Essentials of Business Statistics

The McGraw-Hill Education Series in Operations and Decision Sciences

Supply Chain Management

Benton
Purchasing and Supply Chain Management
Third Edition

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

Burt, Petcavage, and Pinkerton
Supply Management
Eighth Edition

Johnson
Purchasing and Supply Management
Sixteenth Edition

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

Stock and Manrodt
Fundamentals of Supply Chain Management

Project Management

Brown and Hyer
Managing Projects: A Team-Based Approach

Larson and Gray
Project Management: The Managerial Process
Seventh Edition

Service Operations Management

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

Management Science

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

Business Research Methods

Cooper and Schindler
Business Research Methods
Thirteenth Edition

Business Forecasting

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

Linear Statistics and Regression

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

Business Systems Dynamics

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

Operations Management

Cachon and Terwiesch
Operations Management
Second Edition

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

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

Jacobs and Chase
Operations and Supply Chain Management
Fifteenth Edition

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

Stevenson
Operations Management
Thirteenth Edition

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

Business Math

Slater and Wittry
Practical Business Math Procedures
Thirteenth Edition

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

Business Statistics

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

Doane and Seward
Applied Statistics in Business and Economics
Sixth Edition

Doane and Seward
Essential Statistics in Business and Economics
Third Edition

Jaggia and Kelly
Business Statistics: Communicating with Numbers
Third Edition

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

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

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

McGuckian
Connect Master: Business Statistics

Final PDF to printer

2e

Essentials of Business Statistics

Communicating with Numbers

SANJIV JAGGIA
California Polytechnic State University

ALISON KELLY
Suffolk University

Mc Graw Hill Educ

ESSENTIALS OF BUSINESS STATISTICS: COMMUNICATING WITH NUMBERS, SECOND EDITION

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

This book is printed on acid-free paper.

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

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

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

Library of Congress Cataloging-in-Publication Data

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

mheducation.com/highered

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

ABOUT THE AUTHORS

Sanjiv Jaggia

Courtesy of Sanjiv Jaggia

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

Alison Kelly

Courtesy of Alison Kelly

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

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

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

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

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

Brad McDonald, *Northern Illinois University*

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

Andrew Koch, *James Madison University*

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

Michael Gordinier, *Washington University, St. Louis*

Continuing Key Features

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

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

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

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

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

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

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

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

Elizabeth Haran, *Salem State University*

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

Jun Liu, *Georgia Southern University*

Features New to the Second Edition

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

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

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

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

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

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

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

Here are other noteworthy changes:

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

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

Integrated Introductory Cases

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

©Mark Bowden/Getty Images

Introductory Case

Investment Decision

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

SYNOPSIS OF INTRODUCTORY CASE

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

©Ingram Publishing/Getty Images

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

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

Alan Chow, *University of South Alabama*

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

Dane Peterson, *Missouri State University*

and Build Skills to Communicate Results

Writing with Statistics

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

> *Writing with statistics shows that statistics **is more than number crunching.***
>
> Greg Cameron, *Brigham Young University*

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

> *This is an **excellent approach. . . .** The ability to translate numerical information into words that others can understand is critical.*
>
> Scott Bailey, *Troy University*

> ***Excellent.** Students need to become better writers.*
>
> Bob Nauss, *University of Missouri, St. Louis*

WRITING WITH STATISTICS

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

©Image Source, all rights reserved.

TABLE 6.5 Grading Scales with Absolute Grading versus Relative Grading

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

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

Professor Lang wants to use the above information to

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

Many teachers would confess that grading is one of the most difficult tasks of their profession.

Sample Report—Absolute Grading versus Relative Grading

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

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

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

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

Unique Coverage and Presentation . . .

By comparing this chapter with other books, I think that this is one of the best explanations about regression I have seen.
Cecilia Maldonado, *Georgia Southwestern State University*

Unique Coverage of Regression Analysis

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

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

This is easy for students to follow and I do get the feeling . . . the sections are spoken language.
Zhen Zhu, *University of Central Oklahoma*

Written as Taught

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

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

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

that Make the Content More Effective

Integration of Microsoft Excel®

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

Using Excel to Obtain Binomial Probabilities

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

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

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

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

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

Matthew Dean, *University of Southern Maine*

Real-World Exercises and Case Studies that Reinforce the Material

Mechanical and Applied Exercises

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

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

18. Consider the following hypothesis test:

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

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

...nsider the following hypothesis test:

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

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

20. Consider the following hypothesis test:

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

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

21. Consider the following hypotheses:

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

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

    a. If $\bar{x} = 132$ and $n = 50$, what is the conclusion at the 5% significance level?
    b. If $\bar{x} = 108$ and $n = 50$, what is the conclusion at the 10% significance level?

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

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

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

    a. Specify the null and alternative hypotheses to test whether average spending at the rival's store is more than $130.
    b. Calculate the value of the test statistic and the p-value.
    c. At the 5% significance level, what is the conclusion to the test?

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

    a. State the null and the alternative hypotheses for the test.
    b. What is the p-value?
    c. Is the sales manager's claim justified at $\alpha = 0.01$?

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

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

Dave Leupp, *University of Colorado at Colorado Springs*

Their exercises and problems are excellent!

Erl Sorensen, *Bentley University*

Features that Go Beyond the Typical

Conceptual Review

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

CONCEPTUAL REVIEW

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

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

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

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

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

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

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

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

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

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

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

Andrew Koch, *James Madison University*

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

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

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

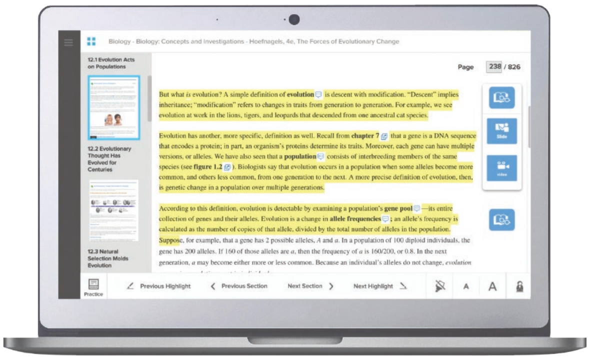
SUCCESSFUL SEMESTERS INCLUDE CONNECT

For Instructors

You're in the driver's seat.

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

65%
Less Time Grading



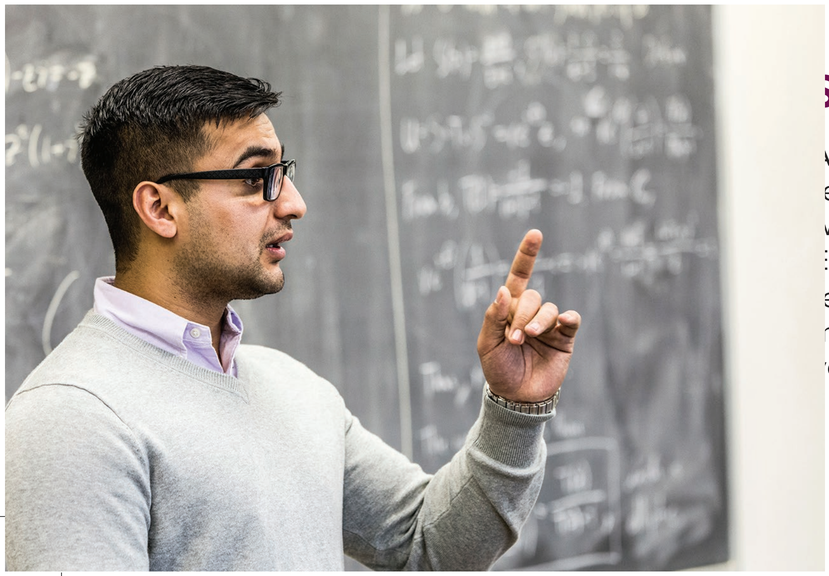
They'll thank you for it.

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

Make it simple, make it affordable.

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

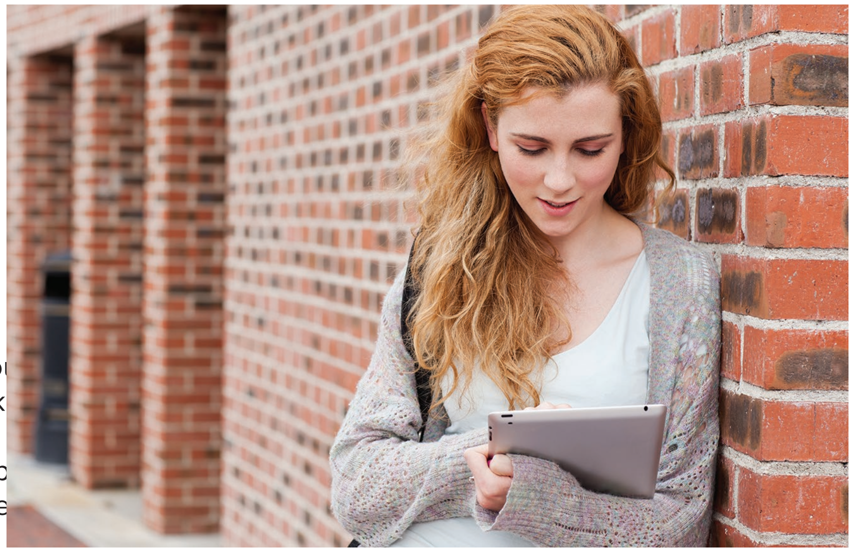
©Hill Street Studios/Tobin Rogers/Blend Images LLC



Solutions for your challenges.

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

For Students



©Shutterstock/wavebreakmedia

Effective, efficient studying.

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

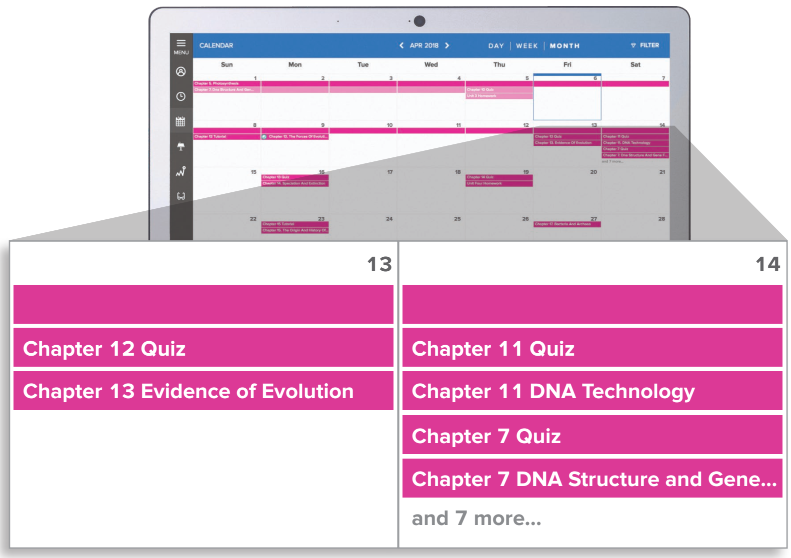
> “I really liked this app—it made it easy to study when you don't have your text-book in front of you.”
>
> - Jordan Cunningham, Eastern Washington University

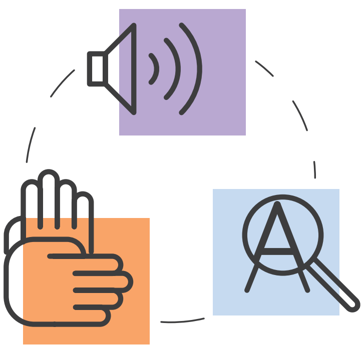
Study anytime, anywhere.

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

No surprises.

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





Learning for everyone.

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

What Resources are Available for Instructors?

Instructor Library

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

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

Tegrity Campus: Lectures 24/7

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

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

ALEKS

ALEKS

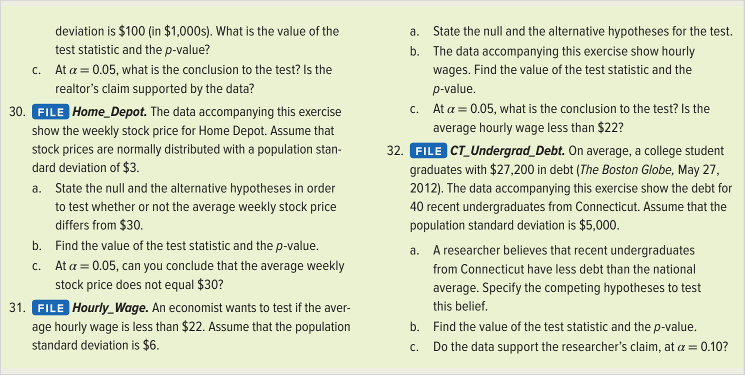
ALEKS is an assessment and learning program that provides individualized instruction in Business Statistics, Business Math, and Accounting. Available online in partnership with McGraw-Hill Education, ALEKS interacts with students much like a skilled human tutor, with the ability to assess precisely a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

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

MegaStat® for Microsoft Excel®

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

What Resources are Available for Students?


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

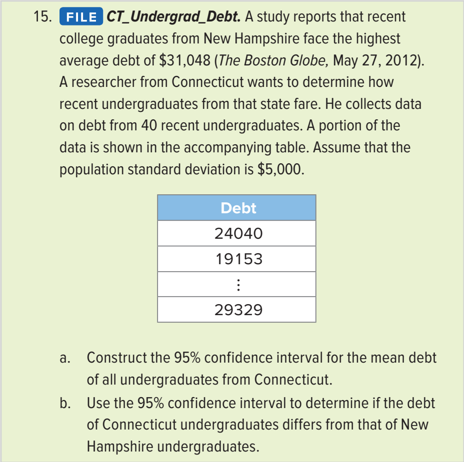
c. At $\alpha = 0.05$, what is the conclusion to the test? Is the realtor's claim supported by the data?

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

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

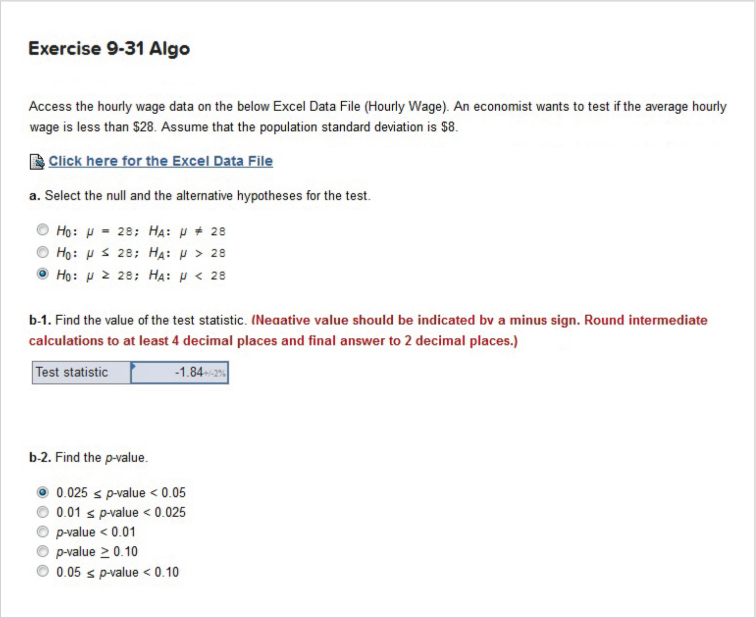
32. FILE ***CT_Undergrad_Debt.*** On average, a college student graduates with $27,200 in debt (*The Boston Globe*, May 27, 2012). The data accompanying this exercise show the debt for 40 recent undergraduates from Connecticut. Assume that the population standard deviation is $5,000.
   a. A researcher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

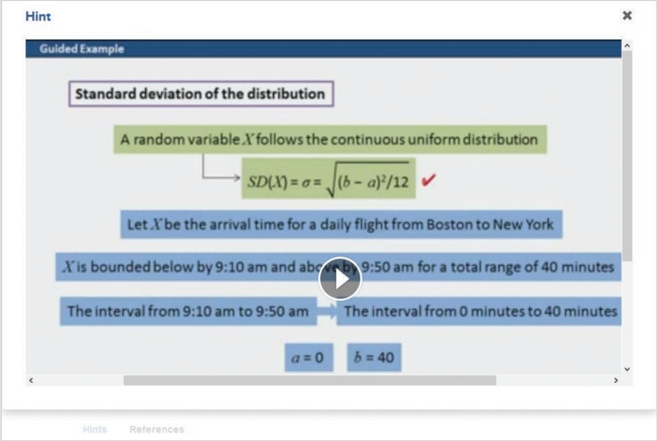
Integration of Excel Data Sets. A convenient feature is the inclusion of an Excel data file link in many problems using data files in their calculation. The link allows students to easily launch into Excel, work the problem, and return to *Connect* to key in the answer and receive feedback on their results.


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

Debt
24040
19153
⋮
29329

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





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

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

McGraw-Hill Customer Care Contact Information

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

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

ACKNOWLEDGMENTS

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

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

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

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

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

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

BRIEF CONTENTS

CONTENTS

Essentials of Business Statistics

2 Tabular and Graphical Methods

Learning Objectives

After reading this chapter you should be able to:

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

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

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

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

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

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

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

©Mitch Diamond/Photodisc/Getty Images

Introductory Case

House Prices in Southern California

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

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

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

Source: www.zillow.com.

FILE *MV_Houses*

Matthew wants to use the sample information to

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

A synopsis of this case is provided in Section 2.2.

LO 2.1

Summarize qualitative data by constructing a frequency distribution.

2.1 SUMMARIZING QUALITATIVE DATA

As we discussed in Chapter 1, nominal and ordinal data are types of qualitative data. Nominal data typically consist of observations that represent labels or names; information related to gender or race are examples. Nominal data are considered the least sophisticated form of data since all we can do with the data is categorize it. Ordinal data are stronger in the sense that we can categorize and order the data. Examples of ordinal data include the ratings of a product or a professor, where 1 represents the worst and 4 represents the best. In order to organize qualitative data, it is often useful to construct a **frequency distribution**.

FREQUENCY DISTRIBUTION FOR QUALITATIVE DATA

A frequency distribution for qualitative data groups data into categories and records the number of observations that fall into each category.

To illustrate the construction of a frequency distribution with nominal data, Table 2.2 shows the weather for the month of February (2010) in Seattle, Washington.

TABLE 2.2 Seattle Weather, February 2010

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
	1 Rainy	2 Rainy	3 Rainy	4 Rainy	5 Rainy	6 Rainy
7 Rainy	8 Rainy	9 Cloudy	10 Rainy	11 Rainy	12 Rainy	13 Rainy
14 Rainy	15 Rainy	16 Rainy	17 Sunny	18 Sunny	19 Sunny	20 Sunny
21 Sunny	22 Sunny	23 Rainy	24 Rainy	25 Rainy	26 Rainy	27 Rainy
28 Sunny						

Source: www.wunderground.com.

©sbk_20d pictures/Moment/Getty Images

We first note that the weather in Seattle is categorized as cloudy, rainy, or sunny. The first column in Table 2.3 lists these categories. Initially, we use a "tally" column to record the number of days that fall into each category. Since the first eight days of February were rainy days, we place the first eight tally marks in the rainy category; the ninth day of February was cloudy, so we place one tally mark in the cloudy category, and so on. Finally, we convert each category's total tally count into its respective numerical value in the frequency column. Since only one tally mark appears in the cloudy category, we record the value 1 as its frequency. Note that if we sum the frequency column, we obtain the sample size. A frequency distribution in its final form does not include the tally column.

TABLE 2.3 Frequency Distribution for Seattle Weather, February 2010

Weather	Tally	
Cloudy	|	1
Rainy	卌 卌 卌 卌	20
Sunny	卌 ||	7
		Total = 28 days

From the frequency distribution, we can now readily observe that the most common type of day in February was rainy since this type of day occurs with the highest frequency. In many applications, we want to compare data sets that differ in size. For example, we might want to compare the weather in February to the weather in March. However, February has 28 days (except during a leap year) and March has 31 days. In this instance, we would convert the frequency distribution to a **relative frequency distribution**. We calculate each category's relative frequency by dividing the respective category's frequency by the total number of observations. The sum of the relative frequencies should equal one, or a value very close to one due to rounding.

In Table 2.4, we convert the frequency distribution from Table 2.3 into a relative frequency distribution. Similarly, we obtain the relative frequency distribution for the month of March; the raw data for March are not shown. March had 3 cloudy days, 18 rainy days, and 10 sunny days. Each of these frequencies was then divided by 31, the number of days in the month of March.

TABLE 2.4 Relative Frequency Distribution for Seattle Weather

Weather	February 2010: Relative Frequency	March 2010: Relative Frequency
Cloudy	1/28 = 0.036	3/31 = 0.097
Rainy	20/28 = 0.714	18/31 = 0.581
Sunny	7/28 = 0.250	10/31 = 0.323
	Total = 1	Total = 1 (subject to rounding)

Source: www.wunderground.com.

We can easily convert relative frequencies into percentages by multiplying by 100. For instance, the percent of cloudy days in February and March equals 3.6% and 9.7%, respectively. From the relative frequency distribution, we can now conclude that the weather in Seattle in both February and March was predominantly rainy. However, the weather in March was a bit nicer in that approximately 32% of the days were sunny, as opposed to only 25% of the days in February.

CALCULATING RELATIVE AND PERCENT FREQUENCIES

The relative frequency for each category of a qualitative variable equals the proportion (fraction) of observations in each category. A category's relative frequency is calculated by dividing its frequency by the total number of observations. The sum of the relative frequencies should equal one.

The percent frequency for each category of a qualitative variable equals the percent (%) of observations in each category; it equals the relative frequency of the category multiplied by 100.

Pie Charts and Bar Charts

LO 2.2

Construct and interpret a pie chart and a bar chart.

We can visualize the information found in frequency distributions by constructing various graphs. Graphical representations often portray the data more dramatically, as well as simplify interpretation. A **pie chart** and a **bar chart** are two widely used graphical representations for qualitative data.

GRAPHICAL DISPLAY OF QUALITATIVE DATA: A PIE CHART

A pie chart is a segmented circle whose segments portray the relative frequencies of the categories of some qualitative variable.

A pie chart is best explained by using an example. Consider Example 2.1.

EXAMPLE 2.1

Is America having a "marriage crisis?" The answer depends on whom you ask, but nearly every study focuses on the women's liberation movement of the late 1960s and 1970s. As more and more women earned college degrees, they entered the workforce and delayed motherhood. Marriage became less necessary for their economic survival. No matter what the reason for the decline in marriage, here are some facts. In 1960, 71% of all adults in the United States were married. Today, barely half of all adults are married, just 52%. Table 2.5 shows the proportions of all adults who were married, widowed, divorced, or single in 1960 compared to those same proportions in 2010. Construct pie charts to graphically depict marital status in the United States in these two time periods.

FILE *Marital_Status*

TABLE 2.5 Marital Status, 1960 versus 2010

Marital Status	1960	2010
Married	0.71	0.52
Single	0.15	0.28
Divorced	0.05	0.14
Widowed	0.09	0.06

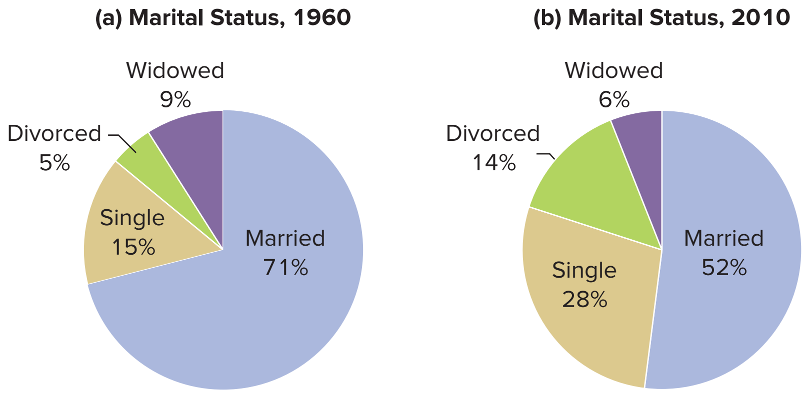
Note: Proportions for each year rounded so that they summed to one.
Source: Pew Research Center analysis of Decennial Census (1960–2000) and American Community Survey data (2008, 2010).

SOLUTION: In order to construct a pie chart, we first draw a circle. We then cut the circle into slices, or sectors such that each sector is proportional to the size of the category we wish to display. For instance, Table 2.5 shows that married adults accounted for 0.71 of all adults in 1960. Since a circle contains 360 degrees, the portion of the circle representing married adults encompasses $0.71 \times 360 = 255.6$ degrees. Similar calculations for the other three categories yield:

Single: $0.15 \times 360 = 54$ degrees
Divorced: $0.05 \times 360 = 18$ degrees
Widowed: $0.09 \times 360 = 32.4$ degrees

The same methodology can be used to construct a pie chart for marital status in 2010. Figure 2.1 shows the resulting pie charts.

FIGURE 2.1 Pie charts for marital status



Another way to graphically depict qualitative data is to construct a **bar chart**.

> **GRAPHICAL DISPLAY OF QUALITATIVE DATA: A BAR CHART**
>
> A bar chart depicts the frequency or the relative frequency for each category of the qualitative variable as a series of horizontal or vertical bars, the lengths of which are proportional to the values that are to be depicted.

We first discuss a vertical bar chart, sometimes referred to as a column chart. Here, we place each category on the horizontal axis and mark the vertical axis with an appropriate range of values for either frequency or relative frequency. The height of each bar is equal to the frequency or the relative frequency of the corresponding category. Typically, we leave space between categories to improve clarity.

Figure 2.2 shows a relative frequency bar chart for the marital status example. It is particularly useful because we can group marital status by year, emphasizing the decline in the proportion of U.S. adults who are married and the rise in the proportion of U.S. adults who are single.

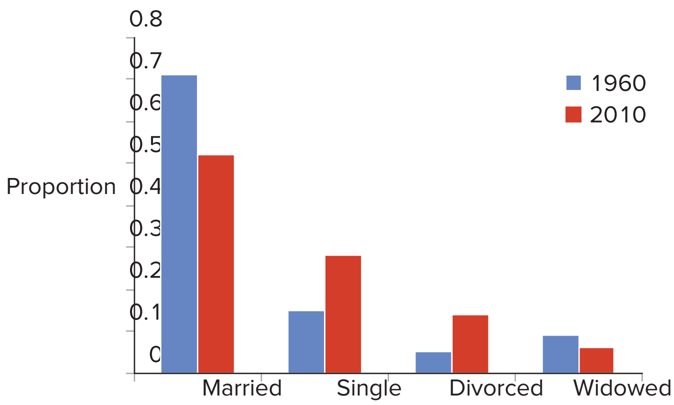


FIGURE 2.2 Marital status of U.S. adults, 1960 versus 2010

For a horizontal bar chart, we simply place each category on the vertical axis and mark the horizontal axis with an appropriate range of values for either frequency or relative frequency. For example, a recent poll asked more than 1,000 Americans: "When traveling in a non-English-speaking country, which word or phrase is most important to know in that country's language?" (Source: *Vanity Fair,* January 2, 2012). Figure 2.3 shows the results of the poll. The phrase "Thank you" earned the largest percentage of votes (38%). Fortunately, only 1% of Americans believed that the phrase "Where is McDonald's?" was of vital importance. The proportions in Figure 2.3 do not sum to one because we exclude responses with uncommon words or phrases.

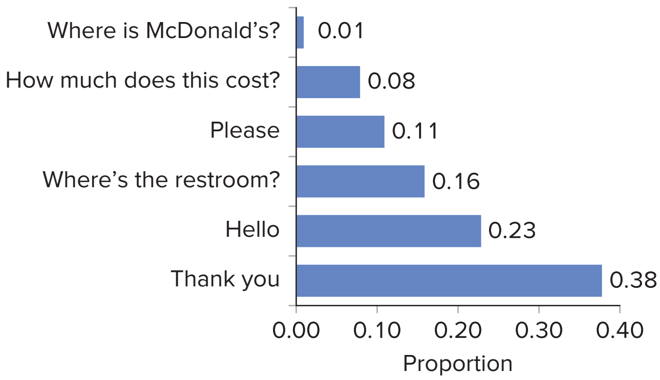


FIGURE 2.3 Results to question: "When traveling in a non-English-speaking country, which word or phrase is most important to know in that country's language?"

Cautionary Comments When Constructing or Interpreting Charts or Graphs

As with many of the statistical methods that we examine throughout this text, the possibility exists for unintentional, as well as purposeful, distortions of graphical information. As a careful researcher, you should follow these basic guidelines:

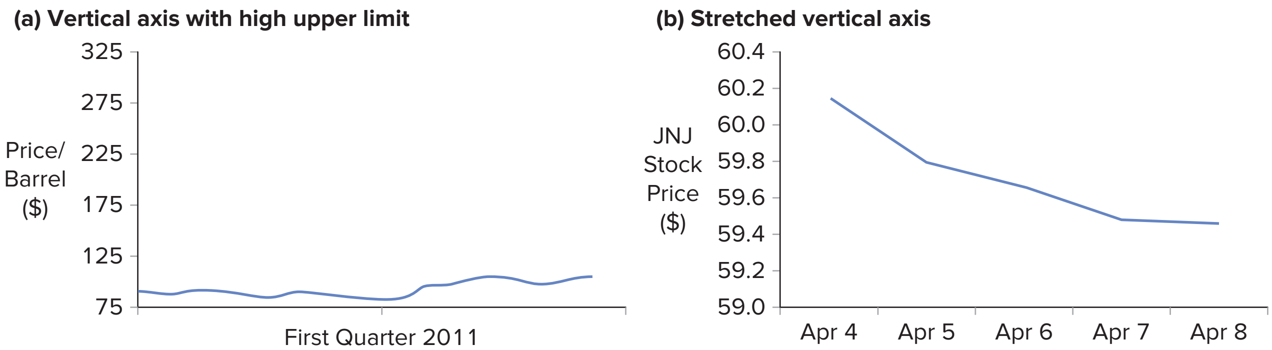
- The simplest graph should be used for a given set of data. Strive for clarity and avoid unnecessary adornments.
- Axes should be clearly marked with the numbers of their respective scales; each axis should be labeled.
- When creating a bar chart, each bar should be of the same width. Differing bar widths create distortions. The same principle holds in the next section when we discuss histograms.
- The vertical axis should not be given a very high value as an upper limit. In these instances, the data may appear compressed so that an increase (or decrease) of the data is not as apparent as it perhaps should be. For example, Figure 2.4(a) plots the daily price for a barrel of crude oil for the first quarter of 2011. Due to Middle East unrest, the price of crude oil rose from a low of \$83.13 per barrel to a high of \$106.19 per barrel, or approximately 28% $\left(= \frac{106.19 - 83.13}{83.13}\right)$. However, since Figure 2.4(a) uses a high value as an upper limit on the vertical axis (\$325), the rise in price appears dampened.

FILE *Crude_Oil*

- The vertical axis should not be stretched so that an increase (or decrease) of the data appears more pronounced than warranted. For example, Figure 2.4(b) charts the daily closing stock price for Johnson & Johnson (JNJ) for the week of April 4, 2011. It is true that the stock price declined over the week from a high of \$60.15 to a low of \$59.46; this amounts to a \$0.69 decrease or an approximate 1% decline. However, since the vertical axis is stretched, the drop in stock price appears more dramatic.

FILE *JNJ*

FIGURE 2.4 Misleading scales on vertical axes



Source: U.S. Energy Information Administration.

Source: www.finance.yahoo.com.

Using Excel to Construct a Pie Chart and a Bar Chart

Excel offers various options for displaying a pie chart. To replicate the pie chart in Figure 2.1(a), follow these steps:

A Pie Chart

FILE *Marital_Status*

A. FILE Open ***Marital_Status*** (Table 2.5).

B. First, select the category names and respective relative frequencies from the year 1960. Leave out the heading (top row). Then, from the menu, choose **Insert > Pie > 2-D Pie**. From the options given, choose the graph on the top left. (If you are having trouble finding this option after selecting **Insert**, look for the circle above **Charts**.)

C. In order to add category names and their respective percentages, select the '+' sign that appears to the right of the pie chart.

A Bar Chart

Excel provides many options for showing a bar chart. To replicate the bar chart in Figure 2.2, follow these steps:

A. FILE Open ***Marital_Status*** (Table 2.5).

B. First, select the category names and respective relative frequencies for the years 1960 and 2010. Leave out the heading (top row). Then, from the menu, choose **Insert > Column > 2-D Column**. From the options given, choose the graph on the top left. (If you are having trouble finding this option after selecting **Insert**, look for the vertical bars above **Charts**.)

C. In the legend to the right of the bar chart, Excel labels the data for the year 1960 as "Series 1" and the data for the year 2010 as "Series 2" by default. In order to edit the legend, select the legend and choose **Design > Select Data**. From the *Legend Entries,* select "Series 1," then select *Edit,* and under *Series Name,* type the new name of 1960. Follow the same steps to rename "Series 2" to 2010.

FILE
Marital_Status

EXERCISES 2.1

1. A local restaurant is committed to providing its patrons with the best dining experience possible. On a recent survey, the restaurant asked patrons to rate the quality of their entrées. The responses ranged from 1 to 5, where 1 indicated a disappointing entrée and 5 indicated an exceptional entrée. The results of the survey are as follows:

3	5	4	4	3	2	3	3	2	5	5	5
5	3	3	2	1	4	5	5	4	2	5	5
5	4	4	3	1	5	2	1	5	4	4	4

   a. Construct frequency and relative frequency distributions that summarize the survey's results.
   b. Are patrons generally satisfied with the quality of their entrées? Explain.

2. First-time patients at North Shore Family Practice are required to fill out a questionnaire that gives the doctor an overall idea of each patient's health. The first question is: "In general, what is the quality of your health?" The patient chooses Excellent, Good, Fair, or Poor. Over the past month, the responses to this question from first-time patients were:

Fair	Good	Fair	Excellent
Good	Good	Good	Poor
Excellent	Excellent	Poor	Good
Fair	Good	Good	Good
Good	Poor	Fair	Excellent
Excellent	Good	Good	Good

   a. Construct frequency and relative frequency distributions that summarize the responses to the questionnaire.
   b. What is the most common response to the questionnaire? How would you characterize the health of first-time patients at this medical practice?

3. A survey asked chief executives at leading U.S. firms the following question: "Where do you expect the U.S. economy to be 12 months from now?" A representative sample of their responses appears below:

Same	Same	Same	Better	Worse
Same	Same	Better	Same	Worse
Same	Better	Same	Better	Same
Worse	Same	Same	Same	Worse
Same	Same	Same	Better	Same

   a. Construct frequency and relative frequency distributions that summarize the responses to the survey. Where did most chief executives expect the U.S. economy to be in 12 months?
   b. Construct a pie chart and a bar chart to summarize your results.

4. AccuWeather.com reported the following weather delays at these major U.S. airline hubs for July 21, 2010:

City	Delay	City	Delay
Atlanta	PM Delays	Mpls./St. Paul	None
Chicago	None	New York	All Day Delays
Dallas/Ft. Worth	None	Orlando	None
Denver	All Day Delays	Philadelphia	All Day Delays
Detroit	AM Delays	Phoenix	None
Houston	All Day Delays	Salt Lake City	None
Las Vegas	All Day Delays	San Francisco	AM Delays
Los Angeles	AM Delays	Seattle	None
Miami	AM Delays	Washington	All Day Delays

jag39519_ch02_018-059 25 09/18/18 06:53 AM

a. Construct frequency and relative frequency distributions that summarize the delays at major U.S. hubs. What was the most common type of delay? Explain.
b. Construct a pie chart and a bar chart to summarize your results.

5. Fifty pro-football rookies were rated on a scale of 1 to 5, based on performance at a training camp as well as on past performance. A ranking of 1 indicated a poor prospect whereas a ranking of 5 indicated an excellent prospect. The following frequency distribution was constructed.

Rating	Frequency
1	4
2	10
3	14
4	18
5	4

a. How many of the rookies received a rating of 4 or better? How many of the rookies received a rating of 2 or worse?
b. Construct the corresponding relative frequency distribution. What percent received a rating of 5?
c. Construct a bar chart for these data.

6. A recent survey asked 5,324 individuals: "What's most important to you when choosing where to live?" The responses are shown in the following relative frequency distribution.

Response	Relative Frequency
Good jobs	0.37
Affordable homes	0.15
Top schools	0.11
Low crime	0.23
Things to do	0.14

Source: Turner, Inc.

a. Construct the corresponding frequency distribution. How many of the respondents chose "low crime" as the most important criterion when choosing where to live?
b. Construct a bar chart for the frequency distribution found in part a.

7. What is the perfect summer trip? A National Geographic Kids survey (*AAA Horizons,* April 2007) asked this question to 316 children ages 8 to 14. Their responses are given in the following frequency distribution.

Top Vacation Choice	Frequency
Cruises	140
Beaches	68
Amusement Parks	68
Big Cities	20
Lakes	12
Summer Camp	8

a. Construct the relative frequency distribution. What percentage of the responses cited "Cruises" as the perfect summer trip?
b. Construct a bar chart for these data.

8. The following table lists U.S. revenue (in $ billions) of the major car-rental companies.

Car-Rental Company	Revenue in 2009
Enterprise	10.7
Hertz	4.7
Avis Budget	4.0
Dollar Thrifty	1.5
Other	1.0

Source: *The Wall Street Journal*, July 30, 2010.

a. Compute the relative market share of the car-rental companies.
b. Hertz accounted for what percentage of sales?
c. Construct a pie chart for these data.

9. In a CBS News survey, 829 respondents were provided with a list of major events and asked which event would happen first. The following percent frequency distribution was constructed.

Event	Percent Frequency
Cure for cancer found	40
End of dependence on oil	27
Signs of life in outer space	12
Peace in Middle East	8
Other	6
None will happen	7

Source: *Vanity Fair,* December 2009.

a. Construct a pie chart and a bar chart for these data.
b. How many people think that a cure for cancer will be found first?

10. A 2010 poll conducted by NBC asked respondents who would win Super Bowl XLV in 2011. The responses by 20,825 people are summarized in the following table.

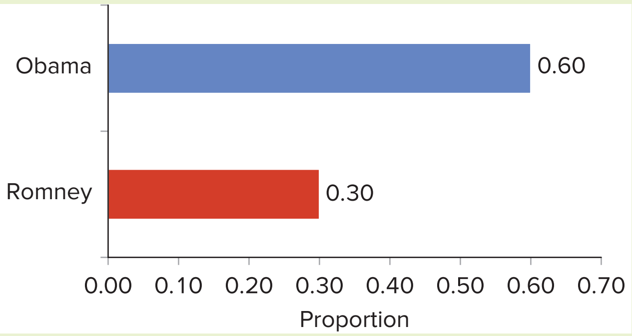
Team	Number of Votes
Atlanta Falcons	4,040
New Orleans Saints	1,880
Houston Texans	1,791
Dallas Cowboys	1,631
Minnesota Vikings	1,438
Indianapolis Colts	1,149
Pittsburgh Steelers	1,141
New England Patriots	1,095
Green Bay Packers	1,076
Others	

a. How many responses were for "Others"?
b. The Green Bay Packers won Super Bowl XLV, defeating the Pittsburgh Steelers by the score of 31–25. What

proportion of respondents felt that the Green Bay Packers would win?

c. Construct a bar chart for these data using relative frequencies.

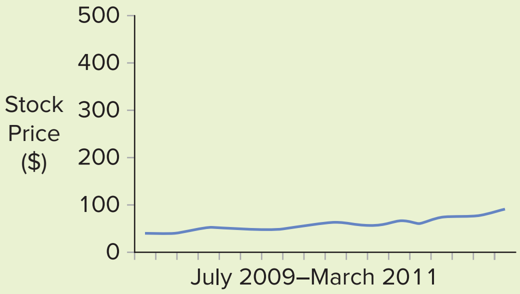
11. In a USA TODAY/Gallup Poll, respondents favored Barack Obama over Mitt Romney in terms of likeability, 60% to 30% (*Los Angeles Times,* July 28, 2012). The following bar chart summarizes the responses.



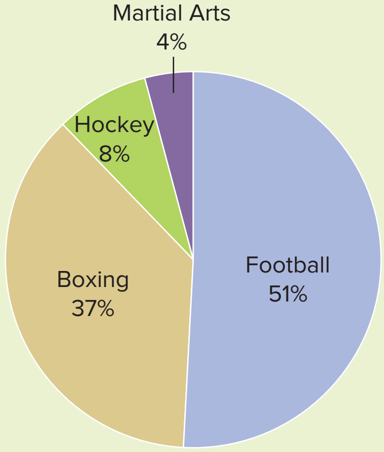
a. What percentage of respondents favored neither Obama nor Romney in terms of likeability?

b. Suppose this survey was based on 500 respondents. How many respondents favored Obama over Romney?

12. The accompanying figure plots the monthly stock price of Caterpillar, Inc., from July 2009 through March 2011. The stock has experienced tremendous growth over this time period, almost tripling in price. Does the figure reflect this growth? If not, why not?



13. A recent survey of 992 people asked: In which professional sport—football, boxing, hockey, or martial arts—is an athlete most likely to sustain an injury that will affect the athlete after he or she retires? (*Vanity Fair,* January 29, 2012.) The following pie chart summarizes the responses.

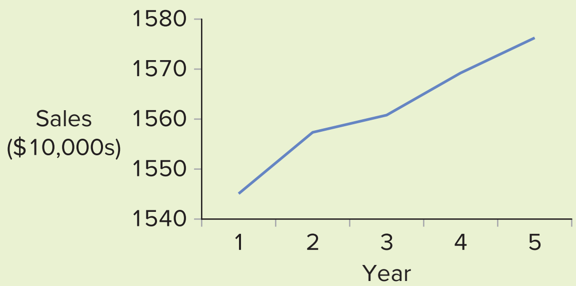


Source: Conde Nast.

a. According to this survey, in which sport was an athlete most likely to sustain an injury with lifelong consequences? In which sport was an athlete least likely to sustain an injury with lifelong consequences?

b. How many respondents believed that professional hockey players were most likely to sustain an injury with lifelong consequences?

14. Annual sales at a small pharmaceutical firm have been rather stagnant over the most recent five-year period, exhibiting only 1.2% growth over this time frame. A research analyst prepares the accompanying graph for inclusion in a sales report.



Does this graph accurately reflect what has happened to sales over the last five years? If not, why not?

2.2 SUMMARIZING QUANTITATIVE DATA

LO 2.3

Summarize quantitative data by constructing a frequency distribution.

With quantitative data, each observation represents a meaningful amount or count. The number of patents held by pharmaceutical firms (count) and household incomes (amount) are examples of quantitative data. Although different in nature from qualitative data, we still use a **frequency distribution** to summarize quantitative data.

Before discussing the mechanics of constructing a frequency distribution, we find it useful to first examine one in its final form, using the house-price data from the introductory case to this chapter. We take the raw data (the observations) from Table 2.1 and construct a frequency distribution with five intervals or **classes**, each of width 100, as shown

jag39519_ch02_018-059 27 09/18/18 06:53 AM

in Table 2.6. We see, for instance, that four houses sold in the first class, where prices ranged from $300,000 up to $400,000. The data are more manageable using a frequency distribution, but some detail is lost because we no longer see the observations.

TABLE 2.6 Frequency Distribution for House-Price Data

Class (in $1,000s)	Frequency
300 up to 400	4
400 up to 500	11
500 up to 600	14
600 up to 700	5
700 up to 800	2
	Total = 36

EXAMPLE 2.2

Based on the frequency distribution in Table 2.6, what is the price range over this time period? What price range exhibited the highest frequency?

SOLUTION: The frequency distribution shows that house prices ranged from $300,000 up to $800,000 over this time period. The most houses (14) sold in the $500,000 up to $600,000 range. Note that only four houses sold in the lowest price range and only two houses sold at the highest price range.

It turns out that reading and understanding a frequency distribution is actually easier than forming one. When we constructed a frequency distribution with qualitative data, the raw data could be categorized in a well-defined way. With quantitative data, we must make certain decisions about the number of classes, as well as the width of each class. We do not apply concrete rules when we define the classes in Table 2.6; however, we are able to follow several guidelines.

TABLE 2.7 Too Many Classes in a Distribution

Class (in $1,000s)	
325 up to 350	2
350 up to 375	1
375 up to 400	1
400 up to 425	3
425 up to 450	5
450 up to 475	3
475 up to 500	0
500 up to 525	5
525 up to 550	5
550 up to 575	3
575 up to 600	1
600 up to 625	0
625 up to 650	1
650 up to 675	4
675 up to 700	0
700 up to 725	1
725 up to 750	1
	Total = 36

Guidelines for Constructing a Frequency Distribution

- *Classes are mutually exclusive.* In other words, classes do not overlap. Each observation falls into one, and only one, class. For instance, suppose a value of 400 appeared in Table 2.1. Given the class divisions in Table 2.6, we would have included this observation in the second class interval. Mathematically, the second class interval is expressed as $400 \leq \text{Price} < 500$. Alternatively, we can define the second interval as $400 < \text{Price} \leq 500$, in which case the value 400 is included in the previous class interval. In short, no matter the specification of the classes, the observation is included in only one of the classes.
- *Classes are exhaustive.* The total number of classes covers the entire sample (or population). In Table 2.6, if we had left off the last class, 700 up to 800, then we would be omitting two observations from the sample.
- *The total number of classes in a frequency distribution usually ranges from 5 to 20.* Smaller data sets tend to have fewer classes than larger data sets. Recall that the goal of constructing a frequency distribution is to summarize the data in a form that accurately depicts the group as a whole. If we have too many classes, then this advantage of the frequency distribution is lost. For instance, suppose we create a frequency distribution for the house-price data with 17 classes, each of width 25, as shown in Table 2.7. Technically, this is a valid frequency distribution, but the summarization advantage of the frequency distribution is lost because there are too

jag39519_ch02_018-059 28 09/18/18 06:53 AM

many class intervals. Similarly, if the frequency distribution has too few classes, then considerable accuracy and detail are lost. Consider a frequency distribution of the house-price data with three classes, each of width 150, as shown in Table 2.8.

TABLE 2.8 Too Few Classes in a Distribution

Class (in $1,000s)	Frequency
300 up to 450	12
450 up to 600	17
600 up to 750	7
	Total = 36

Again, this is a valid frequency distribution. However, we cannot tell whether the 17 houses that sold for $450,000 up to $600,000 fall closer to the price of $450,000, fall closer to the price of $600,000, or are evenly spread within the interval. With only three classes in the frequency distribution, too much detail is lost.

- Once we choose the number of classes for a raw data set, we can then *approximate the width of each class* by using the formula

$$\frac{\text{Largest value} - \text{Smallest value}}{\text{Number of classes}}.$$

Generally, the width of each class is the same for each class interval. If the class width varied, comparisons between the numbers of observations in different intervals would be misleading.

- It is preferable to *define class limits that are easy to recognize and interpret.* Suppose we conclude, as we do in Table 2.6, that we should have five classes in the frequency distribution for the house-price data. Applying the class-width formula with the largest value of 735 and the smallest value of 330 (from Table 2.1) yields $\frac{735 - 330}{5} = 81$. Table 2.9 shows the frequency distribution with five classes and a class width of 81.

TABLE 2.9 Cumbersome Class Width in a Distribution

Class (in $1,000s)	Frequency
330 up to 411	4
411 up to 492	11
492 up to 573	12
573 up to 654	3
654 up to 735	6
	Total = 36

Again, this is a valid frequency distribution, but it proves unwieldy. Recall that one major goal in forming a frequency distribution is to provide more clarity in interpreting the data. Grouping the data in this manner actually makes analyzing the data more difficult. In order to facilitate interpretation of the frequency distribution, it is best to define class limits with ease of recognition in mind. To this end, and as initially shown in Table 2.6, we set the lower limit of the first class at 300 (rather than 330) and obtain the remaining class limits by successively adding 100 (rather than 81).

Once we have clearly defined the classes for a particular data set, the next step is to count and record the number of data points that fall into each class. As we did with the construction of a qualitative frequency distribution, we usually include a tally column to aid in counting (see Table 2.10), but then we remove this column in the final presentation of the frequency distribution. For instance, in Table 2.1, the first observation, 430, falls in the second class, so we place a tally mark in the second class; the next observation, 520, falls in the third class, so we place a tally mark in the third class, and so on. The frequency column shows the numerical value of the respective tally count. Since four tally marks appear in the first class, we record the value 4 as its frequency—the number of observations that fall within the first class. One way to ensure that we have included all the observations in the frequency distribution is to sum the frequency column. This sum should always equal the population or sample size.

TABLE 2.10 Constructing Frequency Distributions for the House-Price Data

Class (in $1,000s)	Tally	Frequency	Cumulative Frequency
300 up to 400	IIII	4	4
400 up to 500	卌 卌 I	11	4 + 11 = 15
500 up to 600	卌 卌 IIII	14	4 + 11 + 14 = 29
600 up to 700	卌	5	4 + 11 + 14 + 5 = 34
700 up to 800	II	2	4 + 11 + 14 + 5 + 2 = 36
		Total = 36	

A frequency distribution indicates how many observations fall within some range. However, we might want to know how many observations fall below the upper limit of a particular class. In these cases, our needs are better served with a **cumulative frequency distribution**.

The last column of Table 2.10 shows values for cumulative frequency. The cumulative frequency of the first class is the same as the frequency of the first class—that is, the value 4. However, the interpretation is different. With respect to the frequency column, the value 4 tells us that four of the houses sold in the $300,000 up to $400,000 range. For the cumulative frequency column, the value 4 tells us that four of the houses sold for less than $400,000. To obtain the cumulative frequency for the second class we add its frequency, 11, with the preceding frequency, 4, and obtain 15. This tells us that 15 of the houses sold for less than $500,000. We solve for the cumulative frequencies of the remaining classes in a like manner. Note that the cumulative frequency of the last class is equal to the sample size of 36. This indicates that all 36 houses sold for less than $800,000.

FREQUENCY AND CUMULATIVE FREQUENCY DISTRIBUTIONS FOR QUANTITATIVE DATA

For quantitative data, a frequency distribution groups data into intervals called classes and records the number of observations that falls within each class. A cumulative frequency distribution records the number of observations that fall below the upper limit of each class.

EXAMPLE 2.3

Using Table 2.10, how many of the houses sold in the $500,000 up to $600,000 range? How many of the houses sold for less than $600,000?

SOLUTION: From the frequency distribution, we find that 14 houses sold in the $500,000 up to $600,000 range. In order to find the number of houses that sold for less than $600,000, we use the cumulative frequency distribution. We readily observe that 29 of the houses sold for less than $600,000.

Suppose we want to compare house prices in Mission Viejo, California, to house prices in another region of the United States. Just as for qualitative data, when making comparisons between two quantitative data sets—especially if the data sets differ in size—a **relative frequency distribution** tends to provide more meaningful information than a frequency distribution.

The second column of Table 2.11 shows the construction of a relative frequency distribution from the frequency distribution in Table 2.10. We take each class's frequency and divide by the total number of observations. For instance, we observed four houses that sold in the lowest range of $300,000 up to $400,000. We take the class frequency of 4 and divide by the sample size, 36, and obtain 0.11. Equivalently, we can say 11% of the houses sold in this price range. We make similar calculations for each class and note that when we sum the column of relative frequencies, we should get a value of one (or, due to rounding, a number very close to one).

TABLE 2.11 Constructing Relative Frequency Distributions for House-Price Data

Class (in $1,000s)	Relative Frequency	Cumulative Relative Frequency
300 up to 400	4/36 = 0.11	0.11
400 up to 500	11/36 = 0.31	0.11 + 0.31 = 0.42
500 up to 600	14/36 = 0.39	0.11 + 0.31 + 0.39 = 0.81
600 up to 700	5/36 = 0.14	0.11 + 0.31 + 0.39 + 0.14 = 0.95
700 up to 800	2/36 = 0.06	0.11 + 0.31 + 0.39 + 0.17 + 0.06 ≈ 1
	Total = 1 (subject to rounding)	

The last column of Table 2.11 shows the **cumulative relative frequency distribution**. The cumulative relative frequency for a particular class indicates the proportion (fraction) of the observations that falls below the upper limit of that particular class. We can calculate the cumulative relative frequency of each class in one of two ways: (1) We can sum successive relative frequencies, or (2) we can divide each class's cumulative frequency by the sample size. In Table 2.11 we show the first way. The value for the first class is the same as the value for its relative frequency—that is, 0.11. For the second class, we add 0.31 to 0.11 and obtain 0.42; this value indicates that 42% of the house prices were less than $500,000. We continue calculating cumulative relative frequencies in this manner until we reach the last class. Here, we get the value one, which means that 100% of the houses sold for less than $800,000.

RELATIVE AND CUMULATIVE RELATIVE FREQUENCY DISTRIBUTIONS

For quantitative data, a relative frequency distribution identifies the proportion (or the fraction) of observations that falls within each class—that is,

$$\text{Class relative frequency} = \frac{\text{Class frequency}}{\text{Total number of observations}}.$$

A cumulative relative frequency distribution records the proportion (or the fraction) of observations that fall below the upper limit of each class.

EXAMPLE 2.4

Using Table 2.11, what percent of the houses sold for at least $500,000 but not more than $600,000? What percent of the houses sold for less than $600,000? What percent of the houses sold for $600,000 or more?

SOLUTION: The relative frequency distribution indicates that 39% of the houses sold for at least $500,000 but not more than $600,000. Further, the cumulative relative frequency distribution indicates that 81% of the houses sold for less than $600,000. This result implies that 19% sold for $600,000 or more.

SYNOPSIS OF INTRODUCTORY CASE

©Brand X Pictures/Stockbyte/Getty Images

During June 2010, Matthew Edwards reviewed the selling prices of 36 house sales in Mission Viejo, California, for a client from Seattle, Washington. After constructing various frequency distributions, he is able to make the following summary conclusions. House prices ranged from $300,000 up to $800,000 over this time period. Most of the houses (14) sold in the $500,000 up to $600,000 range, which is, more or less, the client's price range. Twenty-nine of the houses sold for less than $600,000. Converting the data into percentages so the client can make comparisons with home sales in the Seattle area, Matthew found that 39% of the houses sold for $500,000 up to $600,000. Moreover, 81% of the houses sold for less than $600,000, which implies that 19% sold for $600,000 or more.

LO 2.4

Construct and interpret a histogram, a polygon, and an ogive.

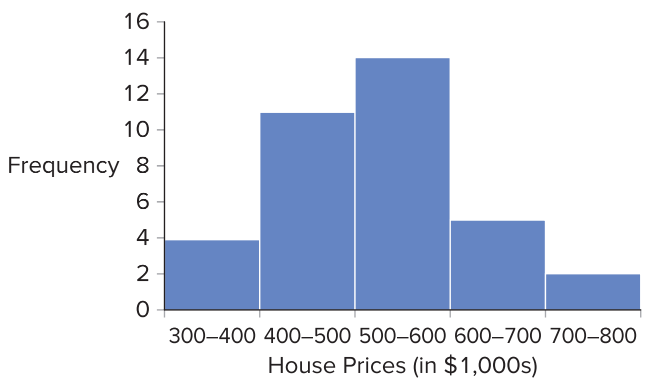
Histograms, Polygons, and Ogives

A **histogram** and a **polygon** are graphical depictions of frequency and relative frequency distributions. The advantage of a visual display is that we can quickly see where most of the observations tend to cluster, as well as the spread and shape of the data. For instance, a histogram and a polygon may reveal whether or not the distribution has a symmetric shape.

> **GRAPHICAL DISPLAY OF QUANTITATIVE DATA: A HISTOGRAM**
>
> A histogram is a series of rectangles where the width and height of each rectangle represent the class width and frequency (or relative frequency) of the respective class.

For quantitative data, a histogram is essentially the counterpart to the vertical bar chart we use for qualitative data. When constructing a histogram, we mark off the class limits along the horizontal axis. The height of each bar represents either the frequency or the relative frequency for each class. No gaps appear between the interval limits. Figure 2.5 shows a histogram for the frequency distribution of house prices shown in Table 2.6. A casual inspection of the histogram reveals that the selling price of houses in this sample ranged from $300,000 to $800,000; however, the most common house price fell in the $500,000 to $600,000 range.

FIGURE 2.5
Frequency histogram for house prices



The only difference between a frequency histogram and a relative frequency histogram is the unit of measurement on the vertical axis. For the frequency histogram, we use the frequency of each class to represent the height; for the relative frequency histogram we use the proportion (or the fraction) of each class to represent the height. In a relative frequency histogram, the area of any rectangle is proportional to the relative frequency of observations falling into that class. Figure 2.6 shows the relative frequency histogram for house prices.

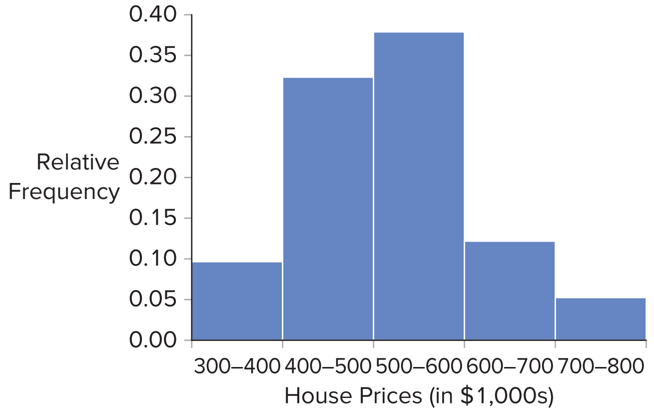
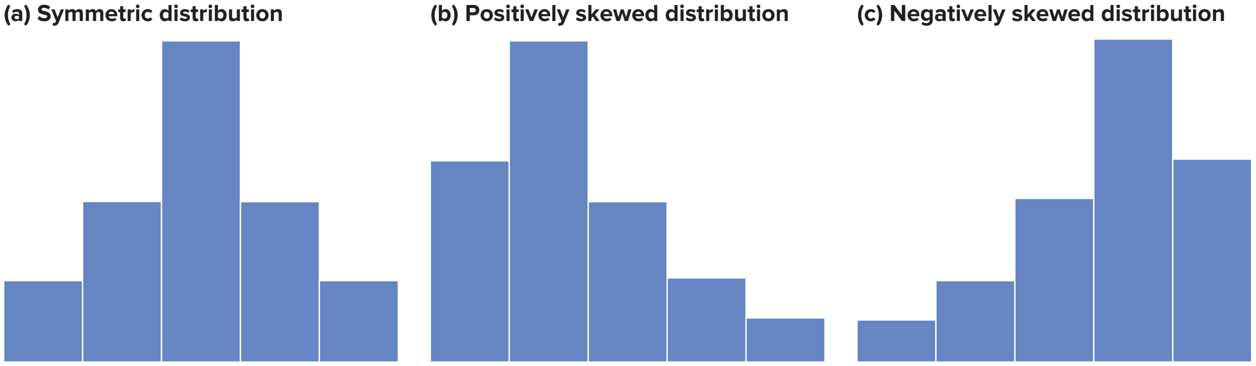


FIGURE 2.6 Relative frequency histogram for house prices

In general, the shape of most data distributions can be categorized as either symmetric or skewed. A symmetric distribution is one that is a mirror image of itself on both sides of its center. That is, the location of values below the center correspond to those above the center. As we will see in later chapters, the smoothed histogram for many data sets approximates a bell-shaped curve, which is indicative of the well-known normal distribution. If the distribution is not symmetric, then it is either positively skewed or negatively skewed.

FIGURE 2.7 Histograms with differing shapes



The histogram in Figure 2.7(a) shows a symmetric distribution. If the edges were smoothed, this histogram would look somewhat bell-shaped. In Figure 2.7(b), the histogram shows a positively skewed, or skewed to the right, distribution with a long tail extending to the right. This attribute reflects the presence of a small number of relatively large values. Finally, the histogram in Figure 2.7(c) indicates a negatively skewed, or skewed to the left, distribution since it has a long tail extending off to the left. Data that follow a negatively skewed distribution have a small number of relatively small values.

Though not nearly as skewed as the data exhibited in Figure 2.7(b), the house-price data in Figure 2.6 exhibit slight positive skew. This is the result of a few, relatively expensive homes in the city. It is common for distributions of house prices and incomes to exhibit positive skewness.

A polygon provides another convenient way of depicting a frequency distribution. It too gives a general idea of the shape of a distribution. Like the histogram, we place either the frequency or the relative frequency of the distribution on the *y*-axis, and the upper and lower limits of each class on the *x*-axis. We plot the midpoint of each class with its corresponding frequency or relative frequency. We then connect neighboring points with a straight line.

GRAPHICAL DISPLAY OF QUANTITATIVE DATA: A POLYGON

A polygon connects a series of neighboring points where each point represents the midpoint of a particular class and its associated frequency or relative frequency.

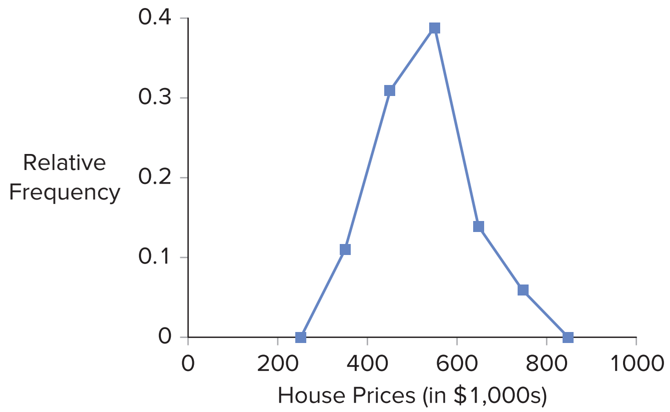
If we choose to construct a polygon for the house-price data, we first calculate the midpoint of each interval; thus, the midpoint for the first interval is $\frac{300 + 400}{2} = 350$, and similarly, the midpoints for the remaining intervals are 450, 550, 650, and 750. We treat each midpoint as the *x*-coordinate and the respective frequency (or relative frequency) as the *y*-coordinate. After plotting the points, we connect neighboring points. In order to close off the graph at each end, we add one interval below the lowest interval (so, 200 up to 300 with midpoint 250) and one interval above the highest interval (so, 800 up to 900 with midpoint 850) and assign each of these classes zero frequencies. Table 2.12 shows the relevant coordinates for plotting a polygon using the house-price data. Here we use relative frequency to represent the *y*-coordinate.

TABLE 2.12 Coordinates for Plotting Relative Frequency Polygon

Classes	x-coordinate (midpoint)	y-coordinate (relative frequency)
(Lower end)	250	0
300–400	350	0.11
400–500	450	0.31
500–600	550	0.39
600–700	650	0.14
700–800	750	0.06
(Upper end)	850	0

Figure 2.8 plots a relative frequency polygon for the house-price data. The distribution appears to approximate the bell-shaped distribution discussed earlier. Only a careful inspection of the right tail suggests that the data are slightly positively skewed.

FIGURE 2.8 Polygon for the house-price data



In many instances, we might want to convey information by plotting an **ogive** (pronounced “ojive”).

> GRAPHICAL DISPLAY OF QUANTITATIVE DATA: AN OGIVE
>
> An ogive connects a series of neighboring points where each point represents the upper limit of a particular class and its associated cumulative frequency or cumulative relative frequency.

An ogive differs from a polygon in that we use the upper limit of each class as the *x*-coordinate and the cumulative frequency or cumulative relative frequency of the corresponding class as the *y*-coordinate. After plotting the points, we connect neighboring points. Lastly, we close the ogive only at the lower end by intersecting the *x*-axis at the lower limit of the first class. Table 2.13 shows the relevant coordinates for plotting an ogive using the house-price data. Here we use the cumulative relative frequency as the *y*-coordinate since the resulting graph tends to have more interpretive appeal. The use of cumulative frequency would not change the shape of the ogive, just the unit of measurement on the *y*-axis.

TABLE 2.13 Coordinates for the Ogive for the House-Price Data

Classes	x-coordinate (upper limit)	y-coordinate (cumulative relative frequency)
(Lower end)	300	0
300–400	400	0.11
400–500	500	0.42
500–600	600	0.81
600–700	700	0.95
700–800	800	1

Figure 2.9 plots the ogive for the house-price data. In general, we can use an ogive to approximate the proportion of values that are less than a specified value on the horizontal axis. Consider an application to the house-price data in Example 2.5.

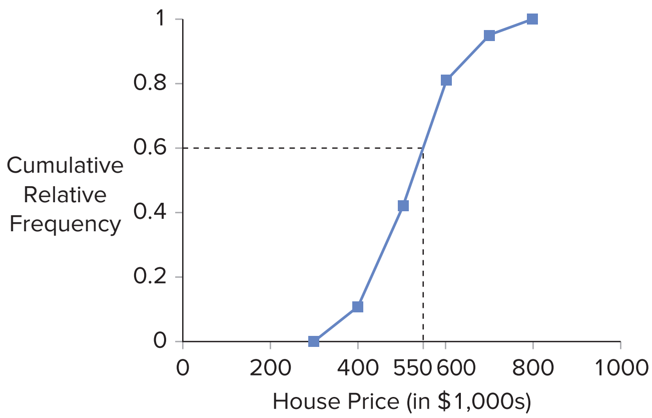


FIGURE 2.9 Ogive for the house-price data

EXAMPLE 2.5

Using Figure 2.9, approximate the percentage of houses that sold for less than $550,000.

SOLUTION: We first draw a vertical line that starts at 550 and intersects the ogive. Then we follow the line to the vertical axis and read the value. We can conclude that approximately 60% of the houses sold for less than $550,000.

Using Excel to Construct a Histogram, a Polygon, and an Ogive

A Histogram Constructed from Raw Data

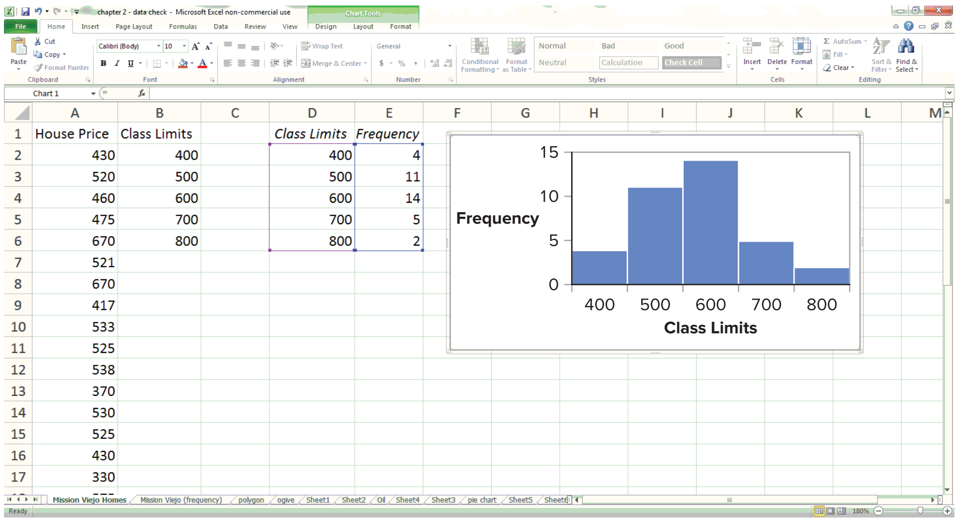
In general, Excel offers two different ways to construct a histogram, depending on whether we have access to the raw data or the frequency distribution. We first replicate Figure 2.5 using the house-price data from Table 2.1 where the data are in raw form. We then replicate Figure 2.5 using the house-price data from Table 2.6 where the data have been converted to a frequency distribution.

FILE
MV_Houses

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

B. See Figure 2.10. In a column next to the data, enter the values of the upper limits for each class, or in this example, 400, 500, 600, 700, and 800; label this column as Class Limits. The reason for these entries is explained in step C. From the menu choose **Data > Data Analysis > Histogram > OK**. (*Note:* If you do not see the **Data Analysis** option under **Data**, you must *add in* this option. From the menu choose **File > Options > Add-Ins** and choose **Go** at the bottom of the dialog box. Select the box to the left of **Analysis Toolpak**, and then click **OK**. If you have installed this option properly, you should now see **Data Analysis** under **Data**.)

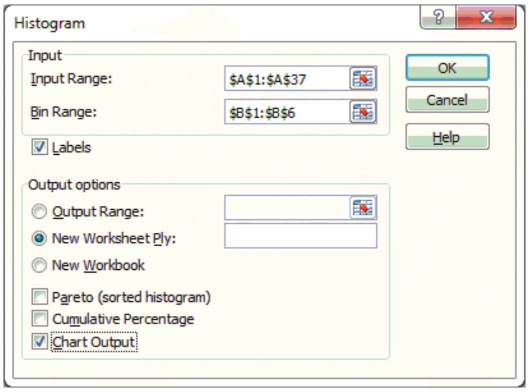
FIGURE 2.10 Constructing a histogram from raw data with Excel



Source: Microsoft Excel

C. See Figure 2.11. In the *Histogram* dialog box, under *Input Range*, select the House Price data. Excel uses the term "bins" for the class limits. If we leave the *Bin Range* box empty, Excel creates evenly distributed intervals using the minimum and maximum values of the input range as end points. This methodology is rarely satisfactory. In order to construct a histogram that is more informative, we use the upper limit of each class as the bin values. Under *Bin Range*, we select the Class Limits data. (Check the *Labels* box if you have included the names House Price and Class Limits as part of the selection.) Under *Output Options,* we choose **Chart Output**, then click **OK**.

FIGURE 2.11 Excel's dialog box for a histogram



Source: Microsoft Excel

D. Since Excel leaves spaces between the rectangles, we right-click on any of the rectangles, choose **Format Data Series** and change the *Gap Width* to 0, then choose **Close**. In the event that the given class limits do not include all the data points, Excel automatically adds another interval labeled "More" to the resulting frequency distribution and histogram. Since we observe zero observations in this interval for this example, we delete this interval for expositional purposes. Excel also defines its classes by excluding the value of the lower limit and including the value of the upper class limit for each interval. For example, if the value 400 appeared in the house-price data, Excel would have accounted for this observation in the first class. If any upper-limit value appeared in the house-price data, we would have adjusted the class limits in the *Bin Range* to 399, 499, and so on, so that Excel's frequency distribution and histogram would be consistent with those that we constructed in Table 2.10 and Figure 2.5. Formatting (regarding axis titles, gridlines, etc.) can be done by selecting **Format > Add Chart Element** from the menu.

A Histogram Constructed from a Frequency Distribution

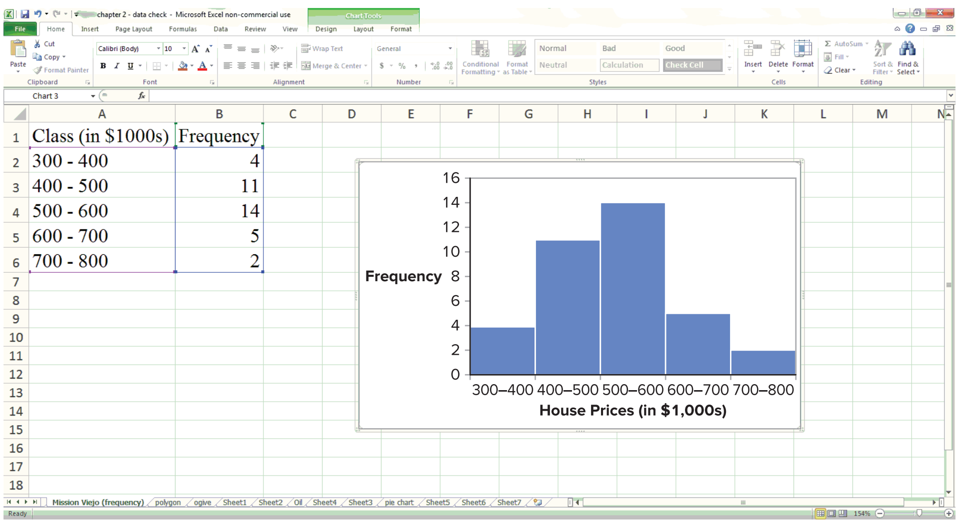
Suppose we do not have the raw data for house prices, but we have the frequency distribution reported in Table 2.6.

A. Open ***MV_Frequency*** (Table 2.6).

B. See Figure 2.12. First, select the classes and respective frequencies. Then, from the menu, choose **Insert > Column > 2-D Column**. From the options, choose the

FILE
MV_Frequency

FIGURE 2.12 Constructing a histogram from a frequency distribution with Excel



Source: Microsoft Excel

graph on the top left. (If you are having trouble finding this option after selecting **Insert**, look for the vertical bars above **Charts**.)

C. In order to remove the spaces between the rectangles, right-click on any of the rectangles, choose **Format Data Series** and change the *Gap Width* to 0, then choose **Close**.

D. Formatting (regarding axis titles, gridlines, etc.) can be done by selecting **Format > Add Chart Element** from the menu.

A Polygon

We replicate the polygon in Figure 2.8.

FILE *Polygon*

A. Open ***Polygon*** (this is a simplified version of the data in Table 2.12).

B. Select the values in the *x* and *y* columns, and choose **Insert > Scatter (X, Y) or Bubble Chart**. From the options given, select the box at the middle right. (If you are having trouble finding this option after selecting **Insert**, look for the graph with data points above **Charts**.)

C. Formatting (regarding axis titles, gridlines, etc.) can be done by selecting **Format > Add Chart Element** from the menu.

An Ogive

We replicate the polygon in Figure 2.9.

FILE *Ogive*

A. Open ***Ogive*** (this is a simplified version of the data in Table 2.13).

B. Select the values in the *x* and *y* columns, and choose **Insert > Scatter (X, Y) or Bubble Chart**. From the options given, select the box at the middle right. (If you are having trouble finding this option after selecting **Insert**, look for the graph with data points above **Charts**.)

C. Formatting (regarding axis titles, gridlines, etc.) can be done by selecting **Format > Add Chart Element** from the menu.

EXERCISES 2.2

Mechanics

15. Consider the following data set:

4.3	10.9	8.7	7.6	6.5	10.9	11.1	14.3	13.2	14.3
3.2	9.8	8.7	5.4	7.6	6.5	10.9	3.2	11.1	11.1
8.7	8.7	4.3	5.4	5.4	12.1	12.1	3.2	8.7	8.7

a. Construct the frequency distribution using classes of 3 up to 5, 5 up to 7, etc.
b. Construct the relative frequency, the cumulative frequency, and the cumulative relative frequency distributions.
c. How many of the observations are at least 7 but less than 9? How many of the observations are less than 9?
d. What percentage of the observations are at least 7 but less than 9? What percentage of the observations are less than 9?
e. Graph the relative frequency histogram.
f. Graph the ogive.

16. Consider the following data set:

4.3	10.9	8.7	7.6	6.5	10.9	11.1	14.3	13.2	14.3
3.2	9.8	8.7	5.4	7.6	6.5	10.9	3.2	11.1	11.1
8.7	8.7	4.3	5.4	5.4	12.1	12.1	3.2	8.7	8.7
10.9	−8.8	28.7	14.3	−4.8	9.8	11.1	5.4	8.7	−2.8
33.2	−3.9	2.1	3.2	22.1	25.4	5.4	29.8	26.5	0.1
−7.6	−4.8	0.1	15.4	−3.8	35.4	21.1	15.4	19.8	23.2
4.3	6.5	−1.9	12.1	24.3	36.5	15.4	3.2	−4.8	2.1

a. Construct the frequency distribution using classes of −10 up to 0, 0 up to 10, etc. How many of the observations are at least 10 but less than 20?
b. Construct the relative frequency distribution and the cumulative relative frequency distribution. What percent of the observations are at least 10 but less than 20? What percent of the observations are less than 20?
c. Graph the relative frequency polygon. Is the distribution symmetric? If not, then how is it skewed?

17. Consider the following frequency distribution:

Class	Frequency
10 up to 20	12
20 up to 30	15
30 up to 40	25
40 up to 50	4

a. Construct the relative frequency distribution. Graph the relative frequency histogram.
b. Construct the cumulative frequency distribution and the cumulative relative frequency distribution.
c. What percent of the observations are at least 30 but less than 40? What percent of the observations are less than 40?

18. Consider the following frequency distribution:

Class	Frequency
1,000 up to 1,100	2
1,100 up to 1,200	7
1,200 up to 1,300	3
1,300 up to 1,400	4

a. Construct the relative frequency distribution. What percent of the observations are at least 1,100 but less than 1,200?
b. Construct the cumulative frequency distribution and the cumulative relative frequency distribution. How many of the observations are less than 1,300?
c. Graph the frequency histogram.

19. Consider the following cumulative frequency distribution:

Class	Cumulative Frequency
15 up to 25	30
25 up to 35	50
35 up to 45	120
45 up to 55	130

a. Construct the frequency distribution. How many observations are at least 35 but less than 45?
b. Graph the frequency histogram.
c. What percent of the observations are less than 45?

20. Consider the following relative frequency distribution:

Class	Relative Frequency
–20 up to –10	0.04
–10 up to 0	0.28
0 up to 10	0.26
10 up to 20	0.22
20 up to 30	0.20

a. Suppose this relative frequency distribution is based on a sample of 50 observations. Construct the frequency distribution. How many of the observations are at least –10 but less than 0?
b. Construct the cumulative frequency distribution. How many of the observations are less than 20?
c. Graph the relative frequency polygon.

21. Consider the following cumulative relative frequency distribution.

Class	Cumulative Relative Frequency
150 up to 200	0.10
200 up to 250	0.35
250 up to 300	0.70
300 up to 350	1

a. Construct the relative frequency distribution. What percent of the observations are at least 250 but less than 300?
b. Graph the ogive.

Applications

22. *Kiplinger's* (August 2007) lists the assets (in billions of $) for the 20 largest stock mutual funds (ranked by size) as follows:

99.8	49.7	86.3	109.2	56.9
88.2	44.1	58.8	176.7	49.9
61.4	128.8	53.6	95.2	92.5
55.0	96.5	45.3	73.0	70.9

a. Construct the frequency distribution using classes of 40 up to 70, 70 up to 100, etc.
b. Construct the relative frequency distribution, the cumulative frequency distribution, and the cumulative relative frequency distribution.
c. How many of the funds had assets of at least $100 but less than $130 (in billions)? How many of the funds had assets less than $160 (in billions)?
d. What percent of the funds had assets of at least $70 but less than $100 (in billions)? What percent of the funds had assets less than $130 (in billions)?
e. Construct the frequency histogram. Comment on the shape of the distribution.

23. The number of text messages sent by 25 13-year-olds over the past month are as follows:

630	516	892	643	627	510	937	909	654
817	760	715	605	975	888	912	952	701
744	793	852	504	562	670	685		

a. Construct the frequency distribution using classes of 500 up to 600, 600 up to 700, etc.
b. Construct the relative frequency distribution, the cumulative frequency distribution, and the cumulative relative frequency distribution.

c. How many of the 13-year-olds sent at least 600 but less than 700 text messages? How many sent less than 800 text messages?

d. What percent of the 13-year-olds sent at least 500 but less than 600 text messages? What percent of the 13-year-olds sent less than 700 text messages?

e. Graph the relative frequency polygon. Comment on the shape of the distribution.

24. AccuWeather.com listed the following high temperatures (in degrees Fahrenheit) for 33 European cities on July 21, 2010.

75	92	81	85	91	73	94	95	81	64	85
62	84	85	81	86	91	79	74	92	91	95
88	87	81	73	76	86	92	83	75	92	83

a. Construct the frequency distribution using classes of 60 up to 70, 70 up to 80, etc.

b. Construct the relative frequency, the cumulative frequency, and the cumulative relative frequency distributions.

c. How many of the cities had high temperatures less than 80°?

d. What percent of the cities had high temperatures of at least 80° but less than 90°? What percent of the cities had high temperatures less than 90°?

e. Construct the relative frequency polygon. Comment on the shape of the distribution.

25. The following relative frequency distribution summarizes the ages of women who had a child in the last year.

Ages	Relative Frequency
15 up to 20	0.10
20 up to 25	0.25
25 up to 30	0.28
30 up to 35	0.24
35 up to 40	0.11
40 up to 45	0.02

Source: *The Statistical Abstract of the United States, 2010.*

a. Assume the relative frequency distribution is based on a sample of 2,000 women. Construct the frequency, the cumulative frequency, and the cumulative relative frequency distributions.

b. What percent of the women were at least 25 but less than 30 years old? What percent of the women were younger than 35 years old?

c. Graph the relative frequency polygon. Comment on the shape of the distribution.

d. Graph the ogive. Using the graph, approximate the age of the middle 50% of the distribution.

26. Fifty cities provided information on vacancy rates (in percent) in local apartments in the following frequency distribution.

Vacancy Rate	Frequency
0 up to 3	5
3 up to 6	10
6 up to 9	20
9 up to 12	10
12 up to 15	5

a. Construct the relative frequency distribution, the cumulative frequency distribution, and the cumulative relative frequency distribution.

b. How many of the cities had a vacancy rate less than 12%? What percent of the cities had a vacancy rate of at least 6% but less than 9%? What percent of the cities had a vacancy rate of less than 9%?

c. Graph the frequency histogram. Comment on the shape of the distribution.

27. The manager of a nightclub near a local university recorded the ages of the last 100 guests in the following cumulative frequency distribution.

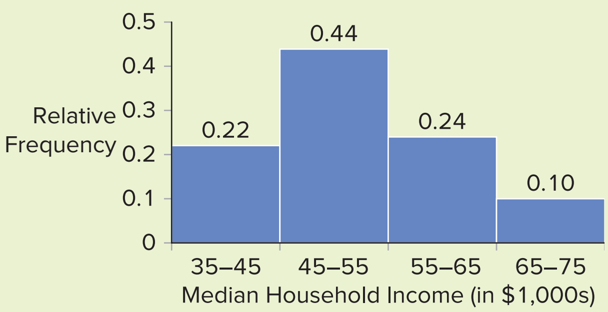
Ages	Cumulative Frequency
18 up to 22	45
22 up to 26	70
26 up to 30	85
30 up to 34	96
34 up to 38	100

a. Construct the frequency, the relative frequency, and the cumulative relative frequency distributions.

b. How many of the guests were at least 26 but less than 30 years old? What percent of the guests were at least 22 but less than 26 years old? What percent of the guests were younger than 34 years old? What percent were 34 years or older?

c. Graph the frequency histogram. Comment on the shape of the distribution.

28. The following relative frequency histogram summarizes the median household income for the 50 states in the United States (*U.S. Census*, 2010).

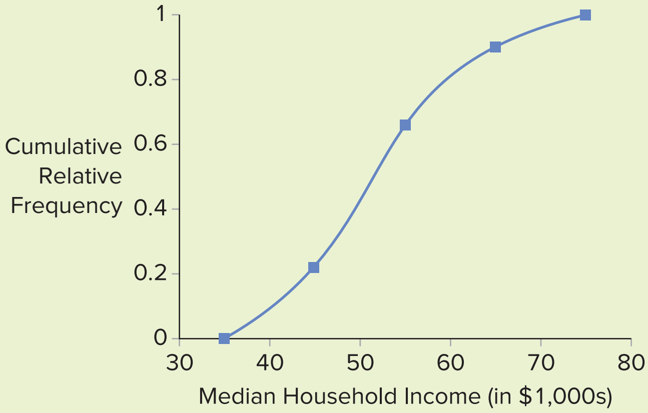


a. Is the distribution symmetric? If not, is it positively or negatively skewed?

b. What percentage of the states had median household income between $45,000 and $55,000?

c. What percentage of the states had median household income between $35,000 and $55,000?

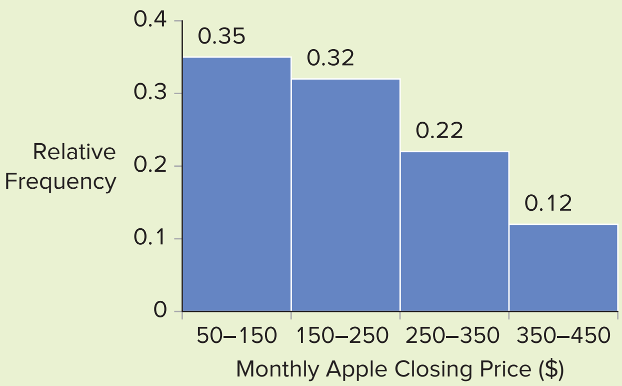
29. The following ogive summarizes the median household income for the 50 states in the United States (*U.S. Census,* 2010).



a. Approximate the percentage of states with median household income less than $50,000.

b. Approximate the percentage of states with median household income more than $60,000.

30. The following histogram summarizes Apple Inc.'s monthly stock price for the years 2007 through 2011 (http://finance.yahoo.com, data retrieved April 20, 2012).

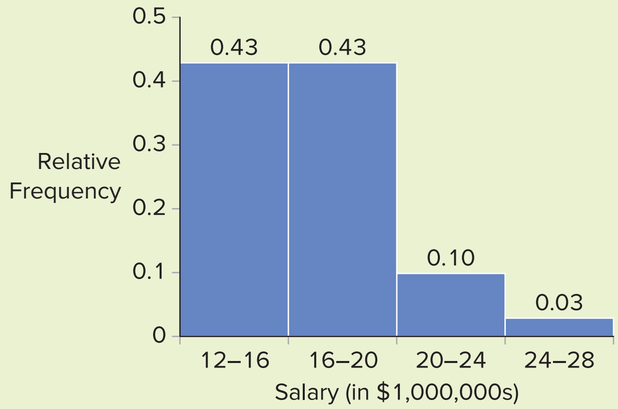


a. Is the distribution symmetric? If not, is it positively or negatively skewed?

b. Over this five-year period, approximate the minimum monthly stock price and the maximum monthly stock price.

c. Over this five-year period, which class had the highest relative frequency?

31. The following histogram summarizes the salaries (in $1,000,000s) for the 30 highest-paid players in the National Basketball Association (NBA) for the 2012 season (www.nba.com, data retrieved March 2012).



a. Is the distribution symmetric? If not, is it positively or negatively skewed?

b. How many NBA players earned between $20,000,000 and $24,000,000?

c. Approximately how many NBA players earned between $12,000,000 and $20,000,000?

32. The following ogive summarizes the salary (in $1,000,000s) for the 30 highest-paid players in the National Basketball Association (NBA) for the 2012 season (www.nba.com, data retrieved March 2012).

a. Approximate the percentage of salaries that were less than $18,000,000.

b. Approximate the number of salaries that were more than $14,000,000.

33. **FILE** ***Math_SAT.*** The following table lists a portion of the average math SAT scores for each state for the year 2009.

State	SAT
Alabama	552
Alaska	516
⋮	⋮
Wyoming	568

Source: www.collegeboard.com.

a. Construct the frequency distribution and graph the frequency histogram using classes of 450 to 500, 501 to 550, etc. Comment on the shape of the distribution. How many of the states had scores between 551 and 600?
b. Construct the relative frequency, the cumulative frequency, and the cumulative relative frequency distributions.
c. How many of the states had average math SAT scores of 550 or less?
d. What percent of the states had average math SAT scores between 551 and 600? What percent of the states had average math SAT scores of 550 or less?

34. **FILE** ***Census.*** The accompanying table shows a portion of median house values (in $) for the 50 states as reported by the U.S. Census Bureau in 2010.

State	House Value
Alabama	117600
Alaska	229100
⋮	⋮
Wyoming	174000

a. Construct the frequency distribution and graph the frequency histogram for the median house values. Use six classes with upper limits of $100,000, $200,000, etc.
b. Is the distribution symmetric? If not, is it positively or negatively skewed?
c. Which class interval had the highest frequency?
d. What percentage of the states had median house values between $300,000 and $400,000?
e. How many of the states had median house values less than $300,000?

35. **FILE** ***Gas_Prices.*** The accompanying table shows a portion of the average price (in $) for a gallon of gas for the 50 states during April 2012.

State	Price
Alabama	4.36
Alaska	3.79
⋮	⋮
Wyoming	3.63

Source: www.AAA.com, data retrieved April 16, 2012.

a. Construct the frequency distribution and graph the frequency histogram for the average gas price. Use six classes with upper limits of $3.70, $3.90, etc.
b. Is the distribution symmetric? If not, is it positively or negatively skewed?
c. Which class interval had the highest frequency?
d. Graph the ogive. Approximate the percentage of states that had an average gas price of $3.90 or less. Approximate the number of states that had an average gas price greater than $3.90.

36. **FILE** ***DJIA_2012.*** For the first three months of 2012, the stock market put up its best first-quarter performance in over a decade (Money.cnn.com, April 9, 2012). The accompanying table shows a portion of the daily price index for the Dow Jones Industrial Average (DJIA) over this period.

Day	DJIA
January 3, 2012	12397
January 4, 2012	12418
⋮	⋮
March 31, 2012	13212

Source: Finance.yahoo.com, data retrieved April 20, 2012.

a. Construct the frequency distribution and the frequency histogram for the DJIA price index. Use five classes with upper limits of 12,500, 12,750, etc. On how many days during this quarter was the DJIA less than 12,500?
b. Graph the relative frequency polygon. Is the distribution symmetric? If not, is it positively or negatively skewed?
c. Graph the ogive. Approximate the percentage of days that the DJIA was less than 13,000.

LO **2.5**

Construct and interpret a stem-and-leaf diagram.

2.3 STEM-AND-LEAF DIAGRAMS

John Tukey (1915–2000), a well-known statistician, provided another visual method for displaying quantitative data. A **stem-and-leaf diagram** is often a preliminary step when analyzing a data set. It is useful in that it gives an overall picture of where the data are centered and how the data are dispersed from the center.

GRAPHICAL DISPLAY OF QUANTITATIVE DATA: A STEM-AND-LEAF DIAGRAM

A stem-and-leaf diagram is constructed by separating each value of a data set into two parts: a *stem,* which consists of the leftmost digits, and a *leaf,* which consists of the last digit.

The best way to explain a stem-and-leaf diagram is to show an example.

EXAMPLE 2.6

Table 2.14 shows the ages of the 25 wealthiest people in the world in 2010. Construct and interpret a stem-and-leaf diagram.

TABLE 2.14 Wealthiest People in the World, 2010

Name	Age	Name	Age
Carlos Slim Helu	70	Li Ka-shing	81
William Gates III	54	Jim Walton	62
Warren Buffet	79	Alice Walton	60
Mukesh Ambani	52	Liliane Bettencourt	87
Lakshmi Mittal	59	S. Robson Walton	66
Lawrence Ellison	65	Prince Alwaleed Alsaud	54
Bernard Arnault	61	David Thomson	52
Eike Batista	53	Michael Otto	66
Amancio Ortega	74	Lee Shau Kee	82
Karl Albrecht	90	Michael Bloomberg	68
Ingvar Kamprad	83	Sergey Brin	36
Christy Walton	55	Charles Koch	74
Stefan Persson	62		

Reprinted by permission of Forbes Media LLC © 2011.

FILE
Wealthiest_People

SOLUTION: For each age, we first decide that the number in the tens spot will denote the stem, thus leaving the number in the ones spot as the leaf. We then identify the lowest and highest values in the data set. Sergey Brin is the youngest member of this group at 36 years of age (stem: 3, leaf: 6) and Karl Albrecht is the oldest at 90 years of age (stem: 9, leaf: 0). These values give us the first and last values in the stem. This means the stems will be 3, 4, 5, 6, 7, 8, and 9, as shown in Panel A of Table 2.15.

TABLE 2.15 Constructing a Stem-and-Leaf Diagram for Example 2.6

Panel A		Panel B		Panel C	
Stem	Leaf	Stem	Leaf	Stem	Leaf
3		3	6	3	6
4		4		4	
5		5	4 2 9 3 5 4 2	5	2 2 3 4 4 5 9
6		6	5 1 2 2 0 6 6 8	6	0 1 2 2 5 6 6 8
7	0	7	0 9 4 4	7	0 4 4 9
8		8	3 1 7 2	8	1 2 3 7
9		9	0	9	0

We then begin with the wealthiest man in the world, Carlos Slim Helu, whose age of 70 gives us a stem of 7 and a leaf of 0. We place a 0 in the row corresponding to a stem of 7, as shown in Panel A of Table 2.15. We continue this process with all the other ages and obtain the values in Panel B. Finally, in Panel C we arrange each individual leaf row in ascending order; this is the stem-and-leaf diagram in its final form.

The stem-and-leaf diagram (Panel C) presents the original 25 values in a more organized form. From the diagram we can readily observe that the ages range from

36 to 90. Wealthy individuals in their sixties make up the largest group in the sample with eight members, while those in their fifties place a close second, accounting for seven members. We also note that the distribution is not perfectly symmetric. A stem-and-leaf diagram is similar to a histogram turned on its side with the added benefit of retaining the original values.

EXERCISES 2.3

Mechanics

37. Consider the following data set:

5.4	4.6	3.5	2.8	2.6	5.5	5.5	2.3	3.2	4.2
4.0	3.0	3.6	4.5	4.7	4.2	3.3	3.2	4.2	3.4

Construct a stem-and-leaf diagram. Is the distribution symmetric? Explain.

38. Consider the following data set:

−64	−52	−73	−82	−85	−80	−79	−65	−50	−71
−80	−85	−75	−65	−77	−87	−72	−83	−73	−80

Construct a stem-and-leaf diagram. Is the distribution symmetric? Explain.

Applications

39. A sample of patients arriving at Overbrook Hospital's emergency room recorded the following body temperature readings over the weekend:

100.4	99.6	101.5	99.8	102.1	101.2	102.3	101.2	102.2	102.4
101.6	101.5	99.7	102.0	101.0	102.5	100.5	101.3	101.2	102.2

Construct and interpret a stem-and-leaf diagram.

40. Suppose the following high temperatures were recorded for major cities in the contiguous United States for a day in July.

84	92	96	91	96	94	93	82	81	76
90	95	84	90	84	98	94	90	83	78
88	96	106	78	92	98	91	84	80	94
94	93	107	87	77	99	94	73	74	92

Construct and interpret a stem-and-leaf diagram.

41. A police officer is concerned with excessive speeds on a portion of Interstate 90 with a posted speed limit of 65 miles per hour. Using his radar gun, he records the following speeds for 25 cars and trucks:

66	72	73	82	80	81	79	65	70	71
80	75	75	65	67	67	72	73	73	80
81	78	71	70	70					

Construct a stem-and-leaf diagram. Are the officer's concerns warranted?

42. Spain was the winner of the 2010 World Cup, beating the Netherlands by a score of 1–0. The ages of the players from both teams were as follows:

Spain									
29	25	23	30	32	25	29	30	26	29
21	28	24	21	27	22	25	21	23	24
Netherlands									
27	22	26	30	35	33	29	25	27	25
35	27	27	26	23	25	23	24	26	39

Construct a stem-and-leaf diagram for each country. Comment on similarities and differences between the two data sets.

2.4 SCATTERPLOTS

LO 2.6

Construct and interpret a scatterplot.

All of the tabular and graphical tools presented thus far have focused on describing one variable. However, in many instances we are interested in the relationship between two variables. People in virtually every discipline examine how one variable may systematically influence another variable. Consider, for instance, how

- Incomes vary with education.
- Sales vary with advertising expenditures.

jag39519_ch02_018-059 44 09/18/18 06:53 AM

- Stock prices vary with corporate profits.
- Crop yields vary with the use of fertilizer.
- Cholesterol levels vary with dietary intake.
- Price varies with reliability.

When examining the relationship between two quantitative variables, a **scatterplot** often proves to be a powerful first step in any analysis.

GRAPHICAL DISPLAY OF TWO QUANTITATIVE VARIABLES: A SCATTERPLOT

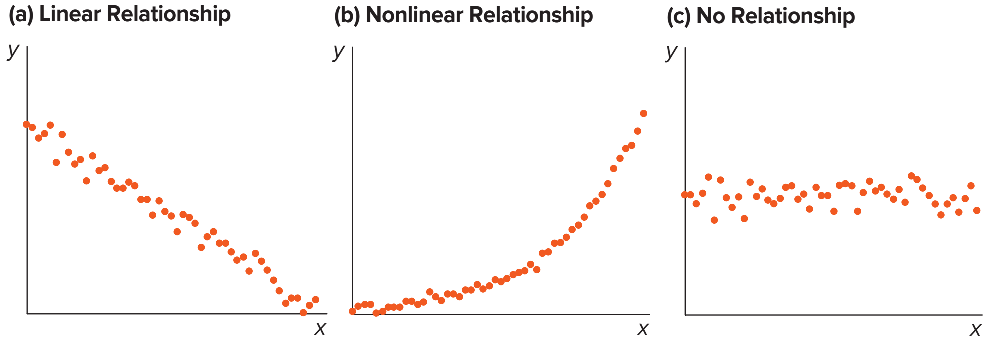
A scatterplot is a graphical tool that helps in determining whether or not two quantitative variables are related in some systematic way. Each point in the diagram represents a pair of observed values of the two variables.

When constructing a scatterplot, we generally refer to one of the variables as x and represent it on the horizontal axis and the other variable as y and represent it on the vertical axis. We then plot each pairing: (x_1, y_1), (x_2, y_2), and so on. Once the data are plotted, the graph may reveal that

- A linear relationship exists between the two variables;
- A nonlinear relationship exists between the two variables; or
- No relationship exists between the two variables.

For example, Figure 2.13(a) shows points on a scatterplot clustered together along a line with a negative slope; we infer that the two variables have a negative linear relationship. Figure 2.13(b) depicts a positive nonlinear relationship; as x increases, y tends to increase at an increasing rate. The points in Figure 2.13(c) are scattered with no apparent pattern; thus, there is no relationship between the two variables.

FIGURE 2.13 Scatterplots depicting relationships between two variables



In order to illustrate a scatterplot, consider the following example.

EXAMPLE 2.7

A social scientist wants to analyze the relationship between educational attainment and income. He collects the data shown in Table 2.16, where Education and Income refer to an individual's years of higher education and annual income (in $1,000s), respectively. Construct and interpret a scatterplot.

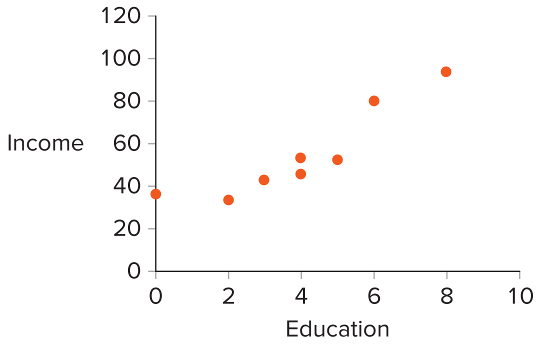
TABLE 2.16 Education and Income for Eight Individuals

Individual	Education	Income
1	3	45
2	4	56
3	6	85
4	2	35
5	5	55
6	4	48
7	8	100
8	0	38

FILE *Edu_Inc*

SOLUTION: We let x and y denote Education and Income, respectively. We plot the first individual's pairing as (3, 45), the second individual's pairing as (4, 56), and so on. Figure 2.14 shows the scatterplot.

FIGURE 2.14 Scatterplot of Income versus Education



As expected, we observe a positive relationship between the two variables; that is, when Education increases, Income tends to increase.

Using Excel to Construct a Scatterplot

In order to demonstrate the construction of a scatterplot, we replicate Figure 2.14.

FILE *Edu_Inc*

A. Open ***Edu_Inc***.

B. Simultaneously select the values in the Education and Income columns, and then choose **Insert > Scatter (X, Y) or Bubble Chart**. From the options, choose the graph at the top left. (If you are having trouble finding this option after selecting **Insert**, look for the graph with data points above **Charts**.)

C. Formatting (regarding axis titles, gridlines, etc.) can be done by selecting **Format >** **Add Chart Element** from the menu.

EXERCISES 2.4

Mechanics

43. Construct a scatterplot with the following data. Describe the relationship between x and y.

x	3	7	12	5	6
y	22	10	5	14	12

44. Construct a scatterplot with the following data. Does a linear relationship exist between x and y?

x	10	4	6	3	7
y	3	2	6	6	4

45. Construct a scatterplot with the following data. Describe the relationship between *x* and *y*.

x	1	2	3	4	5	6	7	8
y	22	20	18	10	5	4	3	2

Applications

46. A statistics instructor wants to examine whether a relationship exists between the hours a student spends studying for the final exam (Hours) and a student's grade on the final exam (Grade). She takes a sample of eight students.

Hours	8	2	3	8	10	15	25	5
Grade	75	47	50	80	85	88	93	55

Construct a scatterplot. What conclusions can you draw from the scatterplot?

47. A study offers evidence that the more weight a woman gains during pregnancy, the higher the risk of having a high-birth-weight baby, defined as at least 8 pounds, 13 ounces, or 4 kilograms (*The Wall Street Journal,* August 5, 2010). High-birth-weight babies are more likely to be obese in adulthood. The weight gain (in kilograms) of eight mothers and the birth weight of their newborns (in kilograms) are recorded in the accompanying table.

Mother's Weight Gain	Newborn's Birth Weight
18	4.0
7	2.5
8	3.0
22	4.5
21	4.0
9	3.5
8	3.0
10	3.5

Construct a scatterplot. Do the results support the findings of the study?

48. In order to diversify risk, investors are often encouraged to invest in assets whose returns have either a negative relationship or no relationship. The annual return data (in %) on two assets is shown in the accompanying table.

Return A	Return B
–20	8
–5	5
18	–1
15	–2
–12	2

Construct a scatterplot. In order to diversify risk, would the investor be wise to include both of these assets in her portfolio? Explain.

49. In an attempt to determine whether a relationship exists between the price of a home (in $1,000s) and the number of days it takes to sell the home, a real estate agent collects data on the recent sales of eight homes.

Price	Days to Sell Home
265	136
225	125
160	120
325	140
430	145
515	150
180	122
423	145

Construct a scatterplot. What can the realtor conclude?

WRITING WITH STATISTICS

The tabular and graphical tools introduced in this chapter are the starting point for most studies and reports that involve statistics. They can help you organize data so you can see patterns and trends in the data, which can then be analyzed by the methods described in later chapters of this text. In this section, we present an example of using tabular and graphical methods in a sample report. Each of the remaining chapters contains a sample report incorporating the concepts developed in that respective chapter.

Camilla Walford is a newly hired journalist for a national newspaper. One of her first tasks is to analyze gas prices in the United States during the week of the Fourth of July holiday.

©Rubberball/Getty Images

She collects average gas prices (in $ per gallon) for the 48 contiguous states and the District of Columbia (DC), a portion of which is shown in Table 2.17.

TABLE 2.17 U.S. Gas Prices, July 2, 2010

FILE
Gas_Prices_2010

State	Price
Alabama	2.59
Arkansas	2.60
⋮	⋮
Wyoming	2.77

Source: *AAA's Daily Fuel Gauge Report*, July 2, 2010.

Camilla wants to use the sample information to

1. Construct frequency distributions to summarize the data.
2. Make summary statements concerning gas prices.
3. Convey the information from the distributions into graphical form.

Sample Report—Gas Prices across the United States

Historically, in the United States, many people choose to take some time off during the Fourth of July holiday period and travel to the beach, the lake, or the mountains. The roads tend to be heavily traveled, making the cost of gas a concern. The following report provides an analysis of gas prices across the nation over this holiday period.

The analysis focuses on the average gas price for the 48 contiguous states and the District of Columbia (henceforth, referenced as 49 states for ease of exposition). The range of gas prices is from a low of $2.52 per gallon (South Carolina) to a high of $3.15 per gallon (California). To find out how gas prices are distributed between these extremes, the data have been organized into several frequency distributions as shown in Table 2.A. The frequency distribution shows that 17 of the 49 states have an average gas price between $2.70 and $2.80 per gallon; or comparably, 35% of the states have an average price in this range as shown by the relative frequency distribution. The cumulative frequency distribution indicates that 35 states have an average price less than $2.80 per gallon. Finally, the cumulative relative frequency distribution shows that the average price in 72% of the states (approximately three-quarters of the sample) is less than $2.80 per gallon.

TABLE 2.A Frequency Distributions for Gas Prices in the United States, July 2, 2010

Average Price ($ per gallon)	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
2.50 up to 2.60	5	0.10	5	0.10
2.60 up to 2.70	13	0.27	18	0.37
2.70 up to 2.80	17	0.35	35	0.72
2.80 up to 2.90	8	0.16	43	0.88
2.90 up to 3.00	4	0.08	47	0.96
3.00 up to 3.10	1	0.02	48	0.98
3.10 up to 3.20	1	0.02	49	1.00
	Sample Size = 49			

Figure 2.A shows a histogram for gas prices. This graph reinforces the fact that the average price of gas nationwide is between $2.50 and $3.20 per gallon. Moreover, gas prices are positively skewed since the distribution runs off to the right; only two states (California and Washington) have gas prices that are more than $3.00 per gallon.

Another useful visual representation of the data is an ogive, shown in Figure 2.B. The ogive is useful for approximating the "middle" price. If we draw a horizontal line to the ogive at the 0.5 relative frequency mark, it intersects the plot at a point corresponding on the horizontal axis to a "middle price" of approximately $2.75. This indicates that gas stations in approximately half of the states charged below this price and half charged above it.

FIGURE 2.A Histogram of average gas prices nationwide

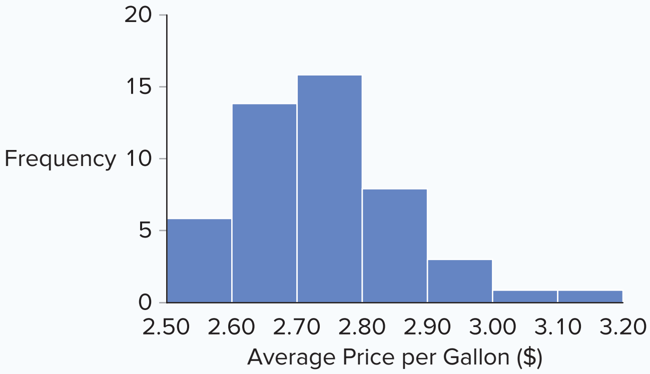
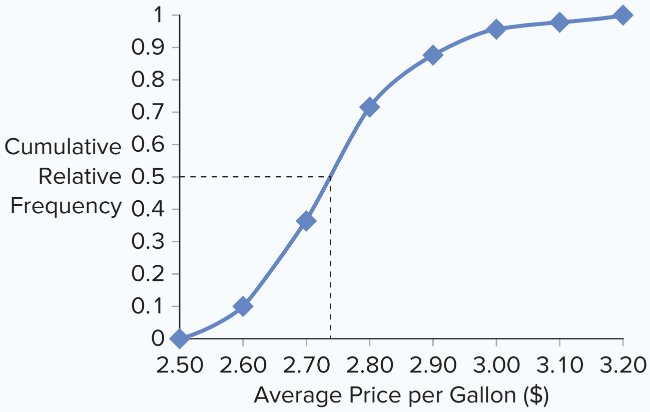


FIGURE 2.B Ogive of average gas prices nationwide



CONCEPTUAL REVIEW

LO 2.1 Summarize qualitative data by constructing a frequency distribution.

For qualitative data, a **frequency distribution** groups data into categories and records the number of observations that fall into each category. A **relative frequency distribution** shows the proportion (or the fraction) of observations in each category.

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

Graphically, we can show a frequency distribution for qualitative data by constructing a **pie chart** or a **bar chart**. A pie chart is a segmented circle that clearly portrays the categories of some qualitative variable. A bar chart depicts the frequency or the relative frequency for each category of the qualitative variable as a series of horizontal or vertical bars, the lengths of which are proportional to the values that are to be depicted.

LO 2.3 Summarize quantitative data by constructing a frequency distribution.

For quantitative data, a **frequency distribution** groups data into intervals called classes, and records the number of observations that falls within each class. A **cumulative frequency distribution** records the number of observations that falls below the upper limit of each class. A **relative frequency distribution** identifies the proportion (or the fraction) of observations that falls within each class. A **cumulative relative frequency distribution** shows the proportion (or the fraction) of observations that falls below the upper limit of each class.

jag39519_ch02_018-059 49 09/18/18 06:53 AM

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

A **histogram** and a **polygon** are graphical representations of a frequency or a relative frequency distribution for quantitative data. An inspection of these graphs reveals where most of the observations tend to cluster, as well as the general shape and spread of the data. An **ogive** is a graphical representation of a cumulative frequency or cumulative relative frequency distribution.

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

A **stem-and-leaf diagram** is another visual method of displaying quantitative data. It is constructed by separating each value of a data set into a *stem,* which consists of the left-most digits, and a *leaf,* which consists of the last digit. Like a histogram and a polygon, a stem-and-leaf diagram gives an overall picture of where the data are centered and how the data are dispersed from the center.

LO 2.6 Construct and interpret a scatterplot.

A **scatterplot** is a graphical tool that helps in determining whether or not two quantitative variables are related in some systematic way. Each point in the diagram represents a pair of observed values of the two variables.

ADDITIONAL EXERCISES AND CASE STUDIES

Exercises

50. A 2003 survey by the Centers for Disease Control and Prevention concluded that smoking is forbidden in nearly 75% of U.S. households (*The Boston Globe,* May 25, 2007). The survey gathered responses from at least 900 households in each state. When residents of Utah were asked whether or not smoking was allowed in their households, a representative sample of responses was as follows:

No	No	No	No	No	No	Yes	No	No	No
No	Yes	No	No	No	No	No	No	No	No

When a similar survey was taken in Kentucky, a representative sample of responses was as follows:

No	No	Yes	No	Yes	No	Yes	Yes	No	No
No	Yes	Yes	No	Yes	No	No	Yes	Yes	No

a. Construct a relative frequency distribution that summarizes the responses of residents from Utah and Kentucky. Comment on the results.
b. Construct a bar chart that summarizes the results for each state.

51. Patrons at a local restaurant were asked to rate their recent experience at the restaurant with respect to its advertised atmosphere of upbeat, comfortable, and clean. Possible responses included Outstanding, Good, OK, and Horrible. The following table shows the responses of 28 patrons:

Horrible	OK	Horrible	Horrible
OK	OK	Horrible	Horrible
Horrible	OK	Horrible	Good
Horrible	Good	Good	Good
Horrible	OK	Horrible	OK
Good	Good	Horrible	Good
Horrible	OK	Horrible	Good

a. Construct a relative frequency distribution that summarizes the responses of the patrons. Briefly summarize your findings. What recommendations would you make to the owner of the restaurant?
b. Construct a pie chart and a bar chart for these data.

52. A survey conducted by CBS News asked parents about the professions they would want their children to pursue. Parents' preferences (in %) are summarized in the following table.

jag39519_ch02_018-059 50 09/18/18 06:53 AM

Profession	Parents' Preference
Doctor, banker, lawyer, or president	65
Internet mogul	13
Humanitarian-aid worker	6
Athlete	9
Movie star, rock star	2
Other	5

Source: *Vanity Fair*, December 2009.

a. Construct a pie chart and a bar chart for these data.
b. How many parents wanted their children to become athletes if the results were based on 550 responses?

53. The one-year return (in %) for 24 mutual funds is as follows:

−14.5	−5.0	−3.7	2.5	−7.9	−11.2
4.8	−16.8	9.0	6.5	8.2	5.3
−12.2	15.9	18.2	25.4	3.4	−1.4
5.5	−4.2	−0.5	6.0	−2.4	10.5

a. Construct the frequency distribution using classes of −20 up to −10, −10 up to 0, etc.
b. Construct the relative frequency, the cumulative frequency, and the cumulative relative frequency distributions.
c. How many of the funds had returns of at least 0% but less than 10%? How many of the funds had returns of 10% or more?
d. What percentage of the funds had returns of at least 10% but less than 20%? What percentage of the funds had returns less than 20%?

54. *The Statistical Abstract of the United States, 2010* provided the following frequency distribution of the number of people (in 1,000s) who live below the poverty level by region.

Region	Number of People
Northeast	6,166
Midwest	7,237
South	15,501
West	8,372

a. Construct the relative frequency distribution. What percentage of people who live below the poverty level live in the Midwest?
b. Construct a pie chart and a bar chart for these data.

55. *Money* magazine (January 2007) reported that an average of 77 million adults in the United States make financial resolutions at the beginning of a new year. Consider the following frequency distribution, which reports the top financial resolutions of 1,026 Americans (MONEY/ICR poll conducted November 8–12, 2006).

Financial Resolution	Frequency
Saving more	328
Paying down debt	257
Making more income	154
Spending less	133
Investing more	103
Saving for a large purchase	41
Don't know	10

a. Construct the relative frequency distribution. What percentage of the respondents indicated that paying down debt was their top financial resolution?
b. Construct the bar chart.

56. A recent poll of 3,057 individuals asked: "What's the longest vacation you plan to take this summer?" The following relative frequency distribution summarizes the results.

Response	Relative Frequency
A few days	0.21
A few long weekends	0.18
One week	0.36
Two weeks	0.25

a. Construct the frequency distribution. How many people are going to take a one-week vacation this summer?
b. Construct the pie chart.

57. A survey conducted by CBS News asked 1,026 respondents: "What would you do with an unexpected tax refund?" The responses (in %) are summarized in the following table;

Response	Percent Frequency
Pay off debts	47
Put it in the bank	30
Spend it	11
I never get a refund	10
Other	2

Source: CBS News Archives.

a. Construct the bar chart.
b. How many people will spend the tax refund?

58. The following table reports the number of people residing in regions in the U.S. as well as the number

of people living below the poverty level in these regions for the year 2013. (All numbers are in 1,000s.)

Region	Total	Below Poverty Level
Northeast	55,478	7,046
Midwest	66,785	8,590
South	116,961	18,870
West	73,742	10,812

Source: www.census.gov/hhes/www/poverty/data/incpovhlth/2013/table3.pdf, data retrieved March 23, 2015.

a. Graph and interpret the pie chart that summarizes the proportion of people who reside in each region.
b. Graph and interpret the pie chart that summarizes the proportion of people living below the poverty level in each region. Is this pie chart consistent with the one you constructed in part (a); that is, in those regions that are relatively less populated, is the proportion of people living below the poverty level less?

59. The manager at a water park constructed the following frequency distribution to summarize attendance in July and August.

Attendance	Frequency
1,000 up to 1,250	5
1,250 up to 1,500	6
1,500 up to 1,750	10
1,750 up to 2,000	20
2,000 up to 2,250	15
2,250 up to 2,500	4

a. Construct the relative frequency, the cumulative frequency, and the cumulative relative frequency distributions.
b. What is the most likely attendance range? How many times was attendance less than 2,000 people?
c. What percentage of the time was attendance at least 1,750 but less than 2,000 people? What percentage of the time was attendance less than 1,750 people? What percentage of the time was attendance 1,750 or more?
d. Construct the relative frequency histogram. Comment on the shape of the distribution.

60. *The Wall Street Journal* (August 28, 2006) asked its readers: "Ideally, how many days a week, if any, would you work from home?" The following relative frequency distribution summarizes the responses from 3,478 readers.

Days Working from Home	Relative Frequency
0	0.12
1	0.18
2	0.30
3	0.15
4	0.07
5	0.19

Construct the pie chart and the bar chart to summarize the data.

61. A researcher conducts a mileage economy test involving 80 cars. The frequency distribution describing average miles per gallon (mpg) appears in the following table.

Average mpg	Frequency
15 up to 20	15
20 up to 25	30
25 up to 30	15
30 up to 35	10
35 up to 40	7
40 up to 45	3

a. Construct the relative frequency, the cumulative frequency, and the cumulative relative frequency distributions.
b. How many of the cars got less than 30 mpg? What percentage of the cars got at least 20 but less than 25 mpg? What percentage of the cars got less than 35 mpg? What percent got 35 mpg or more?
c. Construct the relative frequency histogram. Comment on the shape of the distribution.

62. FILE ***Wealthiest_Americans.*** The accompanying table lists a portion of the ages and net worth (in $ billions) of the wealthiest people in America.

Name	Age	Net Worth
William Gates III	53	50.0
Warren Buffet	79	40.0
⋮	⋮	⋮
Philip Knight	71	9.5

Source: *Forbes*, Special Report, September 2009.

a. What percentage of the wealthiest people in America had net worth more than $20 billion?
b. What percentage of the wealthiest people in America had net worth between $10 billion and $20 billion?
c. Construct a stem-and-leaf diagram on age. Comment on the shape of the distribution.

63. **FILE** ***DOW_PEG.*** The price-to-earnings growth ratio, or PEG ratio, is the market's valuation of a company relative to its earnings prospects. A PEG ratio of 1 indicates that the stock's price is in line with growth expectations. A PEG ratio less than 1 suggests that the stock of the company is undervalued (typical of value stocks), whereas a PEG ratio greater than 1 suggests the stock is overvalued (typical of growth stocks). The accompanying table shows a portion of PEG ratios of companies listed on the Dow Jones Industrial Average.

Company	PEG Ratio
3M (MMM)	1.4
Alcoa (AA)	0.9
⋮	⋮
Walt Disney (DIS)	1.2

Source: www.finance.yahoo, data retrieved April 13, 2011.

Construct the stem-and-leaf diagram on the PEG ratio. Interpret your findings.

64. The following table lists the sale price (in $1,000s) and house type of 20 houses that recently sold in New Jersey.

Price	Type	Price	Type
305	Ranch	568	Colonial
450	Colonial	385	Other
389	Contemporary	310	Contemporary
525	Other	450	Colonial
300	Ranch	400	Other
330	Contemporary	359	Ranch
355	Contemporary	379	Ranch
405	Colonial	509	Colonial
365	Ranch	435	Colonial
415	Ranch	510	Other

a. Construct a frequency distribution for types of houses. Interpret the results.
b. Construct a frequency distribution for house prices. Use six classes, starting with 300, each with a width of 50. Interpret the results.

65. A manager of a local retail store analyzes the relationship between Advertising (in $100s) and Sales (in $1,000s) by reviewing the store's data for the previous six months. Construct a scatterplot and comment on whether or not a relationship exists.

Advertising	Sales
20	15
25	18
30	20
22	16
27	19
26	20

66. The following table lists the National Basketball Association's (NBA's) leading scorers, their average minutes per game (MPG), and their average points per game (PPG) for 2008:

Player	MPG	PPG
D. Wade	38.6	30.2
L. James	37.7	28.4
K. Bryant	36.1	26.8
D. Nowitzki	37.3	25.9
D. Granger	36.2	25.8
K. Durant	39.0	25.3
C. Paul	38.5	22.8
C. Anthony	34.5	22.8
C. Bosh	38.0	22.7
B. Roy	37.2	22.6

Source: www.espn.com.

Construct and interpret a scatterplot of PPG against MPG. Does a relationship exist between the two variables?

CASE STUDIES

CASE STUDY 2.1 There are six broad sectors that comprise the Dow Jones Industrial Average (DJIA). The following table shows a portion of the 30 companies that comprise the DJIA and each company's sector.

FILE
DJIA_Sector

Data for Case Study 2.1 Companies and Sectors of the DJIA

Company	Sector
3M (MMM)	Manufacturing
American Express (AXP)	Finance
⋮	⋮
Walmart (WMT)	Consumer

Source: www.money.cnn.com/data/dow30/, information retrieved March 21, 2015.

In a report, use the sample information to

1. Construct the frequency distribution and the relative frequency distribution for the sectors that comprise the DJIA. Use pie charts for data visualization.
2. Discuss how the various sectors are represented in the DJIA.

CASE STUDY 2.2 When reviewing the overall strength of a particular firm, financial analysts typically examine the net profit margin. This statistic is generally calculated as the ratio of a firm's net profit after taxes (net income) to its revenue, expressed as a percentage. For example, a 20% net profit margin means that a firm has a net income of $0.20 for each dollar of sales. A net profit margin can even be negative if the firm has a negative net income. In general, the higher the net profit margin, the more effective the firm is at converting revenue into actual profit. The net profit margin serves as a good way of comparing firms in the same industry, since such firms generally are subject to the same business conditions. However, financial analysts also use the net profit margin to compare firms in different industries in order to gauge which firms are relatively more profitable. The accompanying table shows a portion of net profit margins (in %) for a sample of clothing retailers.

FILE
Net_Profit_Margins

Data for Case Study 2.2 Net Profit Margin for Clothing Retailers

Firm	Net Profit Margin
Abercrombie & Fitch	1.58
Aéropostale	10.64
⋮	⋮
Wet Seal	16.15

Source: www.finance.yahoo.com, data retrieved July 2010.

In a report, use the sample information to

1. Provide a brief definition of net profit margin and explain why it is an important statistic.
2. Construct appropriate tables (frequency distribution, relative frequency distribution, etc.) and graphs that summarize the clothing industry's net profit margin. Use −5, 0, 5, and so on, for the upper limits of the classes for the distributions.
3. Discuss where the data tend to cluster and how the data are spread from the lowest value to the highest value.
4. Comment on the net profit margin of the clothing industry, as compared to the beverage industry's net profit margin of approximately 10.9% (Source: biz.yahoo, July 2010).

CASE STUDY 2.3 The following table lists a portion of U.S. life expectancy (in years) for the 50 states.

Data for Case Study 2.3 Life Expectancy by State, 2010–2011

Rank	State	Life Expectancy
1	Hawaii	81.5
2	Minnesota	80.9
⋮	⋮	⋮
50	Mississippi	74.8

Source: en.wikipedia.org/wiki/List_of_U.S._states_by_life_expectancy, data retrieved April 25, 2012.

FILE
Life_Expectancy

In a report, use the sample information to

1. Construct appropriate tables (frequency distribution, relative frequency distribution, etc.) and graphs to summarize life expectancy in the United States. Use 75, 76.5, 78, and so on, for the upper limits of the classes for the distributions.
2. Discuss where the data tend to cluster and how the data are spread from the lowest value to the highest value.
3. Comment on the shape of the distribution.

APPENDIX 2.1 Guidelines for Other Software Packages

The following section provides brief commands for Minitab, SPSS, JMP, and R. Copy and paste the specified data file into the relevant software spreadsheet prior to following the commands. When importing data into R, use the menu-driven option File > Import Dataset > From Excel.

Minitab

Pie Chart

FILE
Marital_Status

A. (Replicating Figure 2.1) From the menu, choose **Graph > Pie Chart**. Select **Chart values from a table**, select Marital Status as the **Categorical variable**, and 1960 and 2010 as the **Summary variables**.

B. Choose **Labels**. Select **Titles/Footnotes** and enter Marital Status, 1960 versus 2010. Then select **Slice Labels** and select **Category name** and **Percent**.

C. Choose **Multiple Graphs**, and then select **On the same graph**.

Bar Chart

FILE
Marital_Status

A. (Replicating Figure 2.2) From the menu, choose **Graph > Bar Chart**. From **Bars Represent** select **Values from a Table**, and from **Two-way Table** select **Cluster**.

B. In the *Bar Chart—Two-Way Table—Cluster* dialog box, select 1960 and 2010 as **Graph variables**. Select Marital Status as **Row labels**. Under **Table Arrangement**, choose **Rows are outermost categories and columns are innermost**.

jag39519_ch02_018-059 55 09/18/18 06:53 AM

Histogram

From Raw Data:

FILE
MV_Houses

A. (Replicating Figure 2.5) From the menu, choose **Graph > Histogram > Simple**. Click **OK**.

B. Select House Price as **Graph Variables**. Click **OK**.

C. Double-click *x*-axis and select **Edit Scale**. Under **Major Tick Positions**, choose **Position of Ticks** and enter 300 400 500 600 700 800. Under **Scale Range**, deselect **Auto** for *Minimum* and enter 300. Then deselect **Auto** for *Maximum* and enter 800. Select the **Binning** tab. Under **Interval Type**, select **Cutpoint**. Under **Interval Definition**, select **Midpoint/Cutpoint Definitions** and enter 300 400 500 600 700 800.

From a Frequency Distribution:

FILE
MV_Frequency

A. (Replicating Figure 2.5) From the menu, choose **Graph > Bar Chart**. From **Bars Represent** select **A function of a variable**, and from **One Y** select **Simple**. Click **OK**.

B. Under **Function** select **Sum**. Under **Graph variables** select **Frequency**, and under **Categorical variable** select **Class (in $1,000s)**.

C. Double-click *x*-axis. Under **Space Between Scale Categories**, deselect **Gap between Cluster** and enter 0.

Polygon

FILE
Polygon

A. (Replicating Figure 2.8) From the menu, choose **Graph > Scatterplot > With Connect Line**.

B. Under **Y variables** select y, and under **X variables** select x.

Ogive

FILE
Ogive

A. (Replicating Figure 2.9) From the menu, choose **Graph > Scatterplot > With Connect Line**.

B. Under **Y variables** select y, and under **X variables** select x.

Scatterplot

FILE
Edu_Inc

A. (Replicating Figure 2.14) From the menu, choose **Graph > Scatterplot > Simple**.

B. Under **Y variables** select Income, and under **X variables** select Education.

SPSS

Pie Chart

FILE
Marital_Status

A. (Replicating Figure 2.1) From the menu, choose **Graphs > Legacy Dialogs > Pie**. Under **Data in Chart Are**, select **Values of individual cases**. Click **Define**.

B. Under **Slices Represent**, select 1960. Under **Slices Labels**, select **Variable**, then select Marital Status.

C. Double-click on the graph to open **Chart Editor**, and then choose **Elements > Show Data Labels**. In the **Properties** dialog box, under **Displayed** select Percent and Marital Status.

Bar Chart

FILE
Marital_Status

A. (Replicating Figure 2.2) From the menu, choose **Graphs > Legacy Dialogs > Bar**. Choose **Clustered**. Under **Data in Chart Are**, select **Values of individual cases**. Click **Define**.

B. Under **Bars Represent**, select 1960 and 2010. Under **Category Labels**, select **Variable**, then select Marital Status.

Histogram

A. (Replicating Figure 2.5) From the menu, choose **Graphs > Legacy Dialogs > Histogram**. Under **Variable**, select HousePrice.

B. In the Output window, double-click on Frequency (*y*-axis title), choose the **Scale** tab, and under **Range**, enter 0 as **Minimum**, 15 as **Maximum**, and 5 as **Major Increment**. Then click **Apply**.

C. Double-click on the bars. Choose the **Binning** tab, and under **X Axis**, select **Custom** and **Interval width**, and enter 100 for the interval width. Then click **Apply**.

FILE *MV_Houses*

Polygon

A. (Replicating Figure 2.8) From the menu, choose **Graphs > Legacy Dialogs > Scatter/Dot**. Choose **Simple Scatter** and then click **Define**.

B. Under **Y Axis** select y, and under **X Axis** select x. Click **OK**.

C. In the Output window, double-click on the graph to open the **Chart Editor**, then from the menu choose **Elements > Interpolation Line**.

FILE *Polygon*

Ogive

A. (Replicating Figure 2.9) From the menu, choose **Graphs > Legacy Dialogs > Scatter/Dot**. Choose **Simple Scatter**. Then click **Define**.

B. Under **Y Axis**, select y, and under **X Axis**, select x. Click **OK**.

C. In the Output window, double-click on the graph to open the **Chart Editor**, then from the menu choose **Elements > Interpolation Line**. Then click **Apply**. From the menu choose **Edit > Select X Axis**. Choose the **Scale** tab, and under **Range**, enter 300 as **Minimum**, 800 as **Maximum**, 100 as **Major Increment**, and 300 as **Origin**. Then click **Apply**.

FILE *Ogive*

Scatterplot

A. (Replicating Figure 2.14) From the menu, choose **Graphs > Legacy Dialogs > Scatter/Dot**. Choose **Simple Scatter**. Then click **Define**.

B. Under **Y Axis**, select Income, and under **X Axis**, select Education. Click **OK**.

C. In the Output window, double-click on the graph to open the **Chart Editor**. From the menu, choose **Edit > Select Y Axis**. Under **Range**, enter 0 as **Minimum**, 120 as **Maximum**, and 20 as **Major Increment**. Then click **Apply**. From the menu choose **Edit > Select X Axis**. Choose the **Scale** tab, and under **Range**, enter 0 as **Minimum**, 10 as **Maximum**, and 2 as **Major Increment**. Then click **Apply**.

FILE *Edu_Inc*

JMP

Pie Chart

(Replicating Figure 2.1) From the menu, choose **Graph > Chart**. Under **Select Columns**, select Marital Status, and then under **Cast Selected Columns into Roles**, select **Categories, X, Levels**. Under **Select Columns**, select 1960 and 2010, and then select **Statistics > % of Total**. Under **Options**, choose **Pie Chart**. In order to add percentages to the pie chart, click the red arrow next to **Chart > Label Options > Label by Percent of Total Values**.

FILE *Marital_Status*

Bar Chart

(Replicating Figure 2.2) From the menu, choose **Graph > Chart**. Under **Select Columns**, select **Marital Status**, and then under **Cast Selected Columns into Roles**, select **Categories, X, Levels**. Under Select Columns, select 1960 and 2010, and select **Statistics > Data**. Under **Options**, select **Overlay** and **Bar Chart**.

FILE *Marital_Status*

jag39519_ch02_018-059 57 09/18/18 06:53 AM

Histogram

FILE
MV_Houses

A. (Replicating Figure 2.5) From the menu, choose **Analyze > Distribution**. Under **Select Columns**, select House Price, then under **Cast Selected Columns into Roles**, select **Y, columns**.

B. Right-click on the *y*-axis and select **Axis Settings**. For **Minimum**, enter 300; for **Maximum** enter 800; and for **Increment**, enter 100. (JMP automatically produces a histogram in a vertical layout. In order to change the layout to horizontal, click the red arrow next to **House Price > Display Options > Horizontal Layout**.)

Polygon

FILE
Polygon

A. (Replicating Figure 2.8) From the menu, choose **Graph > Overlay Plot**. Under **Select Columns**, select x, and then under **Cast Selected Columns into Roles**, select **X**. Under **Select Columns**, select y, and then under **Cast Selected Columns into Roles**, select **Y**.

B. Click on the red triangle next to the title **Overlay Plot**. Select **Y Options > Connect Points**.

Ogive

FILE
Ogive

A. (Replicating Figure 2.9) From the menu, choose **Graph > Overlay Plot**. Under **Select Columns**, select x, and then under **Cast Selected Columns into Roles**, select **X**. Under **Select Columns**, select y, and then under **Cast Selected Columns into Roles**, select **Y**.

B. Click on the red triangle next to the title **Overlay Plot.** Select **Y Options > Connect Points**.

Scatterplot

FILE
Edu_Inc

(Replicating Figure 2.14) From the menu, choose **Graph > Overlay Plot**. Under **Select Columns**, select Education, and then under **Cast Selected Columns into Roles**, select **X**. Under **Select Columns**, select Income, and then under **Cast Selected Columns into Roles**, select **Y**.

R

Pie Chart

FILE
Marital_Status

(Replicating Figure 2.1) Use the **pie** function. For options within the function, use *labels* to indicate the names for each category and *main* to designate a title. Enter:

```
> pie(Marital_Status$'1960', labels = Marital_Status$'Marital Status',
  main = "Marital Status, 1960")
```

Histogram

FILE
MV_Houses

(Replicating Figure 2.5) Use the **hist** function. For options within the function, use *breaks* to denote the number of distinct intervals, *main* to designate a title, and *xlab* to label the *x*-axis. Enter:

```
> hist(MV_Houses$'House Price', breaks = 5, main = "Histogram",
  xlab = "House Prices (in $1,000s)")
```

Polygon

A. (Replicating Figure 2.8) Use the **plot** function. For options within the function, use *ylab* and *xlab* to label the *y*-axis and the *x*-axis, respectively. Enter:

FILE
Polygon

```
> plot(Polygon$'y' ~  Polygon$'x', ylab="Relative Frequency",
  xlab="House Prices (in $1,000s)")
```

B. Add lines to the scatterplot using the **lines** function. Enter:

```
> lines(Polygon$'y' ~  Polygon$'x')
```

Ogive

A. (Replicating Figure 2.9) Refer to the instructions for the polygon for specifics about the **plot** and the **line** functions. Enter:

FILE
Ogive

```
> plot(Ogive$'y' ~  Ogive$'x', ylab="Cumulative Relative Frequency",
  xlab="House Prices (in $1,000s)")
> lines(Ogive$'y' ~  Ogive$'x')
```

Scatterplot

(Replicating Figure 2.14) Refer to the instructions for the polygon for specifics about the **plot** function. Enter:

FILE
Edu_Inc

```
> plot(Edu_Inc$'Income' ~  Edu_Inc$'Education', ylab = "Income",
  xlab = "Education")
```