Essentials of Business Statistics

Bowerman • O'Connell • Murphree • Orris

5e

Bruce L. Bowerman
Miami University
Richard T. O'Connell
Miami University
Emily S. Murphree
Miami University
J. B. Orris
Butler University

Essentials of Business Statistics

FIFTH EDITION

with major contributions by
Steven C. Huchendorf
University of Minnesota
Dawn C. Porter
University of Southern California
Patrick J. Schur
Miami University

Mc Graw Hill Education

ESSENTIALS OF BUSINESS STATISTICS, FIFTH 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 4

ISBN 978-0-07-802053-7
MHID 0-07-802053-0

Senior Vice President, Products & Markets: *Kurt L. Strand*
Vice President, Content Production & Technology Services: *Kimberly Meriwether David*
Managing Director: *Douglas Reiner*
Senior Brand Manager: *Thomas Hayward*
Executive Director of Development: *Ann Torbert*
Senior Development Editor: *Wanda J. Zeman*
Senior Marketing Manager: *Heather A. Kazakoff*
Director, Content Production: *Terri Schiesl*
Content Project Manager: *Harvey Yep*
Content Project Manager: *Daryl Horrocks*
Senior Buyer: *Debra R. Sylvester*
Design: *Matthew Baldwin*
Cover Image: © *Bloomberg via Getty Images*
Lead Content Licensing Specialist: *Keri Johnson*
Typeface: *10/12 Times New Roman*
Compositor: *MPS Limited*
Printer: *R. R. Donnelley*

The CIP data for this title has been applied for.

About the Authors

Bruce L. Bowerman Bruce L. Bowerman is professor emeritus of decision sciences at Miami University in Oxford, Ohio. He received his Ph.D. degree in statistics from Iowa State University in 1974, and he has over 41 years of experience teaching basic statistics, regression analysis, time series forecasting, survey sampling, and design of experiments to both undergraduate and graduate students. In 1987 Professor Bowerman received an Outstanding Teaching award from the Miami University senior class, and in 1992 he received an Effective Educator award from the Richard T. Farmer School of Business Administration. Together with Richard T. O'Connell, Professor Bowerman has written 20 textbooks. In his spare time, Professor Bowerman enjoys watching movies and sports, playing tennis, and designing houses.

Richard T. O'Connell Richard T. O'Connell is professor emeritus of decision sciences at Miami University in Oxford, Ohio. He has more than 36 years of experience teaching basic statistics, statistical quality control and process improvement, regression analysis, time series forecasting, and design of experiments to both undergraduate and graduate business students. He also has extensive consulting experience and has taught workshops dealing with statistical process control and process improvement for a variety of companies in the Midwest. In 2000 Professor O'Connell received an Effective Educator award from the Richard T. Farmer School of Business Administration. Together with Bruce L. Bowerman, he has written 20 textbooks. In his spare time, Professor O'Connell enjoys fishing, collecting 1950s and 1960s rock music, and following the Green Bay Packers and Purdue University sports.

Emily S. Murphree Emily S. Murphree is associate professor of statistics in the Department of Mathematics and Statistics at Miami University in Oxford, Ohio. She received her Ph.D. degree in statistics from the University of North Carolina and does research in applied probability. Professor Murphree received Miami's College of Arts and Science Distinguished Educator Award in 1998. In 1996, she was named one of Oxford's Citizens of the Year for her work with Habitat for Humanity and for organizing annual Sonia Kovalevsky Mathematical Sciences Days for area high school girls. Her enthusiasm for hiking in wilderness areas of the West motivated her current research on estimating animal population sizes.

James Burdeane "Deane" Orris J. B. Orris is a professor emeritus of management science at Butler University in Indianapolis, Indiana. He received his Ph.D. from the University of Illinois in 1971, and in the late 1970s with the advent of personal computers, he combined his interest in statistics and computers to write one of the first personal computer statistics packages—MICROSTAT. Over the past 20 years, MICROSTAT has evolved into MegaStat which is an Excel add-in statistics program. He wrote an Excel book, *Essentials: Excel 2000 Advanced,* in 1999 and *Basic Statistics Using Excel and MegaStat* in 2006. He taught statistics and computer courses in the College of Business Administration of Butler University from 1971 until 2013. He is a member of the American Statistical Association and is past president of the Central Indiana Chapter. In his spare time, Professor Orris enjoys reading, working out, and working in his woodworking shop.

FROM THE

In *Essentials of Business Statistics, Fifth Edition,* we provide a modern, practical, and unique framework for teaching an introductory course in business statistics. As in previous editions, we employ real or realistic examples, continuing case studies, and a business improvement theme to teach the material. Moreover, we believe that this fifth edition features more concise and lucid explanations, an improved topic flow, and a judicious use of realistic and compelling examples. Overall, the fifth edition is 32 pages shorter than the fourth edition while covering all previous material as well as additional topics. Below we outline the attributes and new features we think make this book an effective learning tool.

- **Continuing case studies that tie together different statistical topics.** These continuing case studies span not only individual chapters but also groups of chapters. Students tell us that when new statistical topics are developed in the context of familiar cases, their "fear factor" is reduced. Of course, to keep the examples from becoming overtired, we introduce new case studies throughout the book.
- **Business improvement conclusions that explicitly show how statistical results lead to practical business decisions.** After appropriate analysis and interpretation, examples and case studies often result in a business improvement conclusion. To emphasize this theme of business improvement, icons BI are placed in the page margins to identify when statistical analysis has led to an important business conclusion. The text of each conclusion is also highlighted in yellow for additional clarity.
- **Examples exploited to motivate an intuitive approach to statistical ideas.** Most concepts and formulas, particularly those that introductory students find most challenging, are first approached by working through the ideas in accessible examples. Only after simple and clear analysis within these concrete examples are more general concepts and formulas discussed.
- **An improved introduction to business statistics in Chapter 1.** The example introducing data and how data can be used to make a successful offer to purchase a house has been made clearer, and two new and more graphically oriented examples have been added to better introduce quantitative and qualitative variables. Random sampling is introduced informally in the context of more tightly focused case studies. [The technical discussion about how to select random samples and other types of samples is in Chapter 7 (Sampling and Sampling Distributions), but the reader has the option of reading about sampling in Chapter 7 immediately after Chapter 1.] Chapter 1 also includes a new discussion of ethical guidelines for practitioners of statistics. Throughout the book, statistics is presented as a broad discipline requiring not simply analytical skills but also judgment and personal ethics.
- **A more streamlined discussion of the graphical and numerical methods of descriptive statistics.** Chapters 2 and 3 utilize several new examples, including an example leading off Chapter 2 that deals with college students' pizza brand preferences. In addition, the explanations of some of the more complicated topics have been simplified. For example, the discussion of percentiles, quartiles, and box plots has been shortened and clarified.
- **An improved, well-motivated discussion of probability and probability distributions in Chapters 4, 5, and 6.** In Chapter 4, methods for calculating probabilities are more clearly motivated in the context of two new examples. We use the Crystal Cable Case, which deals with studying cable television and Internet penetration rates, to illustrate many probabilistic concepts and calculations. Moreover, students' understanding of the important concepts of conditional probability and statistical independence is sharpened by a new real-world case involving gender discrimination at a pharmaceutical company. The probability distribution, mean, and standard deviation of a discrete random variable are all motivated and explained in a more succinct discussion in Chapter 5. An example illustrates how knowledge of a mean and standard deviation are enough to estimate potential investment returns. Chapter 5 also features an improved introduction to the binomial distribution where the previous careful discussion is supplemented by an illustrative tree diagram. Students can now see the origins of all the factors in the binomial formula more clearly. Chapter 5 ends with a new optional section where joint probabilities and covariances are explained in the context of portfolio diversification. In Chapter 6, continuous probabilities are developed by improved examples. The coffee temperature case introduces the key ideas and is eventually used to help study the normal distribution. Similarly, the elevator waiting time case is used to explore the continuous uniform distribution.

AUTHORS

- **An improved discussion of sampling distributions and statistical inference in Chapters 7 through 12.** In Chapter 7, the discussion of sampling distributions has been modified to more seamlessly move from a small population example involving sampling car mileages to a related large population example. The introduction to confidence intervals in Chapter 8 features a very visual, graphical approach that we think makes finding and interpreting confidence intervals much easier. This chapter now also includes a shorter and clearer discussion of the difference between a confidence interval and a tolerance interval and concludes with a new section about estimating parameters of finite populations. Hypothesis testing procedures (using both the critical value and p-value approaches) are summarized efficiently and visually in summary boxes that are much more transparent than traditional summaries lacking visual prompts. These summary boxes are featured throughout the chapter covering inferences for one mean, one proportion, and one variance (Chapter 9), and the chapter covering inferences for two means, two proportions, and two variances (Chapter 10), as well as in later chapters covering regression analysis. In addition, the discussion of formulating the null and alternative hypotheses has been completely rewritten and expanded, and a new, earlier discussion of the weight of evidence interpretation of p-values is given. Also, a short presentation of the logic behind finding the probability of a Type II error when testing a two-sided alternative hypothesis now accompanies the general formula that can be used to calculate this probability. In Chapter 10 we mention the unrealistic "known variance" case when comparing population means only briefly and move swiftly to the more realistic "unknown variance" case. The discussion of comparing population variances has been shortened and made clearer. In Chapter 11 (Experimental Design and Analysis of Variance) we use a concise but understandable approach to covering one-way ANOVA, the randomized block design, and two-way ANOVA. A new, short presentation of using hypothesis testing to make pairwise comparisons now supplements our usual confidence interval discussion. Chapter 12 covers chi-square goodness-of-fit tests and tests of independence.
- **Streamlined and improved discussions of simple and multiple regression and statistical quality control.** As in the fourth edition, we use the Tasty Sub Shop Case to introduce the ideas of both simple and multiple regression analysis. This case has been popular with our readers. In Chapter 13 (Simple Linear Regression Analysis), the discussion of the simple linear regression model has been slightly shortened, the section on residual analysis has been significantly shortened and improved, and more exercises on residual analysis have been added. After discussing the basics of multiple regression, Chapter 14 has five innovative, advanced sections that are concise and can be covered in any order. These optional sections explain (1) using dummy variables (including an improved discussion of interaction when using dummy variables), (2) using squared and interaction terms, (3) model building and the effects of multicollinearity (including an added discussion of backward elimination), (4) residual analysis in multiple regression (including an improved and slightly expanded discussion of outlying and influential observations), and (5) logistic regression (a new section). Chapter 15, which is on the book's website and deals with process improvement, has been streamlined by relying on a single case, the hole location case, to explain $\bar{X}$ and R charts as well as establishing process control, pattern analysis, and capability studies.
- **Increased emphasis on Excel and MINITAB throughout the text.** The main text features Excel and MINITAB outputs. The end-of-chapter appendices provide improved step-by-step instructions about how to perform statistical analyses using these software packages as well as MegaStat, an Excel add-in.

Bruce L. Bowerman
Richard T. O'Connell
Emily S. Murphree
J. B. Orris

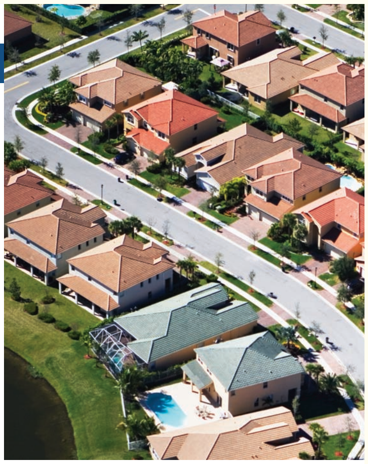
A TOUR OF THIS

Chapter Introductions

Each chapter begins with a list of the section topics that are covered in the chapter, along with chapter learning objectives and a preview of the case study analysis to be carried out in the chapter.

CHAPTER 1

An Introduction to Business Statistics



Learning Objectives

When you have mastered the material in this chapter, you will be able to:

- LO1-1 Define a variable.
- LO1-2 Describe the difference between a quantitative variable and a qualitative variable.
- LO1-3 Describe the difference between cross-sectional data and time series data.
- LO1-4 Construct and interpret a time series (runs) plot.
- LO1-5 Identify the different types of data sources: existing data sources, experimental studies, and observational studies.
- LO1-6 Describe the difference between a population and a sample.
- LO1-7 Distinguish between descriptive statistics and statistical inference.
- LO1-8 Explain the importance of random sampling.
- LO1-9 Identify the ratio, interval, ordinal, and nominative scales of measurement (Optional).

Chapter Outline

The subject of **statistics** involves the study of how to collect, analyze, and interpret data. **Data are facts and figures from which conclusions can be drawn.** Such conclusions are important to the decision making of many professions and organizations. For example, **economists** use conclusions drawn from the latest data on unemployment and inflation to help the government make policy decisions. **Financial planners** use recent trends in stock market prices and economic conditions to make investment decisions. **Accountants** use **sample data** concerning a company's *actual sales revenues* to assess whether the company's *claimed sales revenues* are valid. **Marketing professionals** help businesses decide which products to develop and market by using data that reveal consumer preferences. **Production supervisors** use manufacturing data to evaluate, control, and improve product quality. **Politicians** rely on data from public opinion polls to formulate legislation and to devise campaign strategies. **Physicians and hospitals** use data on the effectiveness of drugs and surgical procedures to provide patients with the best possible treatment.

In this chapter we begin to see how we collect and analyze data. As we proceed through the chapter, we introduce several case studies. These case studies (and others to be introduced later) are revisited throughout later chapters as we learn the statistical methods needed to analyze them. Briefly, we will begin to study three cases:

C

The Cell Phone Case. A bank estimates its cellular phone costs and decides whether to outsource management of its wireless resources by studying the calling patterns of its employees.

The Marketing Research Case. A bottling company investigates consumer reaction to a new bottle design for one of its popular soft drinks.

The Car Mileage Case. To determine if it qualifies for a federal tax credit based on fuel economy, an automaker studies the gas mileage of its new midsize model.

1.1 Data

LO1-1 Define a variable.

Data sets, elements, and variables We have said that data are facts and figures from which conclusions can be drawn. Together, the data that are collected for a particular study are referred to as a **data set.** For example, Table 1.1 is a data set that gives information about the new homes sold in a Florida luxury home development over a recent three-month period. Potential buyers in this housing community could choose either the "Diamond" or the "Ruby" home model design and could have the home built on either a lake lot or a treed lot (with no water access).

In order to understand the data in Table 1.1, note that any data set provides information about some group of individual **elements,** which may be people, objects, events, or other entities. The information that a data set provides about its elements usually describes one or more characteristics of these elements.

Any characteristic of an element is called a **variable.**

For the data set in Table 1.1, each sold home is an element, and four variables are used to describe the homes. These variables are (1) the home model design, (2) the type of lot on which the home was built, (3) the list (asking) price, and (4) the (actual) selling price. Moreover, each home model design came with "everything included"—specifically, a complete, luxury interior package and a choice (at no price difference) of one of three different architectural exteriors. The builder made the list price of each home solely dependent on the model design. however, the builder gave various price reductions for homes build on treed lots.

TABLE 1.1 A Data Set Describing Five Home Sales DS HomeSales

Home	Model Design	Lot Type	List Price	Selling Price
1	Diamond	Lake	$494,000	$494,000
2	Ruby	Treed	$447,000	$398,000
3	Diamond	Treed	$494,000	$440,000
4	Diamond	Treed	$494,000	$469,000
5	Ruby	Lake	$447,000	$447,000

Continuing Case Studies and Business Improvement Conclusions

The main chapter discussions feature real or realistic examples, continuing case studies, and a business improvement theme. The continuing case studies span not only individual chapters but also groups of chapters and tie together different statistical topics. To emphasize the text's theme of business improvement, icons BI are placed in the page margins to identify when statistical analysis has led to an important business improvement conclusion. Each conclusion is also highlighted in yellow for additional clarity. For example, in Chapters 1 and 3 we consider **The Cell Phone Case:**

TABLE 1.4 A Sample of Cellular Usages (in Minutes) for 100 Randomly Selected Employees DS CellUse

75	485	37	547	753	93	897	694	797	477
654	578	504	670	490	225	509	247	597	173
496	553	0	198	507	157	672	296	774	479
0	822	705	814	20	513	546	801	721	273
879	433	420	521	648	41	528	359	367	948
511	704	535	585	341	530	216	512	491	0
542	562	49	505	461	496	241	624	885	259
571	338	503	529	737	444	372	555	290	830
719	120	468	730	853	18	479	144	24	513
482	683	212	418	399	376	323	173	669	611

EXAMPLE 3.5 The Cell Phone Case: Reducing Cellular Phone Costs C

Suppose that a cellular management service tells the bank that if its cellular cost per minute for the random sample of 100 bank employees is over 18 cents per minute, the bank will benefit from automated cellular management of its calling plans. Last month's cellular usages for the 100 randomly selected employees are given in Table 1.4 (page 9), and a dot plot of these usages is given in the page margin. If we add the usages together, we find that the 100 employees used a total of 46,625 minutes. Furthermore, the total cellular cost incurred by the 100 employees is found to be $9,317 (this total includes base costs, overage costs, long distance, and roaming). This works out to an average of $9,317/46,625 = $.1998, or 19.98 cents per minute. Because this average cellular cost per minute exceeds 18 cents per minute, the bank will hire the cellular management service to manage its calling plans.

BI

TEXT'S FEATURES

Figures and Tables

Throughout the text, charts, graphs, tables, and Excel and MINITAB outputs are used to illustrate statistical concepts. For example:

- In Chapter 3 (**Descriptive Statistics: Numerical Methods**), the following figures are used to help explain the **Empirical Rule**. Moreover, in **The Car Mileage Case** an automaker uses the Empirical Rule to find estimates of the "typical," "lowest," and "highest" mileage that a new midsize car should be expected to get in combined city and highway driving. In actual practice, real automakers have provided similar information broken down into separate estimates for city and highway driving—see the Buick LaCrosse new car sticker in Figure 3.14.

FIGURE 3.14 The Empirical Rule and Tolerance Intervals for a Normally Distributed Population

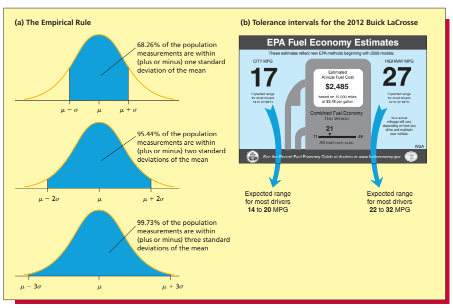
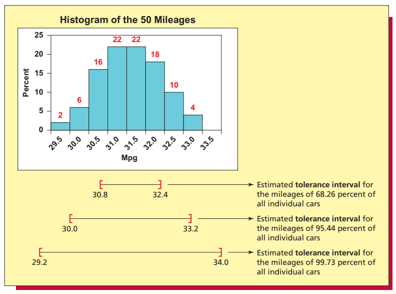


FIGURE 3.15 Estimated Tolerance Intervals in the Car Mileage Case



- In Chapter 7 (**Sampling and Sampling Distributions**), the following figures (and others) are used to help explain the **sampling distribution of the sample mean** and the **Central Limit Theorem**. In addition, the figures describe different applications of random sampling in **The Car Mileage Case**, and thus this case is used as an integrative tool to help students understand sampling distributions.

FIGURE 7.1 A Comparison of Individual Car Mileages and Sample Means

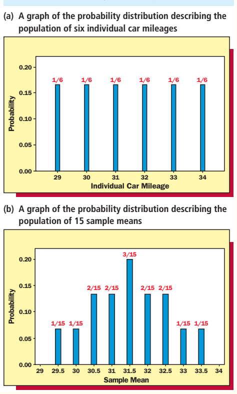


FIGURE 7.2 The Normally Distributed Population of All Individual Car Mileages and the Normally Distributed Population of All Possible Sample Means

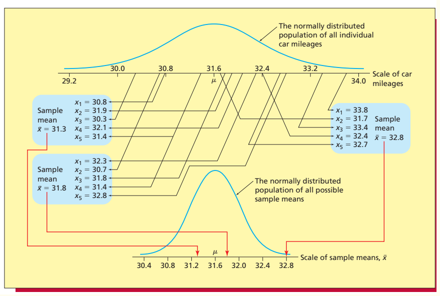


FIGURE 7.3 A Comparison of (1) the Population of All Individual Car Mileages, (2) the Sampling Distribution of the Sample Mean $\bar{x}$ When $n = 5$, and (3) the Sampling Distribution of the Sample Mean $\bar{x}$ When $n = 50$

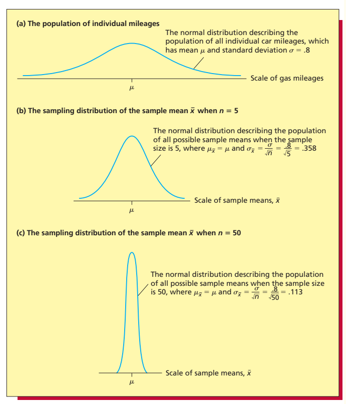
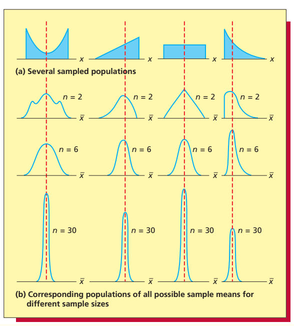


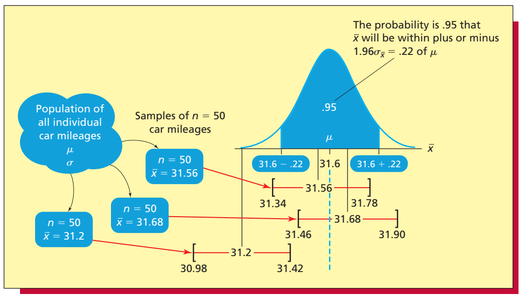
FIGURE 7.5 The Central Limit Theorem Says That the Larger the Sample Size Is, the More Nearly Normally Distributed Is the Population of All Possible Sample Means



A TOUR OF THIS

- In Chapter 8 (**Confidence Intervals**), the following figure (and others) are used to help explain the meaning of a **95 percent confidence interval** for the population mean. Furthermore, in **The Car Mileage Case** an automaker uses a confidence interval procedure specified by the Environmental Protection Agency (EPA) to find the EPA estimate of a new midsize model's true mean mileage.

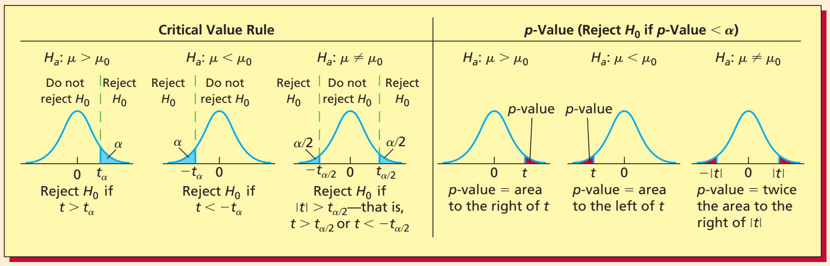
FIGURE 8.2 Three 95 Percent Confidence Intervals for μ



- In Chapter 9 (**Hypothesis Testing**), a five-step hypothesis testing procedure, **new graphical hypothesis testing summary boxes**, and many graphics are used to show how to carry out hypothesis tests.

A t Test about a Population Mean: σ Unknown

Null Hypothesis $H_0: \mu = \mu_0$ **Test Statistic** $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$ $df = n - 1$ **Assumptions** Normal population or Large sample size



The Five Steps of Hypothesis Testing

1. State the null hypothesis H_0 and the alternative hypothesis H_a.
2. Specify the level of significance α.
3. Select the test statistic.

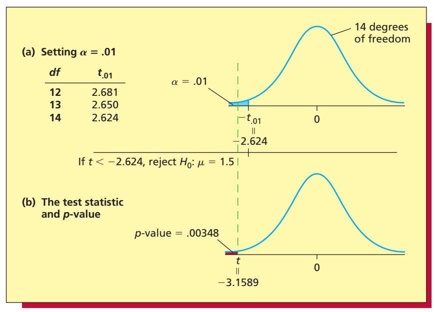
Using a critical value rule:

4. Determine the critical value rule for deciding whether to reject H_0.
5. Collect the sample data, compute the value of the test statistic, and decide whether to reject H_0 by using the critical value rule. Interpret the statistical results.

Using a *p*-value:

4. Collect the sample data, compute the value of the test statistic, and compute the p-value.
5. Reject H_0 at level of significance α if the p-value is less than α. Interpret the statistical results.

FIGURE 9.5 Testing $H_0: \mu = 1.5$ versus $H_a: \mu < 1.5$ by Using a Critical Value and the p-Value




```
Test of mu = 1.5 vs < 1.5
                                              95% Upper
Variable   N    Mean   StDev  SE Mean     Bound      T      P
Ratio     15  1.3433  0.1921   0.0496    1.4307  -3.16  0.003
```

- In Chapters 13 and 14 (**Simple Linear and Multiple Regression**), a substantial number of data plots, Excel and MINITAB outputs, and other graphics are used to teach simple and multiple regression analysis. For example, in **The Tasty Sub Shop Case** a business entrepreneur uses data plotted in Figures 14.1 and 14.2 and the Excel and MINITAB outputs in Figure 14.4 to predict the yearly revenue of a potential Tasty Sub Shop restaurant site on the basis of the population and business activity near the site. Using the **95 percent prediction interval** on the MINITAB output and projected restaurant operating costs, the entrepreneur decides whether to purchase a Tasty Sub Shop franchise for the potential restaurant site.

TEXT'S FEATURES

FIGURE 14.1 Plot of y (Yearly Revenue) versus x_1 (Population Size)

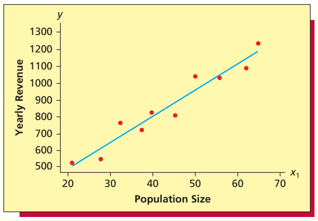


FIGURE 14.2 Plot of y (Yearly Revenue) versus x_2 (Business Rating)

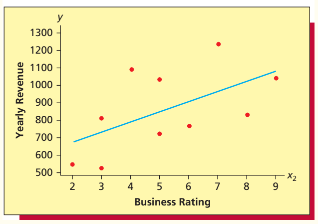


FIGURE 14.4 Excel and MINITAB Outputs of a Regression Analysis of the Tasty Sub Shop Revenue Data in Table 14.1 Using the Model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

(a) The Excel output

Regression Statistics

Multiple R	0.9905
R Square	0.9810 [8]
Adjusted R Square	0.9756 [9]
Standard Error	36.6856 [7]
Observations	10

ANOVA	df	SS	MS	F	Significance F
Regression	2	486355.7 [10]	243177.8	180.689 [13]	9.46E-07 [14]
Residual	7	9420.8 [11]	1345.835		
Total	9	495776.5 [12]			

	Coefficients	Standard Error [4]	t Stat [5]	P-value [6]	Lower 95% [19]	Upper 95% [19]
Intercept	125.289 [1]	40.9333	3.06	0.0183	28.4969	222.0807
population	14.1996 [2]	0.9100	15.60	1.07E-06	12.0478	16.3517
bus_rating	22.8107 [3]	5.7692	3.95	0.0055	9.1686	36.4527

(b) The MINITAB output

```
The regression equation is
revenue = 125 + 14.2 population + 22.8 bus_rating

Predictor        Coef   SE Coef[4]        T[5]      P[6]
Constant       125.29[1]   40.93          3.06     0.018
population    14.1996[2]    0.91          15.6     0.000
bus_rating     22.811[3]   5.769          3.95     0.006
S = 36.6856[7]           R-Sq = 98.10%[8]          R-Sq(adj) = 97.6%[9]

Analysis of Variance
Source            DF        SS          MS          F          P
Regression         2    486356[10]   243178    180.69[13]   0.000[14]
Residual Error     7      9421[11]     1346
Total              9    495777[12]

Predicted Values for New Observations
   New Obs     Fit[15]      SE Fit[16]      95% CI[17]            95% PI[18]
         1     956.6              15   (921.0, 992.2)      (862.8, 1050.4)

Values of Predictors for New Observations
   New Obs     population     bus_rating
         1           47.3              7
```

[1] b_0 [2] b_1 [3] b_2 [4] s_{b_j} = standard error of the estimate b_j [5] t statistics [6] p-values for t statistics [7] s = standard error
[8] R^2 [9] Adjusted R^2 [10] Explained variation [11] SSE = Unexplained variation [12] Total variation [13] F(model) statistic
[14] p-value for F(model) [15] $\hat{y}$ = point prediction when $x_1 = 47.3$ and $x_2 = 7$ [16] $s_{\hat{y}}$ = standard error of the estimate $\hat{y}$
[17] 95% confidence interval when $x_1 = 47.3$ and $x_2 = 7$ [18] 95% prediction interval when $x_1 = 47.3$ and $x_2 = 7$ [19] 95% confidence interval for β_j

Exercises

Many of the exercises in the text require the analysis of real data. Data sets are identified by an icon in the text and are included on the Online Learning Center (OLC): www.mhhe.com/bowermaness5e. Exercises in each section are broken into two parts—"Concepts" and "Methods and Applications"—and there are supplementary and Internet exercises at the end of each chapter.

2.7 Below we give the overall dining experience ratings (Outstanding, Very Good, Good, Average, or Poor) of 30 randomly selected patrons at a restaurant on a Saturday evening. DS RestRating

Outstanding	Good	Very Good	Very Good	Outstanding	Good
Outstanding	Outstanding	Outstanding	Very Good	Very Good	Average
Very Good	Outstanding	Outstanding	Outstanding	Outstanding	Very Good
Outstanding	Good	Very Good	Outstanding	Very Good	Outstanding
Good	Very Good	Outstanding	Very Good	Good	Outstanding

a Find the frequency distribution and relative frequency distribution for these data.
b Construct a percentage bar chart for these data.
c Construct a percentage pie chart for these data.

Chapter Ending Material and Excel/MINITAB/MegaStat® Tutorials

The end-of-chapter material includes a chapter summary, a glossary of terms, important formula references, and comprehensive appendices that show students how to use Excel, MINITAB, and MegaStat.

Chapter Summary

We began this chapter by presenting and comparing several measures of **central tendency.** We defined the **population mean** and we saw how to estimate the population mean by using a **sample mean.** We also defined the **median** and **mode,** and we compared the mean, median, and mode for symmetrical distributions and for distributions that are skewed to the right or left. We then studied measures of **variation** (or *spread*). We defined the **range, variance,** and **standard deviation,** and we saw how to estimate a population variance and standard deviation by using a sample. We learned that a good way to interpret the standard deviation when a population is (approximately) normally distributed is to use the **Empirical Rule,** and we studied **Chebyshev's Theorem,** which gives us intervals containing reasonably large fractions of the population units no matter what the population's shape might be. We also saw that, when a data set is highly skewed, it is best to use **percentiles** and **quartiles** to measure variation, and we learned how to construct a **box-and-whiskers plot** by using the quartiles.

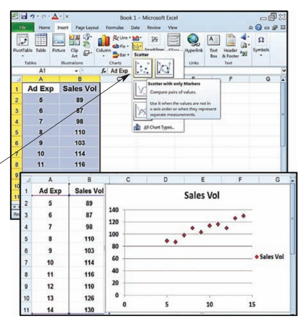
After learning how to measure and depict central tendency and variability, we presented several optional topics. First, we discussed several numerical measures of the relationship between two variables. These included the **covariance,** the **correlation coefficient,** and the **least squares line.** We then introduced the concept of a **weighted mean** and also explained how to compute descriptive statistics for **grouped data.** Finally, we showed how to calculate the **geometric mean** and demonstrated its interpretation.

Glossary of Terms

box-and-whiskers display (box plot): A graphical portrayal of a data set that depicts both the central tendency and variability of the data. It is constructed using Q_1, M_d, and Q_3. (pages 121, 122)
central tendency: A term referring to the middle of a population or sample of measurements. (page 99)
outlier (in a box-and-whiskers display): A measurement less than the lower limit or greater than the upper limit. (page 122)
percentile: The value such that a specified percentage of the measurements in a population or sample fall at or below it. (page 118)
point estimate: A one-number estimate for the value of a population parameter. (page 99)

Constructing a scatter plot of sales volume versus advertising expenditure as in Figure 2.24 on page 67 (data file: SalesPlot.xlsx):

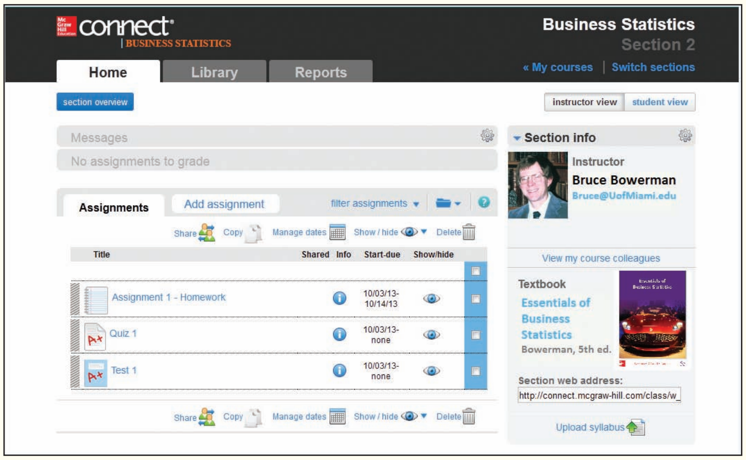
- Enter the advertising and sales data in Table 2.20 on page 67 into columns A and B—advertising expenditures in column A with label "Ad Exp" and sales values in column B with label "Sales Vol." **Note: The variable to be graphed on the horizontal axis must be in the first column** (that is, the left-most column) and **the variable to be graphed on the vertical axis must be in the second column** (that is, the rightmost column).
- Select the entire range of data to be graphed.
- Select **Insert : Scatter : Scatter with only Markers**
- The scatter plot will be displayed in a graphics window. Move the plot to a chart sheet and edit appropriately.



WHAT TECHNOLOGY CONNECTS STUDENTS...

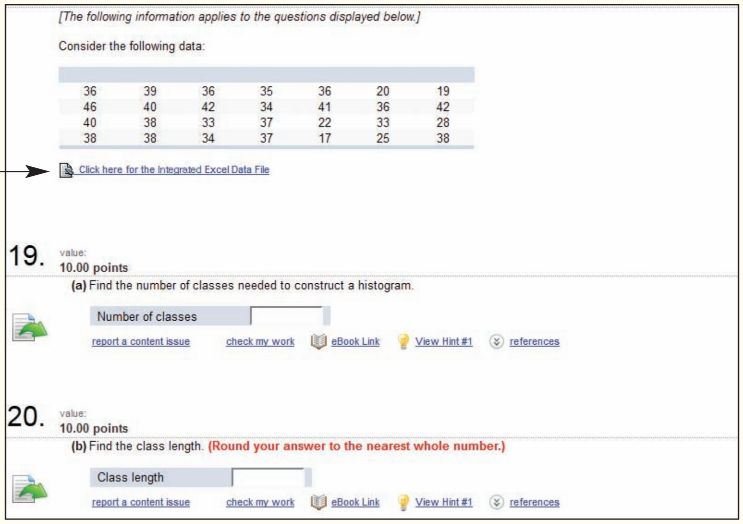
McGraw-Hill ***Connect®*** ***Business Statistics*** is an online assignment and assessment solution that connects students with the tools and resources they'll need to achieve success through faster learning, higher retention, and more efficient studying. It provides instructors with tools to quickly pick content and assignments according to the topics 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 feedback on any questions that students may have missed.



Integration of Excel Data Files. 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.

Excel Data File



Student Resource Library. The *Connect Business Statistics* Student Library is the place for students to access additional resources. The Student Library provides quick access to recorded lectures, practice materials, eBooks, data files, PowerPoint files, and more.

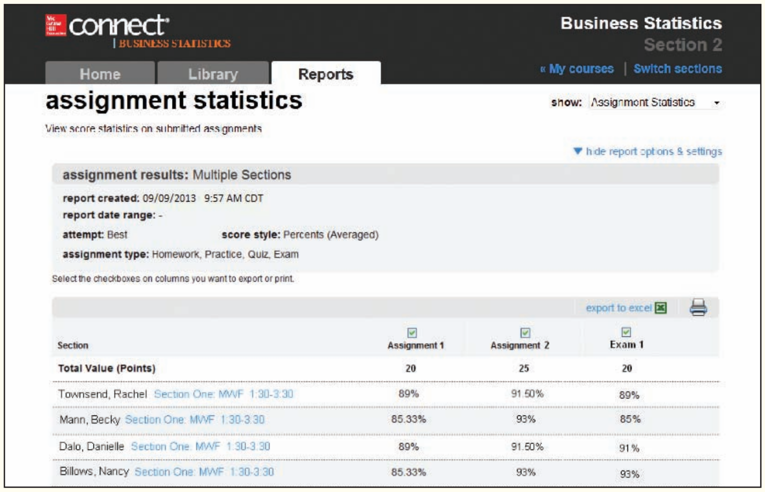
TO SUCCESS IN BUSINESS STATISTICS?

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 side-by-side comparisons with correct answers.
- Access and review each response; manually change grades or leave comments for students to review.

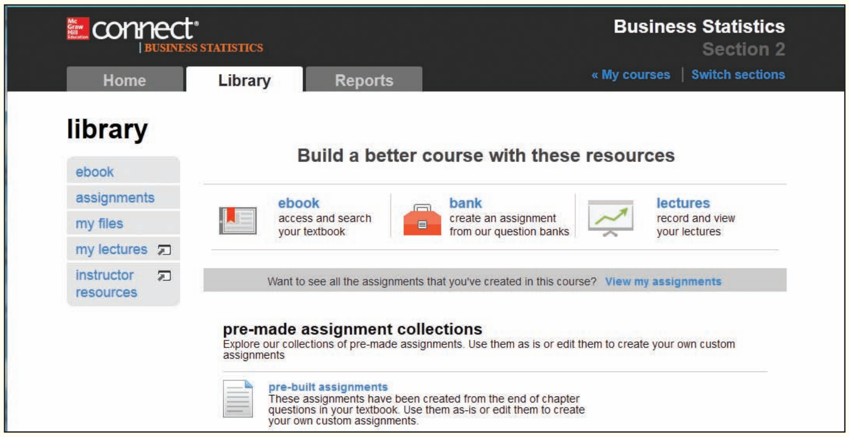
Student Reporting. *Connect Business Statistics* keeps instructors informed about how each student, section, and class is performing, allowing for more productive use of lecture and office hours. The progress-tracking function enables you to:

- 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 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:

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



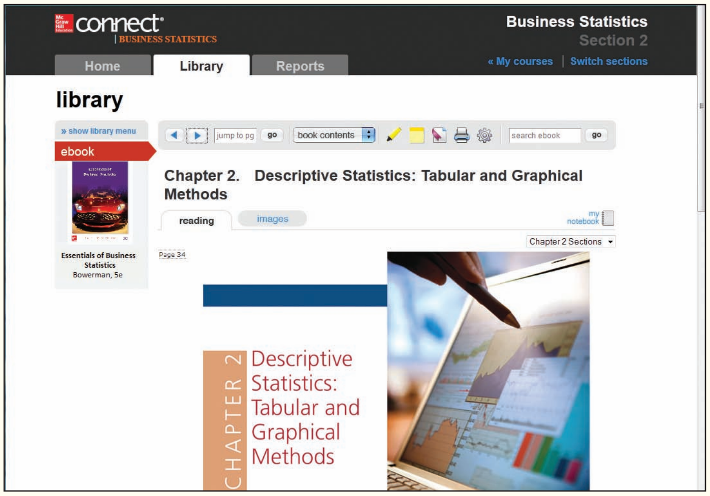
WHAT TECHNOLOGY CONNECTS STUDENTS...

connect plus+ | BUSINESS STATISTICS

Connect® Plus Business Statistics includes a seamless integration of an eBook and *Connect Business Statistics.* Benefits of the rich functionality integrated into the product are outlined below.

Integrated Media-Rich eBook. An integrated media-rich eBook allows students to access media in context with each chapter. Students can highlight, take notes, and access shared instructor highlights and notes to learn the course material.

Dynamic Links. Dynamic links provide a connection between the problems or questions you assign to your students and the location in the eBook where that problem or question is covered.



Powerful Search Function. A powerful search function pinpoints and connects key concepts in a snap. This state-of-the-art, thoroughly tested system supports you in preparing students for the world that awaits. For more information about *Connect*, go to www.mcgrawhillconnect.com or contact your local McGraw-Hill sales representative.

Connect Packaging Options

Connect with 1 Semester Access Card: 0077641159
Connect Plus with 1 Semester Access Card: 0077641183

McGraw-Hill tegrity campus

Tegrity Campus: Lectures 24/7

Tegrity Campus is a service that makes class time available 24/7. With *Tegrity Campus,* you can automatically capture every lecture in a searchable format for students to review when they study and complete assignments. With a simple one-click start-and-stop process, you capture all computer screens and corresponding 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.

TO SUCCESS IN BUSINESS STATISTICS?

WHAT SOFTWARE IS AVAILABLE?

MegaStat® for Microsoft Excel®—Windows® and Mac OS-X: www.mhhe.com/megastat

MegaStat is a full-featured Excel add-in by J. B. Orris of Butler University that is available with this text. The online installer will install the MegaStat add-in for all versions of Microsoft Excel beginning with Excel 2007 and up to Excel 2013. MegaStat 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. Ease-of-use features include AutoExpand 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. MegaStat works with all recent versions of Excel.

MINITAB® (ISBN: 007305237x)

Minitab® Student Version 14 is available to help students solve the business statistics exercises in the text. This software is available in the student version and can be packaged with any McGraw-Hill business statistics text.

McGraw-Hill Customer Care 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 contact 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.

WHAT RESOURCES ARE AVAILABLE FOR INSTRUCTORS?

Online Learning Center: www.mhhe.com/bowermaness5e

The Online Learning Center (OLC) is the text website with online content for both students and instructors. It provides the instructor with a complete Instructor's Manual in Word format, the complete Test Bank in both Word files and computerized EZ Test format, Instructor PowerPoint slides, text art files, an introduction to ALEKS®, an introduction to McGraw-Hill *Connect Business Statistics*®, access to the eBook, and more.

EZ TEST

All test bank questions are available in an EZ Test electronic format. Included are a number of multiple-choice, true/false, and short-answer questions and problems. The answers to all questions are given, along with a rating of the level of difficulty, Bloom's taxonomy question type, and AACSB knowledge category.

Online Course Management

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

- **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.
- **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.
- **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.
- **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: 0077641175)

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.

Online Learning Center: www.mhhe.com/bowermaness5e

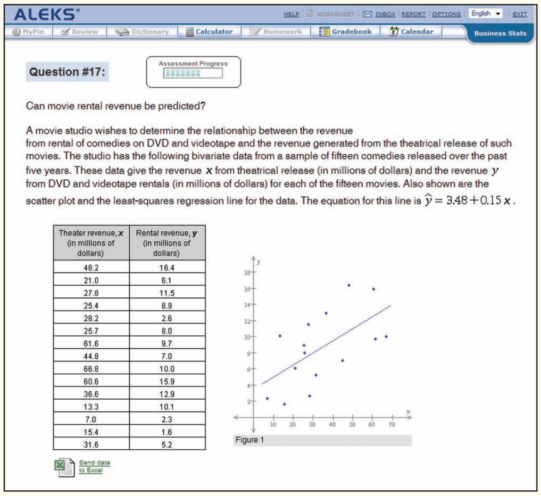
The Online Learning Center (OLC) provides students with the following content:

- Quizzes—self-grading to assess knowledge of the material
- Data sets—import into Excel for quick calculation and analysis
- PowerPoint—gives an overview of chapter content
- Appendixes—quick look-up when the text isn't available

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/Irwin, 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/highered/business. ALEKS is a registered trademark of ALEKS Corporation.

ACKNOWLEDGMENTS

We wish to thank many people who have helped to make this book a reality. We thank Drena Bowerman, who spent many hours cutting and taping and making trips to the copy shop, so that we could complete the manuscript on time. As indicated on the title page, we thank Professor Steven C. Huchendorf, University of Minnesota; Dawn C. Porter, University of Southern California; and Patrick J. Schur, Miami University; for major contributions to this book. We also thank Susan Cramer of Miami University for helpful advice on writing this book.

We also wish to thank the people at McGraw-Hill/Irwin for their dedication to this book. These people include senior brand manager Thomas Hayward, who is an extremely helpful resource to the authors; executive editor Dick Hercher, who persuaded us initially to publish with McGraw-Hill/Irwin; senior development editor Wanda Zeman, who has shown great dedication to the improvement of this book; and content project manager Harvey Yep, who has very capably and diligently guided this book through its production and who has been a tremendous help to the authors. We also thank our former executive editor, Scott Isenberg, for the tremendous help he has given us in developing all of our McGraw-Hill business statistics books.

We also wish to thank the error checkers, Patrick Schur, Miami University of Ohio, Lou Patille, Colorado Heights University, and Peter Royce, University of New Hampshire, who were very helpful. Most importantly, we wish to thank our families for their acceptance, unconditional love, and support.

Many reviewers have contributed to this book, and we are grateful to all of them. They include

Lawrence Acker, Harris-Stowe State University
Ajay K. Aggarwal, Millsaps College
Mohammad Ahmadi, University of Tennessee–Chattanooga
Sung K. Ahn, Washington State University
Imam Alam, University of Northern Iowa
Eugene Allevato, Woodbury University
Mostafa S. Aminzadeh, Towson University
Henry Ander, Arizona State University–Tempe
Randy J. Anderson, California State University–Fresno
Mohammad Bajwa, Northampton Community College
Ron Barnes, University of Houston–Downtown
John D. Barrett, University of North Alabama
Mary Jo Boehms, Jackson State Community College
Pamela A. Boger, Ohio University–Athens
David Booth, Kent State University
Dave Bregenzer, Utah State University
Philip E. Burian, Colorado Technical University–Sioux Falls
Giorgio Canarella, California State University–Los Angeles
Margaret Capen, East Carolina University
Priscilla Chaffe-Stengel, California State University–Fresno
Gary H. Chao, Utah State University
Ali A. Choudhry, Florida International University
Richard Cleary, Bentley College
Bruce Cooil, Vanderbilt University
Sam Cousley, University of Mississippi
Teresa A Dalton, University of Denver
Nit Dasgupta, University of Wisconsin–Eau Claire
Linda Dawson, University of Washington–Tacoma
Jay Devore, California Polytechnic State University
Bernard Dickman, Hofstra University
Joan Donohue, University of South Carolina
Anne Drougas, Dominican University
Mark Eakin, University of Texas–Arlington
Hammou Elbarmi, Baruch College
Ashraf ELHoubi, Lamar University
Soheila Fardanesh, Towson University
Nicholas R. Farnum, California State University–Fullerton
James Flynn, Cleveland State University
Lillian Fok, University of New Orleans
Tom Fox, Cleveland State Community College
Charles A. Gates Jr., Olivet Nazarene University
Linda S. Ghent, Eastern Illinois University
Allen Gibson, Seton Hall University
Scott D. Gilbert, Southern Illinois University
Nicholas Gorgievski, Nichols College
TeWhan Hahn, University of Idaho
Clifford B. Hawley, West Virginia University
Rhonda L. Hensley, North Carolina A&T State University
Eric Howington, Valdosta State University
Zhimin Huang, Adelphi University
Steven C. Huchendorf, University of Minnesota
Dene Hurley, Lehman College–CUNY
C. Thomas Innis, University of Cincinnati
Jeffrey Jarrett, University of Rhode Island
Craig Johnson, Brigham Young University
Valerie M. Jones, Tidewater Community College
Nancy K. Keith, Missouri State University
Thomas Kratzer, Malone University
Alan Kreger, University of Maryland
Michael Kulansky, University of Maryland
Risa Kumazawa, Georgia Southern University
David A. Larson, University of South Alabama
John Lawrence, California State University–Fullerton
Lee Lawton, University of St. Thomas
John D. Levendis, Loyola University–New Orleans
Barbara Libby, Walden University
Carel Ligeon, Auburn University–Montgomery
Kenneth Linna, Auburn University–Montgomery
David W. Little, High Point University
Donald MacRitchie, Framingham State College
Cecelia Maldonado, Georgia Southern State University
Edward Markowski, Old Dominion University
Mamata Marme, Augustana College
Jerrold H. May, University of Pittsburgh
Brad McDonald, Northern Illinois University
Richard A. McGowan, Boston College

ACKNOWLEDGMENTS

Christy McLendon, University of New Orleans
John M. Miller, Sam Houston State University
Richard Miller, Cleveland State University
Robert Mogull, California State University–Sacramento
Jason Molitierno, Sacred Heart University
Steven Rein, California Polytechnic State University
Donna Retzlaff-Roberts, University of South Alabama
Peter Royce, University of New Hampshire
Fatollah Salimian, Salisbury University
Yvonne Sandoval, Pima Community College
Sunil Sapra, California State University–Los Angeles
Patrick J. Schur, Miami University
William L. Seaver, University of Tennessee
Kevin Shanahan, University of Texas–Tyler
Arkudy Shemyakin, University of St. Thomas
Charlie Shi, Daiblo Valley College
Joyce Shotick, Bradley University
Plamen Simeonov, University of Houston Downtown
Bob Smidt, California Polytechnic State University
Rafael Solis, California State University–Fresno
Toni M. Somers, Wayne State University
Ronald L. Spicer, Colorado Technical University–Sioux Falls
Mitchell Spiegel, Johns Hopkins University
Timothy Staley, Keller Graduate School of Management
David Stoffer, University of Pittsburgh
Matthew Stollack, St. Norbert College
Cliff Stone, Ball State University
Courtney Sykes, Colorado State University
Bedassa Tadesse, University of Minnesota–Duluth
Stanley Taylor, California State University–Sacramento
Patrick Thompson, University of Florida
Richard S. Tovar-Silos, Lamar University
Emmanuelle Vaast, Long Island University–Brooklyn
Ed Wallace, Malcolm X College
Bin Wang, Saint Edwards University
Allen Webster, Bradley University
Blake Whitten, University of Iowa
Neil Wilmot, University of Minnesota–Duluth
Susan Wolcott-Hanes, Binghamton University
Mustafa Yilmaz, Northeastern University
Gary Yoshimoto, Saint Cloud State University
William F. Younkin, Miami University
Xiaowei Zhu, University of Wisconsin–Milwaukee

DEDICATION

Bruce L. Bowerman
To my wife, children, sister, and other family members:
Drena
Michael, Jinda, Benjamin, and Lex
Asa and Nicole
Susan
Barney, Fiona, and Radeesa
Daphne, Chloe, and Edgar
Gwyneth and Tony
Callie, Bobby, Marmalade, Randy, and Penney
Clarence, Quincy, Teddy, Julius, Charlie, and Sally

Richard T. O'Connell
To my children and grandchildren:
Christopher, Bradley, Sam, and Joshua

Emily S. Murphree
To Kevin and the Math Ladies

J. B. Orris
To my children:
Amy and Bradley

Chapter-by-Chapter Revisions for 5th Edition

Chapter 1

- Initial example made clearer.
- Two new graphical examples added to better introduce quantitative and qualitative variables.
- Intuitive explanation of random sampling and introduction of 3 major case studies made more concise.
- New subsection on ethical statistical practice.
- Cable cost example updated.
- Data set for coffee temperature case expanded and ready for use in continuous probability distribution chapter.

Chapter 2

- Pizza preference data replaces Jeep preference data in creating bar and pie charts and in business decision making.
- Seven new data sets added.
- Eighteen new exercises replace former exercises.

Chapter 3

- Section on percentiles, quartiles, and box plots completely rewritten, simplified, and shortened.
- Ten new data sets used.
- Nineteen new exercises replace former exercises.

Chapter 4

- Main discussion in chapter rewritten and simplified.
- Cable penetration example (based on Time Warner Cable) replaces newspaper subscription example.
- Employment discrimination case (based on real pharmaceutical company) used in conditional probability section.
- Exercises updated in this and all subsequent chapters.

Chapter 5

- Introduction to discrete probability distributions rewritten, simplified, and shortened.
- Binominal distribution introduced using a tree diagram.
- New optional section on joint distributions and covariance previously found in an appendix.

Chapter 6

- Introduction to continuous probability distributions improved and motivated by coffee temperature data.
- Uniform distribution section now begins with an example.
- Normal distribution motivated by tie-in to coffee temperature data.

Chapter 7

- A more seamless transition from a small population example involving sampling car mileages to a related large population example.
- New optional section deriving the mean and variance of the sample mean.

Chapter 8

- A shorter and clearer discussion of the difference between a confidence interval and a tolerance interval.
- New section on estimating parameters of finite populations.

Chapter 9

- Discussion of formulating the null and alternative hypotheses completely rewritten and expanded.
- New, earlier discussion of the weight of evidence interpretation of p-values.
- Short presentation of the logic behind finding the probability of a Type II error when testing a two-sided alternative hypothesis now accompanies the general formula for calculating this probability.

Chapter 10

- Discussion of comparing population variances made shorter and clearer.

Chapter 11

- New, short presentation of using hypothesis testing to make pairwise comparisons now supplements our usual confidence interval discussion.

Chapter 12

- No significant changes.

Chapter 13

- Discussion of the simple linear regression model slightly shortened.
- Section on residual analysis significantly shortened and improved.
- New exercises on residual analysis.

Chapter 14

- Improved discussion of interaction using dummy variables.
- Discussion of backward elimination added.
- Improved and slightly expanded discussion of outlying and influential observations.
- Section on logistic regression added.
- New supplementary exercises.

Chapter 15

- X bar and R charts presented much more concisely using one example.

Brief Table of Contents

Table of Contents

Chapter 7

Sampling and Sampling Distributions

Chapter 8

Confidence Intervals

Chapter 9

Hypothesis Testing

Chapter 10

Statistical Inferences Based on Two Samples

Chapter 11

Experimental Design and Analysis of Variance

Chapter 12

Chi-Square Tests

Chapter 13
Simple Linear Regression Analysis

Chapter 14
Multiple Regression and Model Building

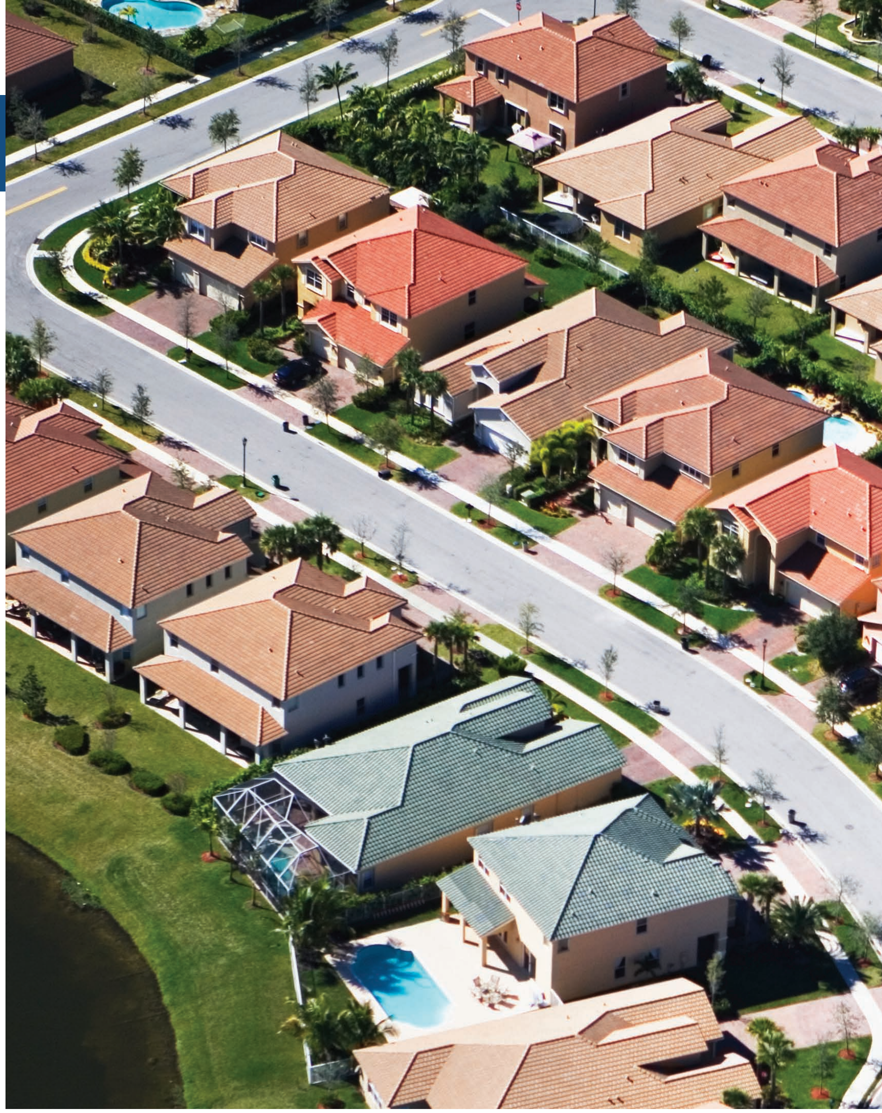
Appendix A

Essentials of Business Statistics

FIFTH EDITION

CHAPTER 1

An Introduction to Business Statistics



Learning Objectives

When you have mastered the material in this chapter, you will be able to:

- **LO1-1** Define a variable.
- **LO1-2** Describe the difference between a quantitative variable and a qualitative variable.
- **LO1-3** Describe the difference between cross-sectional data and time series data.
- **LO1-4** Construct and interpret a time series (runs) plot.
- **LO1-5** Identify the different types of data sources: existing data sources, experimental studies, and observational studies.
- **LO1-6** Describe the difference between a population and a sample.
- **LO1-7** Distinguish between descriptive statistics and statistical inference.
- **LO1-8** Explain the importance of random sampling.
- **LO1-9** Identify the ratio, interval, ordinal, and nominative scales of measurement (Optional).

Chapter Outline

The subject of **statistics** involves the study of how to collect, analyze, and interpret data. **Data are facts and figures from which conclusions can be drawn.** Such conclusions are important to the decision making of many professions and organizations. For example, **economists** use conclusions drawn from the latest data on unemployment and inflation to help the government make policy decisions. **Financial planners** use recent trends in stock market prices and economic conditions to make investment decisions. **Accountants** use **sample data** concerning a company's *actual sales revenues* to assess whether the company's *claimed sales revenues* are valid. **Marketing professionals** help businesses decide which products to develop and market by using data that reveal consumer preferences. **Production supervisors** use manufacturing data to evaluate, control, and improve product quality. **Politicians** rely on data from public opinion polls to formulate legislation and to devise campaign strategies. **Physicians and hospitals** use data on the effectiveness of drugs and surgical procedures to provide patients with the best possible treatment.

In this chapter we begin to see how we collect and analyze data. As we proceed through the chapter, we introduce several case studies. These case studies (and others to be introduced later) are revisited throughout later chapters as we learn the statistical methods needed to analyze them. Briefly, we will begin to study three cases:

C

The Cell Phone Case. A bank estimates its cellular phone costs and decides whether to outsource management of its wireless resources by studying the calling patterns of its employees.

The Marketing Research Case. A bottling company investigates consumer reaction to a new bottle design for one of its popular soft drinks.

The Car Mileage Case. To determine if it qualifies for a federal tax credit based on fuel economy, an automaker studies the gas mileage of its new midsize model.

1.1 Data

LO1-1 Define a variable.

Data sets, elements, and variables We have said that data are facts and figures from which conclusions can be drawn. Together, the data that are collected for a particular study are referred to as a **data set.** For example, Table 1.1 is a data set that gives information about the new homes sold in a Florida luxury home development over a recent three-month period. Potential buyers in this housing community could choose either the "Diamond" or the "Ruby" home model design and could have the home built on either a lake lot or a treed lot (with no water access).

In order to understand the data in Table 1.1, note that any data set provides information about some group of individual **elements,** which may be people, objects, events, or other entities. The information that a data set provides about its elements usually describes one or more characteristics of these elements.

Any characteristic of an element is called a **variable.**

For the data set in Table 1.1, each sold home is an element, and four variables are used to describe the homes. These variables are (1) the home model design, (2) the type of lot on which the home was built, (3) the list (asking) price, and (4) the (actual) selling price. Moreover, each home model design came with "everything included"—specifically, a complete, luxury interior package and a choice (at no price difference) of one of three different architectural exteriors. The builder made the list price of each home solely dependent on the model design. However, the builder gave various price reductions for homes built on treed lots.

TABLE 1.1 A Data Set Describing Five Home Sales DS HomeSales

Home	Model Design	Lot Type	List Price	Selling Price
1	Diamond	Lake	$494,000	$494,000
2	Ruby	Treed	$447,000	$398,000
3	Diamond	Treed	$494,000	$440,000
4	Diamond	Treed	$494,000	$469,000
5	Ruby	Lake	$447,000	$447,000

The data in Table 1.1 are real (with some minor modifications to protect privacy) and were provided by a business executive—a friend of the authors—who recently received a promotion and needed to move to central Florida. While searching for a new home, the executive and his family visited the luxury home community and decided they wanted to purchase a Diamond model on a treed lot. The list price of this home was $494,000, but the developer offered to sell it for an "incentive" price of $469,000. Intuitively, the incentive price's $25,000 savings off list price seemed like a good deal. However, the executive resisted making an immediate decision. Instead, he decided to collect data on the selling prices of new homes recently sold in the community and use the data to assess whether the developer might accept a lower offer. In order to collect "relevant data," the executive talked to local real estate professionals and learned that new homes sold in the community during the previous three months were a good indicator of current home value. Using real estate sales records, the executive also learned that five of the community's new homes had sold in the previous three months. The data given in Table 1.1 are the data that the executive collected about these five homes.

LO1-2 Describe the difference between a quantitative variable and a qualitative variable.

TABLE 1.2
2012 MLB Payrolls
DS MLB

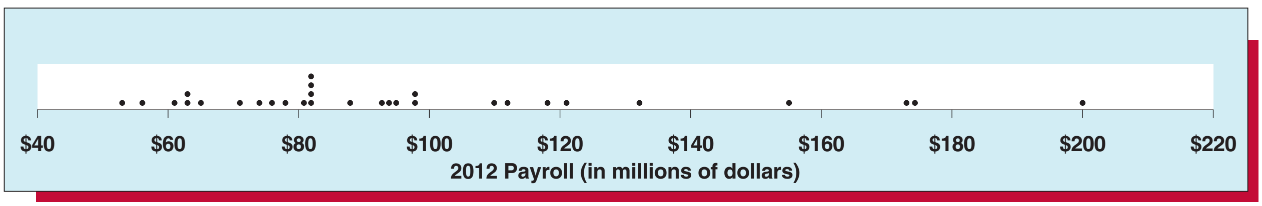
Team	2012 Payroll
New York Yankees	$200
Philadelphia Phillies	$174
Boston Red Sox	$173
Los Angeles Angels	$155
Detroit Tigers	$132
Texas Rangers	$121
San Francisco Giants	$118
Miami Marlins	$112
St. Louis Cardinals	$110
Milwaukee Brewers	$98
Chicago White Sox	$98
Los Angeles Dodgers	$95
Minnesota Twins	$94
New York Mets	$93
Chicago Cubs	$88
Atlanta Braves	$82
Cincinnati Reds	$82
Seattle Mariners	$82
Washington Nationals	$82
Baltimore Orioles	$81
Colorado Rockies	$78
Toronto Blue Jays	$76
Arizona Diamondbacks	$74
Cleveland Indians	$71
Tampa Bay Rays	$65
Pittsburgh Pirates	$63
Kansas City Royals	$63
Houston Astros	$61
San Diego Padres	$56
Oakland Athletics	$53

Source: http://baseball.about.com/od/newsrumors/a/2012-Baseball-Team-Payrolls.htm (accessed September 12, 2013).

Quantitative and qualitative variables In order to understand the conclusions the business executive reached using the data in Table 1.1, we need to further discuss variables. For any variable describing an element in a data set, we carry out a **measurement** to assign a value of the variable to the element. For example, in the real estate example, real estate sales records gave the actual selling price of each home to the nearest dollar. In another example, a credit card company might measure the time it takes for a cardholder's bill to be paid to the nearest day. Or, in a third example, an automaker might measure the gasoline mileage obtained by a car in city driving to the nearest one-tenth of a mile per gallon by conducting a mileage test on a driving course prescribed by the Environmental Protection Agency (EPA). If the possible values of a variable are numbers that represent quantities (that is, "how much" or "how many"), then the variable is said to be **quantitative.** For example, (1) the actual selling price of a home, (2) the payment time of a bill, (3) the gasoline mileage of a car, and (4) the 2012 payroll of a Major League Baseball team are all quantitative variables. Considering the last example, Table 1.2 in the page margin gives the 2012 payroll (in millions of dollars) for each of the 30 Major League Baseball (MLB) teams. Moreover, Figure 1.1 portrays the team payrolls as a **dot plot.** In this plot, each team payroll is shown as a dot located on the real number line—for example, the leftmost dot represents the payroll for the Oakland Athletics. In general, the values of a quantitative variable are numbers on the real line. In contrast, if we simply record into which of several categories an element falls, then the variable is said to be **qualitative** or **categorical.** Examples of categorical variables include (1) a person's gender, (2) whether a person who purchases a product is satisfied with the product, (3) the type of lot on which a home is built, and (4) the color of a car.[1] Figure 1.2 illustrates the categories we might use for the qualitative variable "car color." This figure is a **bar chart** showing the 10 most popular (worldwide) car colors for 2012 and the percentages of cars having these colors.

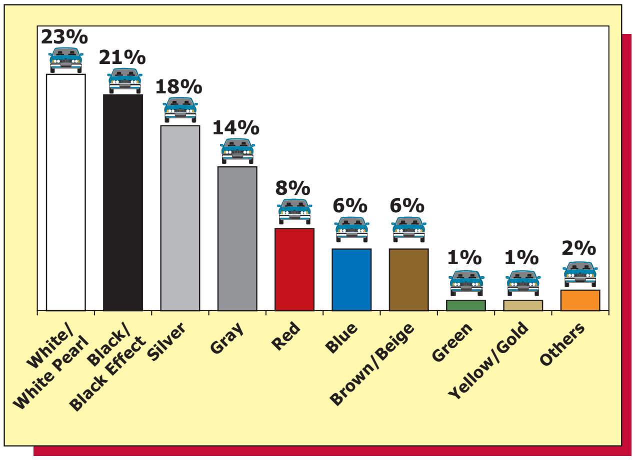
Of the four variables describing the home sales data in Table 1.1, two variables—list price and selling price—are quantitative, and two variables—model design and lot type—are qualitative. Furthermore, when the business executive examined Table 1.1, he noted that homes on lake lots had sold at their list price, but homes on treed lots had not. Because the executive and his family wished to purchase a Diamond model on a treed lot, the executive also noted that two Diamond

FIGURE 1.1 A Dot Plot of 2012 MLB Payrolls (Payroll Is a Quantitative Variable)



[1]Optional Section 1.5 discusses two types of quantitative variables (ratio and interval) and two types of qualitative variables (ordinal and nominative).

FIGURE 1.2 The Ten Most Popular Car Colors in the World for 2012 (Car Color Is a Qualitative Variable)



Source: http://www.autoweek.com/article/20121206/carnews01/121209911 (accessed September 12, 2013).

FIGURE 1.3 Time Series Plot of the Average Basic Cable Rates in the U.S. from 1999 to 2009 DS BasicCable

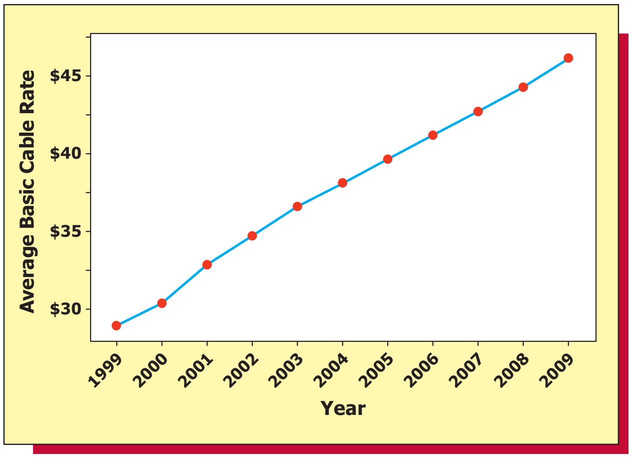


TABLE 1.3 The Average Basic Cable Rates in the U.S. from 1999 to 2009 DS BasicCable

Year	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Cable Rate	$ 28.92	30.37	32.87	34.71	36.59	38.14	39.63	41.17	42.72	44.28	46.13

Source: U.S. Energy Information Administration, http://www.eia.gov/

models on treed lots had sold in the previous three months. One of these Diamond models had sold for the incentive price of $469,000, but the other had sold for a lower price of $440,000. Hoping to pay the lower price for his family's new home, the executive offered $440,000 for the Diamond model on the treed lot. Initially, the home builder turned down this offer, but two days later the builder called back and accepted the offer. The executive had used data to buy the new home for $54,000 less than the list price and $29,000 less than the incentive price!

Cross-sectional and time series data Some statistical techniques are used to analyze *cross-sectional data,* while others are used to analyze *time series data.* **Cross-sectional data** are data collected at the same or approximately the same point in time. For example, suppose that a bank wishes to analyze last month's cell phone bills for its employees. Then, because the cell phone costs given by these bills are for different employees in the same month, the cell phone costs are cross-sectional data. **Time series data** are data collected over different time periods. For example, Table 1.3 presents the average basic cable television rate in the United States for each of the years 1999 to 2009. Figure 1.3 is a **time series plot**—also called a **runs plot**—of these data. Here we plot each cable rate on the vertical scale versus its corresponding time index (year) on the horizontal scale. For instance, the first cable rate ($28.92) is plotted versus 1999, the second cable rate ($30.37) is plotted versus 2000, and so forth. Examining the time series plot, we see that the cable rates increased substantially from 1999 to 2009. Finally, because the five homes in Table 1.1 were sold over a three-month period that represented a relatively stable real estate market, we can consider the data in Table 1.1 to essentially be cross-sectional data.

LO1-3 Describe the difference between cross-sectional data and time series data.

LO1-4 Construct and interpret a time series (runs) plot.

1.2 Data Sources

Primary data are data collected by an individual directly through personally planned **experimentation** or **observation. Secondary data** are data taken from an **existing source.**

LO1-5 Identify the different types of data sources: existing data sources, experimental studies, and observational studies.

Existing sources Sometimes we can use data *already gathered* by public or private sources. The Internet is an obvious place to search for electronic versions of government publications, company reports, and business journals, but there is also a wealth of information available in the reference section of a good library or in county courthouse records.

If a business wishes to find demographic data about regions of the United States, a natural source is the U.S. Census Bureau's website at http://www.census.gov. Other useful websites for economic and financial data include the Federal Reserve at http://research.stlouisfed.org/fred2/ and the Bureau of Labor Statistics at http://stats.bls.gov/.

However, given the ease with which anyone can post documents, pictures, weblogs, and videos on the World Wide Web, not all sites are equally reliable. Some of the sources will be more useful, exhaustive, and error-free than others. Fortunately, search engines prioritize the lists and provide the most relevant and highly used sites first.

Obviously, performing such web searches costs next to nothing and takes relatively little time, but the tradeoff is that we are also limited in terms of the type of information we are able to find. Another option may be to use a private data source. Most companies keep employee records and information about their customers, products, processes, and advertising results. If we have no affiliation with these companies, however, these data may be difficult to obtain.

Another alternative would be to contact a data collection agency, which typically incurs some kind of cost. You can either buy subscriptions or purchase individual company financial reports from agencies like Bloomberg and Dow Jones & Company. If you need to collect specific information, some companies, such as ACNielsen and Information Resources, Inc., can be hired to collect the information for a fee.

Experimental and observational studies There are many instances when the data we need are not readily available from a public or private source. In cases like these, we need to collect the data ourselves. Suppose we work for a soft drink company and want to assess consumer reactions to a new bottled water. Because the water has not been marketed yet, we may choose to conduct taste tests, focus groups, or some other market research. When projecting political election results, telephone surveys and exit polls are commonly used to obtain the information needed to predict voting trends. New drugs for fighting disease are tested by collecting data under carefully controlled and monitored experimental conditions. In many marketing, political, and medical situations of these sorts, companies sometimes hire outside consultants or statisticians to help them obtain appropriate data. Regardless of whether newly minted data are gathered in-house or by paid outsiders, this type of data collection requires much more time, effort, and expense than are needed when data can be found from public or private sources.

When initiating a study, we first define our variable of interest, or **response variable.** Other variables, typically called **factors,** that may be related to the response variable of interest will also be measured. When we are able to set or manipulate the values of these factors, we have an **experimental study.** For example, a pharmaceutical company might wish to determine the most appropriate daily dose of a cholesterol-lowering drug for patients having cholesterol levels that are too high. The company can perform an experiment in which one sample of patients receives a placebo; a second sample receives some low dose; a third a higher dose; and so forth. This is an experiment because the company controls the amount of drug each group receives. The optimal daily dose can be determined by analyzing the patients' responses to the different dosage levels given.

When analysts are unable to control the factors of interest, the study is **observational.** In studies of diet and cholesterol, patients' diets are not under the analyst's control. Patients are often unwilling or unable to follow prescribed diets; doctors might simply ask patients what they eat and then look for associations between the factor *diet* and the response variable *cholesterol level*.

Asking people what they eat is an example of performing a **survey.** In general, people in a survey are asked questions about their behaviors, opinions, beliefs, and other characteristics. For instance, shoppers at a mall might be asked to fill out a short questionnaire which seeks their opinions about a new bottled water. In other observational studies, we might simply observe the behavior of people. For example, we might observe the behavior of shoppers as they look at a store display, or we might observe the interactions between students and teachers.

Exercises for Sections 1.1 and 1.2

CONCEPTS

connect

1.1 Define what we mean by a *variable,* and explain the difference between a quantitative variable and a qualitative (categorical) variable.

1.2 Below we list several variables. Which of these variables are quantitative and which are qualitative? Explain.

- **a** The dollar amount on an accounts receivable invoice.
- **b** The net profit for a company in 2013.
- **c** The stock exchange on which a company's stock is traded.
- **d** The national debt of the United States in 2013.
- **e** The advertising medium (radio, television, or print) used to promote a product.

1.3 **(1)** Discuss the difference between cross-sectional data and time series data. **(2)** If we record the total number of cars sold in 2012 by each of 10 car salespeople, are the data cross-sectional or time series data? **(3)** If we record the total number of cars sold by a particular car salesperson in each of the years 2008, 2009, 2010, 2011, and 2012, are the data cross-sectional or time series data?

1.4 Consider a medical study that is being performed to test the effect of smoking on lung cancer. Two groups of subjects are identified; one group has lung cancer and the other one doesn't. Both are asked to fill out a questionnaire containing questions about their age, sex, occupation, and number of cigarettes smoked per day. **(1)** What is the response variable? **(2)** Which are the factors? **(3)** What type of study is this (experimental or observational)?

METHODS AND APPLICATIONS

1.5 Consider the five homes in Table 1.1 (page 3). What do you think you would have to pay for a Ruby model on a treed lot?

1.6 Consider the five homes in Table 1.1 (page 3). What do you think you would have to pay for a Diamond model on a lake lot? For a Ruby model on a lake lot?

1.7 The number of Bismark X-12 electronic calculators sold at Smith's Department Stores over the past 24 months have been: 197, 211, 203, 247, 239, 269, 308, 262, 258, 256, 261, 288, 296, 276, 305, 308, 356, 393, 363, 386, 443, 308, 358, and 384. Make a time series plot of these data. That is, plot 197 versus month 1, 211 versus month 2, and so forth. What does the time series plot tell you? DS CalcSale

1.3 Populations and Samples

LO1-6 Describe the difference between a population and a sample.

We often collect data in order to study a population.

A **population** is the set of all elements about which we wish to draw conclusions.

Examples of populations include (1) all of last year's graduates of Dartmouth College's Master of Business Administration program, (2) all current MasterCard cardholders, and (3) all Buick LaCrosses that have been or will be produced this year.

We usually focus on studying one or more variables describing the population elements. If we carry out a measurement to assign a value of a variable to each and every population element, we have a *population of measurements* (sometimes called *observations*). If the population is small, it is reasonable to do this. For instance, if 150 students graduated last year from the Dartmouth College MBA program, it might be feasible to survey the graduates and to record all of their starting salaries. In general:

If we examine all of the population measurements, we say that we are conducting a **census** of the population.

Often the population that we wish to study is very large, and it is too time-consuming or costly to conduct a census. In such a situation, we select and analyze a subset (or portion) of the population elements.

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

For example, suppose that 8,742 students graduated last year from a large state university. It would probably be too time-consuming to take a census of the population of all of their starting salaries. Therefore, we would select a sample of graduates, and we would obtain and record their starting salaries. When we measure a characteristic of the elements in a sample, we have a **sample of measurements.**