decisions. This becomes increasingly challenging with the rapid growth of data volume fueled by social media and automatic data collection. **Value** derived from big data is perhaps the most important aspect of any statistical project. Having a plethora of data does not guarantee that useful insights or measurable improvements will be generated. Organizations must develop a methodical plan for formulating business questions, curating the right data, and unlocking the hidden potential in big data.

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 users 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 portals that consumers use. And because different social media are used for different purposes, data collected from these sites may very well reflect different populations of Internet users; this distinction is especially important from a strategic business standpoint. Therefore, Facebook data are simply a very large sample.

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

STRUCTURED, UNSTRUCTURED, AND BIG DATA

Structured data are data that reside in a predefined, row-column format, while unstructured data do not conform to a predefined, row-column format. Big data is a term used to describe a massive amount 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 necessary imply complete (population) data.

In this textbook, we will focus on traditional statistical methods applied to structured data. Sophisticated tools to analyze unstructured data are beyond the scope of this textbook.

Data on the Web

Comstock Images/Jupiterimages

The explosion in the field of statistics and data analytics is partly due to the growing availability of vast amounts of data and improved computational power. Many experts believe that 90% of the data in the world today were created in the last two years alone. These days, it has become easy to access data 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 1.1

Applications

1. A few years ago, it came as a surprise when Apple's iPhone 4 was found to have a problem. Users complained of weak reception, and sometimes even dropped calls, when they cradled the phone in their hands in a particular way. A survey at a local store found that 2% of iPhone 4 users experienced this reception problem.
   a. Describe the relevant population.
   b. Is 2% associated with the population or the sample?
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 old. 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. Describe the relevant population.
4. Recent college graduates with an engineering degree continue to earn high salaries. An online search revealed that the average annual salary for an entry-level position in engineering is $65,000.
   a. What is the relevant population?
   b. Do you think the average salary of $65,000 is computed from the population? Explain.
5. Research suggests that depression significantly increases the risk of developing dementia later in life. Suppose that in a study involving 949 elderly persons, it was found 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. Are the numbers 22% and 17% associated with the population or a sample?