Access Code(s) for Purchased Term(s):

Your access code is not your password

Store your access info for safekeeping here!

Course Name: _____

Course URL (if applicable): _____

Username (Login): _____

Password: _____

DO NOT DISCARD. SAVE THIS CARD TO ACCESS YOUR ONLINE PRODUCT.

GET MORE WITH THE MINDTAP MOBILE APP!

- Stay organized — never miss another assignment.
- Study whenever and wherever you want.
- Get a better grade with mobile quizzes and flashcards.

Technical Support
cengage.com/support

Additional Support
cengagebrain.com/shop/FAQ.html#Digital Products

CENGAGE Learning

To learn more about Cengage Learning, visit **www.cengage.com**

J00001

ISBN-13: 978-1-337-09419-1
ISBN-10: 1-337-09419-6

MindTap® Business Statistics for **Statistics for Business & Economics, Revised 13e**

ISBN: 978-1-337-09419-1

Anderson/Sweeney/Williams/Camm/Cochran

CENGAGE Learning

LET'S GET STARTED!

ACTIVATE INSTRUCTOR-ASSIGNED MATERIALS FOR THIS COURSE ONLINE USING THE ENCLOSED ACCESS CODE.

MindTap® Business Statistics, 2 Semesters
Printed Access Card for

Statistics for Business & Economics, Revised 13e

AGREEMENT URLs:

- CengageBrain Service Agreement: **www.cengagebrain.com/shop/terms.html**
- MindTap Service Agreement: **www.cengage.com/mindtap-service-agreement**
- Cengage Learning Privacy Statement: **www.cengage.com/privacy**

IMPORTANT: READ THIS BEFORE OPENING

The Access Code contained in this package (whether activated or not) cannot be resold or otherwise transferred, and cannot be returned once opened. MindTap and MindTap content may only be accessed within the country in which you have established your account. In addition, the Access Code and the online product or service accessible through activation of that Access Code is also governed by the additional terms contained in the agreements (which may include terms of use or terms of service) located at the Agreement URLs (Universal Resource Locator) listed above (the "Agreements"). Please read the Agreements carefully before opening this package, as opening the package will indicate your acceptance of the foregoing (including the Agreements). If you do not agree to the foregoing (including the Agreements), then Cengage is unwilling to provide the Access Code and online product or service to you, and you should promptly return the unopened package together with proof of purchase to the place of purchase and your purchase will be refunded.

CONGRATULATIONS!

You're about to unlock the content needed to ACE this course.

A+

You only need two things to get started:

1. The **registration link** from your instructor
2. The unique **access code** to the right

Have both? Good! Time to get started.

1. Visit your instructor-provided registration link.
2. Create an account or sign in to CengageBrain.
3. When prompted, enter the access code from this package.

You now have access to your course!

Download the MindTap mobile app for convenient access anytime, anywhere! Available on both The App Store and GooglePlay.

Did something go wrong? We can help:

My instructor didn't give me a registration link.
You should ask your instructor for the registration link or visit http://login.cengagebrain.com and enter your access code.

Access code already in use?
If your access code came with a used book, the access code may have been used before. Access codes are single use. Return to where you purchased this product for assistance.

Need Technical Support?
See back panel for complete support information.

REVISED EDITION

Anderson * Sweeney * Williams * Camm * Cochran

Statistics for Business & Economics 13e

iStockphoto.com/alienforce; iStockphoto.com/TommL

Statistics for Business & Economics

13e Revised

David R. Anderson
University of Cincinnati

Dennis J. Sweeney
University of Cincinnati

Thomas A. Williams
Rochester Institute of Technology

Jeffrey D. Camm
Wake Forest University

James J. Cochran
University of Alabama

CENGAGE Learning

Australia • Brazil • Mexico • Singapore • United Kingdom • United States

CENGAGE Learning®

***Statistics for Business and Economics*, Thirteenth Edition, Revised**
David R. Anderson, Dennis J. Sweeney, Thomas A. Williams, Jeffrey D. Camm, James J. Cochran

Vice President, General Manager: Social Science and Qualitative Business: Erin Joyner
Product Director: Mike Schenk
Product Team Manager: Joe Sabatino
Senior Product Manager: Aaron Arnsparger
Content Developer: Anne Merrill
Senior Marketing Manager: Nate Anderson
Digital Content Designer: Brandon Foltz
Product Assistant: Audrey Jacobs
Senior Content Project Manager: Colleen Farmer
Manufacturing Planner: Ron Montgomery
Production Service: MPS Limited
Senior Art Director: Michelle Kunkler
Internal Designer: Beckmeyer Design
Cover Designer: Beckmeyer Design
Cover Image: iStockphoto.com/alienforce
Intellectual Property
Analyst: Brittani Morgan
Project Manager: Nick Barrows

Library of Congress Control Number: 2016954099

Package ISBN: 978-1-337-09416-0

Book only ISBN: 978-1-305-88188-4

Cengage Learning
20 Channel Center Street
Boston, MA 02210
USA

Cengage Learning is a leading provider of customized learning solutions with employees residing in nearly 40 different countries and sales in more than 125 countries around the world. Find your local representative at **www.cengage.com**

Cengage Learning products are represented in Canada by Nelson Education, Ltd.

To learn more about Cengage Learning Solutions, visit **www.cengage.com**

Purchase any of our products at your local college store or at our preferred online store **www.cengagebrain.com**

Printed in United States of America
Print Number: 01 Print Year: 2017

Dedicated to
Marcia, Cherri, Robbie, Karen, and Teresa

Brief Contents

Contents

Preface

This text is the revised 13th edition of *STATISTICS FOR BUSINESS AND ECONOMICS.* The revised edition updates the material in *STATISTICS FOR BUSINESS ECONOMICS* 13e for use with Microsoft Excel 2016 and Minitab 17. Current users of the 13th edition will find changes to the chapter-ending appendices, which now describe Excel 2016 and Minitab 17 procedures. In addition to the updated the chapter-ending appendices, we have updated the appendix to the book entitled Microsoft Excel 2016 and Tools for Statistical Analysis. This appendix provides an introduction to Excel 2016 and its tools for statistical analysis. Several of Excel's statistical functions have been upgraded and improved.

The remainder of this preface describes the authors' objectives in writing *STATISTICS FOR BUSINESS AND ECONOMICS* and the major changes that were made in developing the 13th edition. The purpose of the text is to give students, primarily those in the fields of business administration and economics, a conceptual introduction to the field of statistics and its many applications. The text is applications-oriented and written with the needs of the nonmathematician in mind; the mathematical prerequisite is understanding of algebra.

Applications of data analysis and statistical methodology are an integral part of the organization and presentation of the text material. The discussion and development of each technique is presented in an application setting, with the statistical results providing insights to decisions and solutions to problems.

Although the book is applications oriented, we have taken care to provide sound methodological development and to use notation that is generally accepted for the topic being covered. Hence, students will find that this text provides good preparation for the study of more advanced statistical material. A bibliography to guide further study is included as an appendix.

The text introduces the student to the software packages of Minitab 17 and Microsoft® Office Excel 2016 and emphasizes the role of computer software in the application of statistical analysis. Minitab is illustrated as it is one of the leading statistical software packages for both education and statistical practice. Excel is not a statistical software package, but the wide availability and use of Excel make it important for students to understand the statistical capabilities of this package. Minitab and Excel procedures are provided in appendices so that instructors have the flexibility of using as much computer emphasis as desired for the course.

Changes in the Thirteenth Edition

We appreciate the acceptance and positive response to the previous editions of *Statistics for Business and Economics.* Accordingly, in making modifications for this new edition, we have maintained the presentation style and readability of those editions. There have been many changes made throughout the text to enhance its educational effectiveness. The most substantial changes in the new edition are summarized here.

Content Revisions

- **Data and Statistics—Chapter 1.** We have expanded our section on data mining to include a discussion of big data. We have added a new section on analytics. We have also placed greater emphasis on the distinction between observed and experimental data.

- **Descriptive Statistics: Tabular and Graphical Displays—Chapter 2.** We have added instructions on how to use Excel's recommended charts option to Appendix 2.2 at the end of this chapter. This new Excel functionality produces a gallery of suggested charts based on the data selected by the user and can help students identify the most appropriate chart(s) to use to depict their data.
- **Descriptive Statistics: Numerical Measures—Chapter 3.** We now use the method for calculating percentiles that is recommended by the National Institute of Standards and Technology (NIST). In addition to being the standard recommended by NIST, this approach is also used by a wide variety of software. The NIST recommended approach for calculating percentiles is used throughout the textbook wherever percentiles are used (for example, when creating a box plot or when calculating quantiles or an interquartile range).
- **Introduction to Probability—Chapter 4.** The discussion on experiments has been updated to draw a more clear distinction between random and designed experiments. This distinction makes it easier to understand the differences in the discussion of experiments in the probability chapters (Chapters 4, 5, and 6) and the experimental design chapter (Chapter 13).
- **Software.** We have revised all step-by-step instructions in the software appendices and all figures throughout the book that feature software output to reflect Excel 2016 and Minitab 17. This provides students exposure to and experience with the current versions of two of the most commonly used software for statistical analysis in business. In this latest edition, we no longer provide discussion of the use of StatTools.
- **Case Problems.** We have added two new case problems in this addition; the total number of cases is 33. One new probability modeling case has been added to Chapter 5, and one new simple linear regression case appears in Chapter 14. The 33 case problems in this book provide students the opportunity to work on more complex problems, analyze larger data sets, and prepare managerial reports based on the results of their analyses.
- **Examples and Exercises Based on Real Data.** We continue to make a substantial effort to update our text examples and exercises with the most current real data and referenced sources of statistical information. In this edition, we have added more than 180 new examples and exercises based on real data and referenced sources. Using data from sources also used by *The Wall Street Journal*, *USA Today*, *Barron's*, and others, we have drawn from actual studies and applications to develop explanations and create exercises that demonstrate the many uses of statistics in business and economics. We believe that the use of real data from interesting and relevant problems helps generate more student interest in the material and enables the student to learn about both statistical methodology and its application. The 13th edition contains more than 350 examples and exercises based on real data.

Features and Pedagogy

Authors Anderson, Sweeney, Williams, Camm, and Cochran have continued many of the features that appeared in previous editions. Important ones for students are noted here.

Methods Exercises and Applications Exercises

The end-of-section exercises are split into two parts, Methods and Applications. The Methods exercises require students to use the formulas and make the necessary computations. The Applications exercises require students to use the chapter material in real-world situations. Thus, students first focus on the computational "nuts and bolts" and then move on to the subtleties of statistical application and interpretation.

Self-Test Exercises

Certain exercises are identified as "Self-Test Exercises." Completely worked-out solutions for these exercises are provided in Appendix D. Students can attempt the Self-Test Exercises and immediately check the solution to evaluate their understanding of the concepts presented in the chapter.

Margin Annotations and Notes and Comments

Margin annotations that highlight key points and provide additional insights for the student are a key feature of this text. These annotations, which appear in the margins, are designed to provide emphasis and enhance understanding of the terms and concepts being presented in the text.

At the end of many sections, we provide Notes and Comments designed to give the student additional insights about the statistical methodology and its application. Notes and Comments include warnings about or limitations of the methodology, recommendations for application, brief descriptions of additional technical considerations, and other matters.

Data Files Accompany the Text

Over 200 data files are available on the website that accompanies the text. In previous editions, we provided data files in both Excel and Minitab formats. In this edition, to be more efficient, we provide the data files in only one format, CSV (comma separated values). In the appendices to Chapter 2, we provide instructions on how to open CSV files in both Excel and Minitab. DATAfile logos are used in the text to identify the data sets that are available on the website. Data sets for all case problems as well as data sets for larger exercises are included. In this edition, instead of supplying both Minitab and Excel data files, we provide data files in a single format (CSV format). This format is accessible to both Minitab and Excel. We give step-by-step instructions on how to open these files in Minitab and Excel in Appendices 2.1 and 2.2 at the end of Chapter 2.

Acknowledgments

We would like to acknowledge the work of our reviewers, who provided comments and suggestions of ways to continue to improve our text. Thanks to

AbouEl-Makarim Aboueissa, University of Southern Maine

Kathleen Arano
Fort Hays State University

Musa Ayar
Uw-baraboo/Sauk County

Kathleen Burke
SUNY Cortland

YC Chang
University of Notre Dame

David Chen
Rosemont College and Saint Joseph's University

Margaret E. Cochran
Northwestern State University of Louisiana

Thomas A. Dahlstrom
Eastern University

Anne Drougas
Dominican University

Fesseha Gebremikael Strayer University/ Calhoun Community College

Malcolm C. Gold
University of Wisconsin—Marshfield/Wood County

Joel Goldstein
Western Connecticut State University

Jim Grant
Lewis & Clark College

Reidar Hagtvedt
University of Alberta School of Business

Clifford B. Hawley
West Virginia University

Vance A. Hughey
Western Nevada College

Tony Hunnicutt
Ouachita Technical College

Stacey M. Jones
Albers School of Business and Economics, Seattle University

Dukpa Kim
University of Virginia

Rajaram Krishnan
Earlham College

Robert J. Lemke
Lake Forest College

Philip J. Mizzi
Arizona State University

Mehdi Mohaghegh
Norwich University

Mihail Motzev
Walla Walla University

Somnath Mukhopadhyay
The University of Texas at El Paso

Kenneth E. Murphy
Chapman University

Ogbonnaya John Nwoha
Grambling State University

Claudiney Pereira
Tulane University

J. G. Pitt
University of Toronto

Scott A. Redenius
Brandeis University

Sandra Robertson
Thomas Nelson Community College

Sunil Sapra
California State University, Los Angeles

Kyle Vann Scott
Snead State Community College

Rodney E. Stanley
Tennessee State University

Jennifer Strehler
Oakton Community College

Ronald Stunda
Valdosta State University

Cindy van Es
Cornell University

Jennifer VanGilder
Ursinus College

Jacqueline Wroughton
Northern Kentucky University

Dmitry Yarushkin
Grand View University

David Zimmer
Western Kentucky University

We continue to owe a debt to our many colleagues and friends for their helpful comments and suggestions in the development of this and earlier editions of our text. Among them are:

Mohammad Ahmadi
University of Tennessee at Chattanooga

Lari Arjomand
Clayton College and State University

Robert Balough
Clarion University

Philip Boudreaux
University of Louisiana

Mike Bourke
Houston Baptist University

James Brannon
University of Wisconsin—Oshkosh

John Bryant
University of Pittsburgh

Peter Bryant
University of Colorado

Terri L. Byczkowski
University of Cincinnati

Robert Carver
Stonehill College

Richard Claycombe
McDaniel College

Robert Cochran
University of Wyoming

Robert Collins
Marquette University

David W. Cravens
Texas Christian University

Tom Dahlstrom
Eastern College

Gopal Dorai
William Patterson University

Nicholas Farnum
California State University—Fullerton

Donald Gren
Salt Lake Community College

Paul Guy
California State University—Chico

Clifford Hawley
West Virginia University

Jim Hightower
California State University, Fullerton

Alan Humphrey
University of Rhode Island

Ann Hussein
Philadelphia College of Textiles and Science

C. Thomas Innis
University of Cincinnati

Ben Isselhardt
Rochester Institute of Technology

Jeffery Jarrett
University of Rhode Island

Ronald Klimberg
St. Joseph's University

David A. Kravitz
George Mason University

David Krueger
St. Cloud State University

John Leschke
University of Virginia

Martin S. Levy
University of Cincinnati

John S. Loucks
St. Edward's University

David Lucking-Reiley
Vanderbilt University

Bala Maniam
Sam Houston State University

Don Marx
University of Alaska, Anchorage

Tom McCullough
University of California—Berkeley

Ronald W. Michener
University of Virginia

Glenn Milligan
Ohio State University

Mitchell Muesham
Sam Houston State University

Roger Myerson
Northwestern University

Richard O'Connell
Miami University of Ohio

Alan Olinsky
Bryant College

Ceyhun Ozgur
Valparaiso University

Tom Pray
Rochester Institute of Technology

Harold Rahmlow
St. Joseph's University

H. V. Ramakrishna
Penn State University at Great Valley

Tom Ryan
Case Western Reserve University

Bill Seaver
University of Tennessee

Alan Smith
Robert Morris College

Willbann Terpening
Gonzaga University

Ted Tsukahara
St. Mary's College of California

Hroki Tsurumi
Rutgers University

David Tufte
University of New Orleans

Victor Ukpolo
Austin Peay State University

Ebenge Usip
Youngstown State University

Cindy Van Es
Cornell University

Jack Vaughn
University of Texas-El Paso

Andrew Welki
John Carroll University

Ari Wijetunga
Morehead State University

J. E. Willis
Louisiana State University

Mustafa Yilmaz
Northeastern University

Gary Yoshimoto
St. Cloud State University

Yan Yu
University of Cincinnati

Charles Zimmerman
Robert Morris College

We thank our associates from business and industry who supplied the Statistics in Practice features. We recognize them individually by a credit line in each of the articles. We are also indebted to our product manager, Aaron Arnsparger; our content developer, Anne Merrill; our content project manager, Colleen Farmer; our project manager at MPS Limited, Manoj Kumar; digital content designer, Brandon Foltz; and others at Cengage for their editorial counsel and support during the preparation of this text.

David R. Anderson
Dennis J. Sweeney
Thomas A. Williams
Jeffrey D. Camm
James J. Cochran

About the Authors

David R. Anderson. David R. Anderson is Professor Emeritus of Quantitative Analysis in the College of Business Administration at the University of Cincinnati. Born in Grand Forks, North Dakota, he earned his B.S., M.S., and Ph.D. degrees from Purdue University. Professor Anderson has served as Head of the Department of Quantitative Analysis and Operations Management and as Associate Dean of the College of Business Administration at the University of Cincinnati. In addition, he was the coordinator of the College's first Executive Program.

At the University of Cincinnati, Professor Anderson has taught introductory statistics for business students as well as graduate-level courses in regression analysis, multivariate analysis, and management science. He has also taught statistical courses at the Department of Labor in Washington, D.C. He has been honored with nominations and awards for excellence in teaching and excellence in service to student organizations.

Professor Anderson has coauthored 10 textbooks in the areas of statistics, management science, linear programming, and production and operations management. He is an active consultant in the field of sampling and statistical methods.

Dennis J. Sweeney. Dennis J. Sweeney is Professor Emeritus of Quantitative Analysis and Founder of the Center for Productivity Improvement at the University of Cincinnati. Born in Des Moines, Iowa, he earned a B.S.B.A. degree from Drake University and his M.B.A. and D.B.A. degrees from Indiana University, where he was an NDEA Fellow. Professor Sweeney has worked in the management science group at Procter & Gamble and spent a year as a visiting professor at Duke University. Professor Sweeney served as Head of the Department of Quantitative Analysis and as Associate Dean of the College of Business Administration at the University of Cincinnati.

Professor Sweeney has published more than 30 articles and monographs in the area of management science and statistics. The National Science Foundation, IBM, Procter & Gamble, Federated Department Stores, Kroger, and Cincinnati Gas & Electric have funded his research, which has been published in *Management Science, Operations Research, Mathematical Programming, Decision Sciences,* and other journals.

Professor Sweeney has coauthored 10 textbooks in the areas of statistics, management science, linear programming, and production and operations management.

Thomas A. Williams. Thomas A. Williams is Professor Emeritus of Management Science in the College of Business at Rochester Institute of Technology. Born in Elmira, New York, he earned his B.S. degree at Clarkson University. He did his graduate work at Rensselaer Polytechnic Institute, where he received his M.S. and Ph.D. degrees.

Before joining the College of Business at RIT, Professor Williams served for seven years as a faculty member in the College of Business Administration at the University of Cincinnati, where he developed the undergraduate program in Information Systems and then served as its coordinator. At RIT he was the first chairman of the Decision Sciences Department. He teaches courses in management science and statistics, as well as graduate courses in regression and decision analysis.

Professor Williams is the coauthor of 11 textbooks in the areas of management science, statistics, production and operations management, and mathematics. He has been a consultant for numerous *Fortune* 500 companies and has worked on projects ranging from the use of data analysis to the development of large-scale regression models.

Jeffrey D. Camm. Jeffrey D. Camm is the Inmar Presidential Chair and Associate Dean of Analytics in the School of Business at Wake Forest University. Born in Cincinnati, Ohio, he holds a B.S. from Xavier University (Ohio) and a Ph.D. from Clemson University. Prior to joining the faculty at Wake Forest, he was on the faculty of the University of Cincinnati. He has also been a visiting scholar at Stanford University and a visiting professor of business administration at the Tuck School of Business at Dartmouth College.

Dr. Camm has published over 30 papers in the general area of optimization applied to problems in operations management and marketing. He has published his research in *Science*, *Management Science*, *Operations Research*, *Interfaces*, and other professional journals. Dr. Camm was named the Dornoff Fellow of Teaching Excellence at the University of Cincinnati and he was the 2006 recipient of the INFORMS Prize for the Teaching of Operations Research Practice. A firm believer in practicing what he preaches, he has served as an operations research consultant to numerous companies and government agencies. From 2005 to 2010 he served as editor-in-chief of *Interfaces* and is currently on the editorial board of *INFORMS Transactions on Education*.

James J. Cochran. James J. Cochran is Professor of Applied Statistics and the Rogers-Spivey Faculty Fellow at the University of Alabama. Born in Dayton, Ohio, he earned his B.S., M.S., and M.B.A. degrees from Wright State University and a Ph.D. from the University of Cincinnati. He has been at the University of Alabama since 2014 and has been a visiting scholar at Stanford University, Universidad de Talca, the University of South Africa and Pole Universitaire Leonard de Vinci.

Professor Cochran has published over three dozen papers in the development and application of operations research and statistical methods. He has published his research *in Management Science, The American Statistician, Communications in Statistics—Theory and Methods, Annals of operations Research, European Journal of Operational Research, Journal of Combinatorial Optimization. Interfaces, Statistics and Probability Letters,* and other professional journals. He was the 2008 recipient of the INFORMS Prize for the Teaching of Operations Research Practice and the 2010 recipient of the Mu Sigma Rho Statistical Education Award. Professor Cochran was elected to the International Statistics Institute in 2005 and named a Fellow of the American Statistical Association in 2011. He also received the Founders Award in 2014 and the Karl E. Peace Award in 2015 from the American Statistical Association. A strong advocate for effective statistics and operations research education as a means of improving the quality of applications to real problems, Professor Cochran has organized and chaired teaching effectiveness workshops in Montevideo, Uruguay; Cape Town, South Africa; Cartagena, Colombia; Jaipur, India; Buenos Aires, Argentina; Nairobi, Kenya; Buea, Cameroon; Kathmandu, Nepal; Osijek, Croatia; Havana, Cuba; and Ulaanbaatar, Mongolia. He has served as an operations research consultant to numerous companies and not-for-profit organizations. He served as editor-in-chief of *INFORMS Transactions on Education* from 2006 to 2012 and is on the editorial board of *Interfaces, International Transactions in Operational Research,* and *Significance.*

CHAPTER 1

Data and Statistics

CONTENTS

STATISTICS *in* PRACTICE

*BLOOMBERG BUSINESSWEEK**

NEW YORK, NEW YORK

With a global circulation of more than 1 million, *Bloomberg Businessweek* is one of the most widely read business magazines in the world. Bloomberg's 1700 reporters in 145 service bureaus around the world enable *Bloomberg Businessweek* to deliver a variety of articles of interest to the global business and economic community. Along with feature articles on current topics, the magazine contains articles on international business, economic analysis, information processing, and science and technology. Information in the feature articles and the regular sections helps readers stay abreast of current developments and assess the impact of those developments on business and economic conditions.

Most issues of *Bloomberg Businessweek,* formerly *BusinessWeek,* provide an in-depth report on a topic of current interest. Often, the in-depth reports contain statistical facts and summaries that help the reader understand the business and economic information. Examples of articles and reports include the impact of businesses moving important work to cloud computing, the crisis facing the U.S. Postal Service, and why the debt crisis is even worse than we think. In addition, *Bloomberg Businessweek* provides a variety of statistics about the state of the economy, including production indexes, stock prices, mutual funds, and interest rates.

Bloomberg Businessweek also uses statistics and statistical information in managing its own business. For example, an annual survey of subscribers helps the company learn about subscriber demographics, reading habits, likely purchases, lifestyles, and so on. *Bloomberg Businessweek* managers use statistical summaries from the survey to provide better services to subscribers and advertisers. One recent North American subscriber survey indicated that 90% of *Bloomberg Businessweek* subscribers use a personal computer at home and that 64% of *Bloomberg Businessweek* subscribers are involved with computer purchases at work. Such statistics alert *Bloomberg Businessweek* managers to subscriber interest in articles about new developments in computers. The results of the subscriber survey are also made available to potential advertisers. The high percentage of subscribers using personal computers at home and the high percentage of subscribers involved with computer purchases at work would be an incentive for a computer manufacturer to consider advertising in *Bloomberg Businessweek.*

In this chapter, we discuss the types of data available for statistical analysis and describe how the data are obtained. We introduce descriptive statistics and statistical inference as ways of converting data into meaningful and easily interpreted statistical information.

© Imaginechina/Corbis

Bloomberg Businessweek uses statistical facts and summaries in many of its articles.

*The authors are indebted to Charlene Trentham, Research Manager, for providing this Statistics in Practice.

Frequently, we see the following types of statements in newspapers and magazines:

- Against the U.S. dollar, the euro has lost nearly 30% of its value in the last year; the Australian dollar lost almost 20% (*The Economist*, April 25th–May 1st, 2015).
- A survey conducted by the Pew Research Center reported that 68% of Internet users believe current laws are not good enough in protecting people's privacy online (*The Wall Street Journal*, March 24, 2014).

- VW Group's U.S. sales continue to slide, with total sales off by 13% from last January, to 36,930 vehicles (*Panorama*, March 2014).
- A poll of 1,320 corporate recruiters indicated that 68% of the recruiters ranked communication skills as one of the top five most important skills for new hires (*Bloomberg Businessweek* April 13–April 19, 2015).
- The California State Teachers' Retirement System has $154.3 billion under management (*Bloomberg Businessweek*, January 21–January 27, 2013).
- At a Sotheby's art auction held on February 5, 2013, Pablo Picasso's painting *Woman Sitting Near a Window* sold for $45 million (*The Wall Street Journal*, February 15, 2013).
- Over the past three months, the industry average for sales incentives per vehicle by GM, Chrysler, Ford, Toyota, and Honda was $2336 (*The Wall Street Journal*, February 14, 2013).

The numerical facts in the preceding statements—30%, 20%, 68%, 13%, 36,930, 1320, 68%, $154.3 billion, $45 million, $2336—are called **statistics**. In this usage, the term *statistics* refers to numerical facts such as averages, medians, percentages, and maximums that help us understand a variety of business and economic situations. However, as you will see, the field, or subject, of statistics involves much more than numerical facts. In a broader sense, statistics is the art and science of collecting, analyzing, presenting, and interpreting data. Particularly in business and economics, the information provided by collecting, analyzing, presenting, and interpreting data gives managers and decision makers a better understanding of the business and economic environment and thus enables them to make more informed and better decisions. In this text, we emphasize the use of statistics for business and economic decision making.

Chapter 1 begins with some illustrations of the applications of statistics in business and economics. In Section 1.2 we define the term *data* and introduce the concept of a data set. This section also introduces key terms such as *variables* and *observations,* discusses the difference between quantitative and categorical data, and illustrates the uses of cross-sectional and time series data. Section 1.3 discusses how data can be obtained from existing sources or through survey and experimental studies designed to obtain new data. The important role that the Internet now plays in obtaining data is also highlighted. The uses of data in developing descriptive statistics and in making statistical inferences are described in Sections 1.4 and 1.5. The last four sections of Chapter 1 provide an introduction to business analytics and the role statistics plays in it, an introduction to big data and data mining, the role of the computer in statistical analysis, and a discussion of ethical guidelines for statistical practice.

Applications in Business and Economics

In today's global business and economic environment, anyone can access vast amounts of statistical information. The most successful managers and decision makers understand the information and know how to use it effectively. In this section, we provide examples that illustrate some of the uses of statistics in business and economics.

Accounting

Public accounting firms use statistical sampling procedures when conducting audits for their clients. For instance, suppose an accounting firm wants to determine whether the amount of accounts receivable shown on a client's balance sheet fairly represents the actual amount of accounts receivable. Usually the large number of individual accounts receivable makes

reviewing and validating every account too time-consuming and expensive. As common practice in such situations, the audit staff selects a subset of the accounts called a sample. After reviewing the accuracy of the sampled accounts, the auditors draw a conclusion as to whether the accounts receivable amount shown on the client's balance sheet is acceptable.

Finance

Financial analysts use a variety of statistical information to guide their investment recommendations. In the case of stocks, analysts review financial data such as price/earnings ratios and dividend yields. By comparing the information for an individual stock with information about the stock market averages, an analyst can begin to draw a conclusion as to whether the stock is a good investment. For example, *The Wall Street Journal* (June 6, 2015) reported that the average dividend yield for the S&P 500 companies was 2%. Microsoft showed a dividend yield of 1.95%. In this case, the statistical information on dividend yield indicates a lower dividend yield for Microsoft than the average dividend yield for the S&P 500 companies. This and other information about Microsoft would help the analyst make an informed buy, sell, or hold recommendation for Microsoft stock.

Marketing

Electronic scanners at retail checkout counters collect data for a variety of marketing research applications. For example, data suppliers such as ACNielsen and Information Resources, Inc., purchase point-of-sale scanner data from grocery stores, process the data, and then sell statistical summaries of the data to manufacturers. Manufacturers spend hundreds of thousands of dollars per product category to obtain this type of scanner data. Manufacturers also purchase data and statistical summaries on promotional activities such as special pricing and the use of in-store displays. Brand managers can review the scanner statistics and the promotional activity statistics to gain a better understanding of the relationship between promotional activities and sales. Such analyses often prove helpful in establishing future marketing strategies for the various products.

Production

Today's emphasis on quality makes quality control an important application of statistics in production. A variety of statistical quality control charts are used to monitor the output of a production process. In particular, an x-bar chart can be used to monitor the average output. Suppose, for example, that a machine fills containers with 12 ounces of a soft drink. Periodically, a production worker selects a sample of containers and computes the average number of ounces in the sample. This average, or x-bar value, is plotted on an x-bar chart. A plotted value above the chart's upper control limit indicates overfilling, and a plotted value below the chart's lower control limit indicates underfilling. The process is termed "in control" and allowed to continue as long as the plotted x-bar values fall between the chart's upper and lower control limits. Properly interpreted, an x-bar chart can help determine when adjustments are necessary to correct a production process.

Economics

Economists frequently provide forecasts about the future of the economy or some aspect of it. They use a variety of statistical information in making such forecasts. For instance, in forecasting inflation rates, economists use statistical information on such indicators as the Producer Price Index, the unemployment rate, and manufacturing capacity utilization. Often these statistical indicators are entered into computerized forecasting models that predict inflation rates.

Information Systems

Information systems administrators are responsible for the day-to-day operation of an organization's computer networks. A variety of statistical information helps administrators assess the performance of computer networks, including local area networks (LANs), wide area networks (WANs), network segments, intranets, and other data communication systems. Statistics such as the mean number of users on the system, the proportion of time any component of the system is down, and the proportion of bandwidth utilized at various times of the day are examples of statistical information that help the system administrator better understand and manage the computer network.

Applications of statistics such as those described in this section are an integral part of this text. Such examples provide an overview of the breadth of statistical applications. To supplement these examples, practitioners in the fields of business and economics provided chapter-opening Statistics in Practice articles that introduce the material covered in each chapter. The Statistics in Practice applications show the importance of statistics in a wide variety of business and economic situations.

1.2 Data

Data are the facts and figures collected, analyzed, and summarized for presentation and interpretation. All the data collected in a particular study are referred to as the **data set** for the study. Table 1.1 shows a data set containing information for 60 nations that participate in the World Trade Organization. The World Trade Organization encourages the free flow of international trade and provides a forum for resolving trade disputes.

Elements, Variables, and Observations

Elements are the entities on which data are collected. Each nation listed in Table 1.1 is an element with the nation or element name shown in the first column. With 60 nations, the data set contains 60 elements.

A **variable** is a characteristic of interest for the elements. The data set in Table 1.1 includes the following five variables:

- WTO Status: The nation's membership status in the World Trade Organization; this can be either as a member or an observer.
- Per Capita GDP ($): The total market value ($) of all goods and services produced by the nation divided by the number of people in the nation; this is commonly used to compare economic productivity of the nations.
- Trade Deficit ($1000s): The difference between the total dollar value of the nation's imports and the total dollar value of the nation's exports.
- Fitch Rating: The nation's sovereign credit rating as appraised by the Fitch Group[1]; the credit ratings range from a high of AAA to a low of F and can be modified by + or −.
- Fitch Outlook: An indication of the direction the credit rating is likely to move over the upcoming two years; the outlook can be negative, stable, or positive.

Measurements collected on each variable for every element in a study provide the data. The set of measurements obtained for a particular element is called an **observation**. Referring to Table 1.1, we see that the first observation (Armenia) contains the following measurements:

[1]The Fitch Group is one of three nationally recognized statistical rating organizations designated by the U.S. Securities and Exchange Commission. The other two are Standard and Poor's and Moody's investor service.

TABLE 1.1 DATA SET FOR 60 NATIONS IN THE WORLD TRADE ORGANIZATION

DATA file
Nations

Data sets such as Nations are available on the website for this text.

Nation	WTO Status	Per Capita GDP ($)	Trade Deficit ($1000s)	Fitch Rating	Fitch Outlook
Armenia	Member	5,400	2,673,359	BB−	Stable
Australia	Member	40,800	−33,304,157	AAA	Stable
Austria	Member	41,700	12,796,558	AAA	Stable
Azerbaijan	Observer	5,400	−16,747,320	BBB−	Positive
Bahrain	Member	27,300	3,102,665	BBB	Stable
Belgium	Member	37,600	−14,930,833	AA+	Negative
Brazil	Member	11,600	−29,796,166	BBB	Stable
Bulgaria	Member	13,500	4,049,237	BBB−	Positive
Canada	Member	40,300	−1,611,380	AAA	Stable
Cape Verde	Member	4,000	874,459	B+	Stable
Chile	Member	16,100	−14,558,218	A+	Stable
China	Member	8,400	−156,705,311	A+	Stable
Colombia	Member	10,100	−1,561,199	BBB−	Stable
Costa Rica	Member	11,500	5,807,509	BB+	Stable
Croatia	Member	18,300	8,108,103	BBB−	Negative
Cyprus	Member	29,100	6,623,337	BBB	Negative
Czech Republic	Member	25,900	−10,749,467	A+	Positive
Denmark	Member	40,200	−15,057,343	AAA	Stable
Ecuador	Member	8,300	1,993,819	B−	Stable
Egypt	Member	6,500	28,486,933	BB	Negative
El Salvador	Member	7,600	5,019,363	BB	Stable
Estonia	Member	20,200	802,234	A+	Stable
France	Member	35,000	118,841,542	AAA	Stable
Georgia	Member	5,400	4,398,153	B+	Positive
Germany	Member	37,900	−213,367,685	AAA	Stable
Hungary	Member	19,600	−9,421,301	BBB−	Negative
Iceland	Member	38,000	−504,939	BB+	Stable
Ireland	Member	39,500	−59,093,323	BBB+	Negative
Israel	Member	31,000	6,722,291	A	Stable
Italy	Member	30,100	33,568,668	A+	Negative
Japan	Member	34,300	31,675,424	AA	Negative
Kazakhstan	Observer	13,000	−33,220,437	BBB	Positive
Kenya	Member	1,700	9,174,198	B+	Stable
Latvia	Member	15,400	2,448,053	BBB−	Positive
Lebanon	Observer	15,600	13,715,550	B	Stable
Lithuania	Member	18,700	3,359,641	BBB	Positive
Malaysia	Member	15,600	−39,420,064	A−	Stable
Mexico	Member	15,100	1,288,112	BBB	Stable
Peru	Member	10,000	−7,888,993	BBB	Stable
Philippines	Member	4,100	15,667,209	BB+	Stable
Poland	Member	20,100	19,552,976	A−	Stable
Portugal	Member	23,200	21,060,508	BBB−	Negative
South Korea	Member	31,700	−37,509,141	A+	Stable
Romania	Member	12,300	13,323,709	BBB−	Stable
Russia	Observer	16,700	−151,400,000	BBB	Positive
Rwanda	Member	1,300	939,222	B	Stable
Serbia	Observer	10,700	8,275,693	BB−	Stable
Seychelles	Observer	24,700	666,026	B	Stable
Singapore	Member	59,900	−27,110,421	AAA	Stable
Slovakia	Member	23,400	−2,110,626	A+	Stable
Slovenia	Member	29,100	2,310,617	AA−	Negative

South Africa	Member	11,000	3,321,801	BBB+	Stable
Sweden	Member	40,600	−10,903,251	AAA	Stable
Switzerland	Member	43,400	−27,197,873	AAA	Stable
Thailand	Member	9,700	2,049,669	BBB	Stable
Turkey	Member	14,600	71,612,947	BB+	Positive
UK	Member	35,900	162,316,831	AAA	Negative
Uruguay	Member	15,400	2,662,628	BB	Positive
USA	Member	48,100	784,438,559	AAA	Stable
Zambia	Member	1,600	−1,805,198	B+	Stable

Member, 5,400, 2,673,359, BB−, and Stable. The second observation (Australia) contains the following measurements: Member, 40,800, −33,304,157, AAA, Stable, and so on. A data set with 60 elements contains 60 observations.

Scales of Measurement

Data collection requires one of the following scales of measurement: nominal, ordinal, interval, or ratio. The scale of measurement determines the amount of information contained in the data and indicates the most appropriate data summarization and statistical analyses.

When the data for a variable consist of labels or names used to identify an attribute of the element, the scale of measurement is considered a **nominal scale**. For example, referring to the data in Table 1.1, the scale of measurement for the WTO Status variable is nominal because the data "member" and "observer" are labels used to identify the status category for the nation. In cases where the scale of measurement is nominal, a numerical code as well as a nonnumerical label may be used. For example, to facilitate data collection and to prepare the data for entry into a computer database, we might use a numerical code for the WTO Status variable by letting 1 denote a member nation in the World Trade Organization and 2 denote an observer nation. The scale of measurement is nominal even though the data appear as numerical values.

The scale of measurement for a variable is considered an **ordinal scale** if the data exhibit the properties of nominal data and in addition, the order or rank of the data is meaningful. For example, referring to the data in Table 1.1, the scale of measurement for the Fitch Rating is ordinal because the rating labels which range from AAA to F can be rank ordered from best credit rating AAA to poorest credit rating F. The rating letters provide the labels similar to nominal data, but in addition, the data can also be ranked or ordered based on the credit rating, which makes the measurement scale ordinal. Ordinal data can also be recorded by a numerical code, for example, your class rank in school.

The scale of measurement for a variable is an **interval scale** if the data have all the properties of ordinal data and the interval between values is expressed in terms of a fixed unit of measure. Interval data are always numerical. College admission SAT scores are an example of interval-scaled data. For example, three students with SAT math scores of 620, 550, and 470 can be ranked or ordered in terms of best performance to poorest performance in math. In addition, the differences between the scores are meaningful. For instance, student 1 scored $620 - 550 = 70$ points more than student 2, while student 2 scored $550 - 470 = 80$ points more than student 3.

The scale of measurement for a variable is a **ratio scale** if the data have all the properties of interval data and the ratio of two values is meaningful. Variables such as distance, height, weight, and time use the ratio scale of measurement. This scale requires that a zero value be included to indicate that nothing exists for the variable at the zero point.

For example, consider the cost of an automobile. A zero value for the cost would indicate that the automobile has no cost and is free. In addition, if we compare the cost of $30,000 for one automobile to the cost of $15,000 for a second automobile, the ratio property shows that the first automobile is $30,000/$15,000 = 2 times, or twice, the cost of the second automobile.

Categorical and Quantitative Data

Data can be classified as either categorical or quantitative. Data that can be grouped by specific categories are referred to as **categorical data**. Categorical data use either the nominal or ordinal scale of measurement. Data that use numeric values to indicate how much or how many are referred to as **quantitative data**. Quantitative data are obtained using either the interval or ratio scale of measurement.

The statistical method appropriate for summarizing data depends upon whether the data are categorical or quantitative.

A **categorical variable** is a variable with categorical data, and a **quantitative variable** is a variable with quantitative data. The statistical analysis appropriate for a particular variable depends upon whether the variable is categorical or quantitative. If the variable is categorical, the statistical analysis is limited. We can summarize categorical data by counting the number of observations in each category or by computing the proportion of the observations in each category. However, even when the categorical data are identified by a numerical code, arithmetic operations such as addition, subtraction, multiplication, and division do not provide meaningful results. Section 2.1 discusses ways of summarizing categorical data.

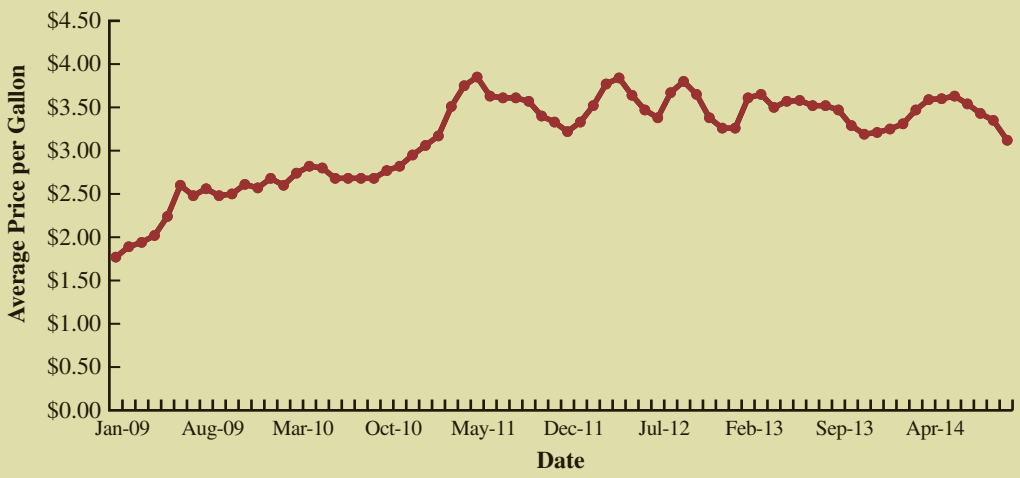
Arithmetic operations provide meaningful results for quantitative variables. For example, quantitative data may be added and then divided by the number of observations to compute the average value. This average is usually meaningful and easily interpreted. In general, more alternatives for statistical analysis are possible when data are quantitative. Section 2.2 and Chapter 3 provide ways of summarizing quantitative data.

Cross-Sectional and Time Series Data

For purposes of statistical analysis, distinguishing between cross-sectional data and time series data is important. **Cross-sectional data** are data collected at the same or approximately the same point in time. The data in Table 1.1 are cross-sectional because they describe the five variables for the 60 World Trade Organization nations at the same point in time. **Time series data** are data collected over several time periods. For example, the time series in Figure 1.1 shows the U.S. average price per gallon of conventional regular gasoline between 2009 and 2014. Between January 2009 and May 2011, the average price per gallon continued to climb steadily. Since then prices have shown more fluctuation, reaching an average price per gallon of $3.12 in October 2014.

Graphs of time series data are frequently found in business and economic publications. Such graphs help analysts understand what happened in the past, identify any trends over time, and project future values for the time series. The graphs of time series data can take on a variety of forms, as shown in Figure 1.2. With a little study, these graphs are usually easy to understand and interpret. For example, Panel (A) in Figure 1.2 is a graph that shows the Dow Jones Industrial Average Index from 2004 to 2014. In November 2004, the popular stock market index was near 10,000. The index rose to slightly over 14,000 in October 2007. However, notice the sharp decline in the time series after the high in 2007. By February 2009, poor economic conditions had caused the Dow Jones Industrial Average Index to return to the 7000 level. This was a scary and discouraging period for investors. However, by late 2009, the index was showing a recovery by reaching 10,000. The index has climbed steadily since then and was above 17,500 in November 2014.

FIGURE 1.1 U.S. AVERAGE PRICE PER GALLON FOR CONVENTIONAL REGULAR GASOLINE



Source: Energy Information Administration, U.S. Department of Energy, March 2014.

The graph in Panel (B) shows the net income of McDonald's Inc. from 2005 to 2013. The declining economic conditions in 2008 and 2009 were actually beneficial to McDonald's as the company's net income rose to all-time highs. The growth in McDonald's net income showed that the company was thriving during the economic downturn as people were cutting back on the more expensive sit-down restaurants and seeking less-expensive alternatives offered by McDonald's. McDonald's net income continued to new all-time highs in 2010 and 2011, decreased slightly in 2012, and increased in 2013.

Panel (C) shows the time series for the occupancy rate of hotels in South Florida over a one-year period. The highest occupancy rates, 95% and 98%, occur during the months of February and March when the climate of South Florida is attractive to tourists. In fact, January to April of each year is typically the high-occupancy season for South Florida hotels. On the other hand, note the low occupancy rates during the months of August to October, with the lowest occupancy rate of 50% occurring in September. High temperatures and the hurricane season are the primary reasons for the drop in hotel occupancy during this period.

NOTES AND COMMENTS

1. An observation is the set of measurements obtained for each element in a data set. Hence, the number of observations is always the same as the number of elements. The number of measurements obtained for each element equals the number of variables. Hence, the total number of data items can be determined by multiplying the number of observations by the number of variables.
2. Quantitative data may be discrete or continuous. Quantitative data that measure how many (e.g., number of calls received in 5 minutes) are discrete. Quantitative data that measure how much (e.g., weight or time) are continuous because no separation occurs between the possible data values.