

MATHEMATICS

A PRACTICAL ODYSSEY

SEVENTH EDITION

DAVID B. JOHNSON

DIABLO VALLEY COLLEGE
PLEASANT HILL, CALIFORNIA

THOMAS A. MOWRY

DIABLO VALLEY COLLEGE
PLEASANT HILL, CALIFORNIA



Australia • Brazil • Japan • Korea • Mexico • Singapore • Spain • United Kingdom • United States

Mathematics: A Practical Odyssey, 7e**David B. Johnson / Thomas A. Mowry**

Publisher: Charles Van Wagner

Acquisitions Editor: Marc Bove

Developmental Editor: Stefanie Beeck

Assistant Editor: Shaun Williams

Media Editor: Guanglei Zhang

Marketing Manager: Ashley Pickering

Marketing Assistant: Angela Kim

Marketing Communications Manager: Mary Anne Payumo

Senior Content Project Manager: Carol Samet

Creative Director: Rob Hugel

Art Director: Vernon Boes

Print Buyer: Linda Hunt

Rights Acquisitions Specialist: Don Schlotman

Production Service: Anne Seitz/Hearthside Publishing Services

Text Designer: Terri Wright

Photo Researcher: Sarah Bonner, Bill Smith Group

Copy Editor: Barbara Willette/Hearthside Publishing Services

Illustrator: Jade Myers/Hearthside Publishing Services

Cover Designer: Terri Wright

Cover Image: All Royalty-free Photodisc/Getty Images unless specified below: Book illustrations from *Alice's Adventures in Wonderland* by Lewis Carroll: "Oh! The Duchess!" by John Tenniel.

© Lebrecht Music & Arts/Corbis; Tweedledee and Tweedledum, © The Granger Collection.

Elections at Rockefeller Square, © AP Images/ Kathy Willens Colored shapes, © Bard Sadowski/iStockphoto.com Passion flower © Jennifer Weinberg/Alamy Abstract, Royalty-free Dex

Image Jackpot balls, © Wasabi/Alamy Gambling table, © Bill Bachman/Photoedit Seattle

space needle © Brian Yarvin, The Image Works Hobbits from Middle Earth, Lord of the

Rings, © Warner Brothers/Photofest Abacus, © Stock Connection/Superstock Background

image: © Maciej Frowlow, Brand X Pictures/ Getty Images

Compositor: MPS Limited, a Macmillan Company

© 2012, 2007 Brooks/Cole, Cengage Learning

ALL RIGHTS RESERVED. No part of this work covered by the copyright herein may be reproduced, transmitted, stored, or used in any form or by any means graphic, electronic, or mechanical, including but not limited to photocopying, recording, scanning, digitizing, taping, Web distribution, information networks, or information storage and retrieval systems, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the publisher.

For product information and technology assistance, contact us at **Cengage Learning Customer & Sales Support, 1-800-354-9706.**

For permission to use material from this text or product, submit all requests online at **www.cengage.com/permissions.**

Further permissions questions can be e-mailed to **permissionrequest@cengage.com.**

Library of Congress Control Number: 2010938098

Student Edition:

ISBN-13: 978-0-538-49505-9

ISBN-10: 0-538-49505-7

Brooks/Cole20 Davis Drive
Belmont, CA 94002-3098
USA

Cengage Learning is a leading provider of customized learning solutions with office locations around the globe, including Singapore, the United Kingdom, Australia, Mexico, Brazil, and Japan. Locate your local office at **www.cengage.com/global.**

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

To learn more about **Brooks/Cole**, visit **www.cengage.com Brooks/Cole.**

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

Printed in the United States of America

1 2 3 4 5 6 7 14 13 12 11 10

CONTENTS



1 Logic 1

- 1.1** Deductive versus Inductive Reasoning 3
- 1.2** Symbolic Logic 20
- 1.3** Truth Tables 30
- 1.4** More on Conditionals 43
- 1.5** Analyzing Arguments 51
- Chapter 1 Review 63



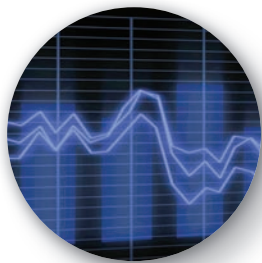
2 Sets and Counting 67

- 2.1** Sets and Set Operations 69
- 2.2** Applications of Venn Diagrams 81
- 2.3** Introduction to Combinatorics 94
- 2.4** Permutations and Combinations 102
- 2.5** Infinite Sets 117
- Chapter 2 Review 127



3 Probability 131

- 3.1** History of Probability 132
- 3.2** Basic Terms of Probability 140
- 3.3** Basic Rules of Probability 157
- 3.4** Combinatorics and Probability 169
- 3.5** Expected Value 181
- 3.6** Conditional Probability 191
- 3.7** Independence; Trees in Genetics 206
- Chapter 3 Review 218



4 Statistics 223

- 4.1** Population, Sample, and Data 224
- 4.2** Measures of Central Tendency 248
- 4.3** Measures of Dispersion 260
- 4.4** The Normal Distribution 277
- 4.5** Polls and Margin of Error 295
- 4.6** Linear Regression 309
- Chapter 4 Review 326



5 Finance 329

- 5.1** Simple Interest 330
- 5.2** Compound Interest 342
- 5.3** Annuities 356
- 5.4** Amortized Loans 369
- 5.5** Annual Percentage Rate with a TI's TVM Application 388
- 5.6** Payout Annuities 394
- Chapter 5 Review 403



6 Voting and Apportionment 407

- 6.1** Voting Systems 408
- 6.2** Methods of Apportionment 429
- 6.3** Flaws of Apportionment 457
- Chapter 6 Review 470



7 Number Systems and Number Theory 473

- 7.1** Place Systems 475
- 7.2** Addition and Subtraction in Different Bases 490
- 7.3** Multiplication and Division in Different Bases 496
- 7.4** Prime Numbers and Perfect Numbers 501
- 7.5** Fibonacci Numbers and the Golden Ratio 513
- Chapter 7 Review 524



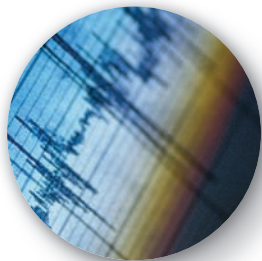
8 Geometry 527

- 8.1** Perimeter and Area 529
- 8.2** Volume and Surface Area 545
- 8.3** Egyptian Geometry 556
- 8.4** The Greeks 567
- 8.5** Right Triangle Trigonometry 579
- 8.6** Linear Perspective 595
- 8.7** Conic Sections and Analytic Geometry 605
- 8.8** Non-Euclidean Geometry 617
- 8.9** Fractal Geometry 628
- 8.10** The Perimeter and Area of a Fractal 648
- Chapter 8 Review 657



9 Graph Theory 663

- 9.1** A Walk through Königsberg 664
- 9.2** Graphs and Euler Trails 670
- 9.3** Hamilton Circuits 683
- 9.4** Networks 696
- 9.5** Scheduling 713
- Chapter 9 Review 725



10 Exponential and Logarithmic Functions 731

- 10.0A** Review of Exponentials and Logarithms 732
- 10.0B** Review of Properties of Logarithms 743
- 10.1** Exponential Growth 756
- 10.2** Exponential Decay 771
- 10.3** Logarithmic Scales 788
- Chapter 10 Review 805



11 Matrices and Markov Chains 807

- 11.0** Review of Matrices 808
- 11.1** Introduction to Markov Chains 821
- 11.2** Systems of Linear Equations 833
- 11.3** Long-Range Predictions with Markov Chains 840
- 11.4** Solving Larger Systems of Equations 845
- 11.5** More on Markov Chains 850
- Chapter 11 Review 855



12 Linear Programming 859

- 12.0** Review of Linear Inequalities 860
- 12.1** The Geometry of Linear Programming 871
- 12.2** Introduction to the Simplex Method 12-2
- 12.3** The Simplex Method: Complete Problems 12-8
- Chapter 12 Review 887



13 The Concepts and History of Calculus 13-1

- 13.0** Review of Ratios, Parabolas, and Functions 13-2
- 13.1** The Antecedents of Calculus 13-12
- 13.2** Four Problems 13-22
- 13.3** Newton and Tangent Lines 13-35
- 13.4** Newton on Falling Objects and the Derivative 13-42
- 13.5** The Trajectory of a Cannonball 13-54
- 13.6** Newton and Areas 13-68
- 13.7** Conclusion 13-76
- Chapter 13 Review 13-83

Appendices

- A** Using a Scientific Calculator A-1
- B** Using a Graphing Calculator A-9
- C** Graphing with a Graphing Calculator A-19
- D** Finding Points of Intersection with a Graphing Calculator A-23
- E** Dimensional Analysis A-25
- F** Body Table for the Standard Normal Distribution A-30
- G** Selected Answers to Odd Exercises A-31

Index I-1

*Sections in color can be found at www.cengage.com/math/johnson

PROBABILITY

3



© Wasabi/Alamy

Uncertainty is a part of our lives. When we wake up, we check the weather report to see whether it might rain. We check the traffic report to see whether we might get stuck in a jam. We check to see whether the interest rate has changed for that car, boat, or home loan. **Probability theory** is the branch of mathematics that analyzes uncertainty.



Dennis Flaherty/photodisc/Jupiter Images

WHAT WE WILL DO IN THIS CHAPTER

GAMBLING IS ALL ABOUT PROBABILITIES:

- Gambling involves uncertainty. Which number will be the winning number?
- All casinos use probability theory to set house odds so that they will make a profit.
- We will use probability theory to determine which casino games of chance are better for you to play and which to avoid.
- Probability theory's roots in gambling are deep: Probability theory got its start when a French nobleman and gambler asked his mathematician friend why he had lost so much money at dice.

THE SCIENCE OF GENETICS COULDN'T EXIST WITHOUT PROBABILITIES:

- Humans inherit traits, such as hair color or an increased risk for a certain disease. A child will inherit some traits but not others, and geneticists use probability theory to analyze this uncertainty.

continued

WHAT WE WILL DO IN THIS CHAPTER — *continued*

- We will use probability theory to analyze the inheritance of hair color—the color of the parents' hair determines a range of possibilities for the color of their child's hair.
- We will also analyze the inheritance of sickle-cell anemia, cystic fibrosis, and other inherited diseases. Your future children might be at risk for these diseases because they can be inherited from disease-free parents.
- Genetics got its start when Gregor Mendel, a nineteenth-century monk, used probability theory to analyze the effect of randomness on heredity.

BELIEVE IT OR NOT, YOUR CELL PHONE USES PROBABILITIES:

- Your cell phone's predictive text feature ("T9" and "iTap" are two examples) allows you to type text messages with a single keypress for each letter, rather than the multitap approach required by older phones. When you type "4663," you could be wanting the word "good" or "home" or "gone." The software looks "4663" up in a dictionary and selects the word that you use with the highest probability.

BUSINESSES USE PROBABILITIES:

- All insurance companies use probability theory to set fees so that the companies will make a profit.
- Investment firms use probability theory to assess financial risk.
- Medical firms use probability theory in diagnosing ailments and determining appropriate treatments.

3.1 History of Probability

OBJECTIVES

- Understand the types of problems that motivated the invention of probability theory
- Become familiar with dice, cards, and the game of roulette

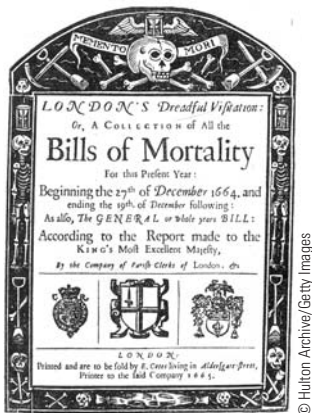
Origins in Gambling

Probability theory is generally considered to have begun in 1654 when the French nobleman (and successful gambler) Antoine Gombauld, the Chevalier de Méré, asked his mathematician friend Blaise Pascal why he had typically made money with one dice bet but lost money with another bet that he thought was

similar. Pascal wrote a letter about this to another prominent French mathematician, Pierre de Fermat. In answering the Chevalier's question, the two launched probability theory.

More than a hundred years earlier, in 1545, the Italian physician, mathematician, and gambler Gerolamo Cardano had used probabilities to help him win more often. His *Liber de ludo aleae* (*Book on Games of Chance*) was the first book on probabilities and gambling. It also contained a section on effective cheating methods. Cardano's work was ignored until interest in Pascal's and Fermat's work prompted its publication.

For a long time, probability theory was not viewed as a serious branch of mathematics because of its early association with gambling. This started to change when Jacob Bernoulli, a prominent Swiss mathematician, wrote *Ars Conjectandi* (*The Art of Conjecture*), published in 1713. In it, Bernoulli developed the mathematical theory of probabilities. Its focus was on gambling, but it also suggested applications of probability to government, economics, law, and genetics.



This *Bills of Mortality* (records of death) was published shortly after John Graunt's analysis.

Other Early Uses: Insurance and Genetics

In 1662, the Englishman John Graunt published his *Natural and Political Observations on the Bills of Mortality*, which used probabilities to analyze death records and was the first important incidence of the use of probabilities for a purpose other than gambling. (It preceded Bernoulli's book.) English firms were selling the first life insurance policies, and they used mortality tables to set fees that were appropriate to the risks involved but still allowed the companies to make a profit.

In the 1860s, Gregor Mendel, an Austrian monk, experimented with pea plants in the abbey garden. His experiments and theories allowed for randomness in the passing on of traits from parent to child, and he used probabilities to analyze the effect of that randomness. In combining biology and mathematics, Mendel founded the science of genetics.

Roulette

Probability theory was originally created to aid gamblers, and games of chance are the most easily and universally understood topic to which probability theory can be applied. Learning about probability and games of chance may convince you not to be a gambler. As we will see in Section 3.5, there is not a single good bet in a casino game of chance.

Roulette is the oldest casino game still being played. Its invention has variously been credited to Pascal, the ancient Chinese, a French monk, and the Italian mathematician Don Pasquale. It became popular when a French policeman introduced it to Paris in 1765 in an attempt to take the advantage away from dishonest gamblers. When the Monte Carlo casino opened in 1863, roulette was the most popular game, especially among the aristocracy. The American roulette wheel has thirty-eight numbered compartments around its circumference. Thirty-six of these compartments are numbered from 1 to 36 and are colored red or black. The remaining two are numbered 0 and 00 and are colored green. Players place their bets by putting their chips on an appropriate spot on the roulette table (see Figure 3.1) on page 134. The dealer spins the wheel and then drops a ball onto the spinning wheel. The ball eventually comes to rest in one of the compartments, and that

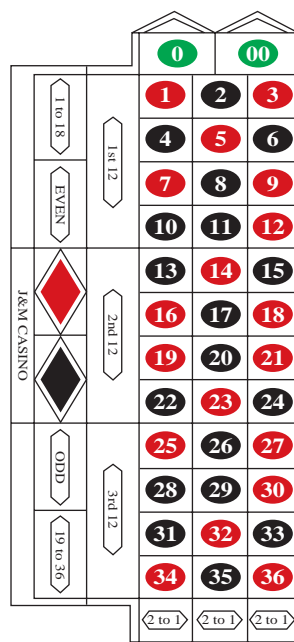


Roulette quickly became a favorite game of the French upper class, as shown in this 1910 photo of Monte Carlo.

compartment’s number is the winning number. (See on page 183 for a photo of the roulette wheel and table.)

For example, if a player wanted to bet \$10 that the ball lands in compartment number 7, she would place \$10 worth of chips on the number 7 on the roulette table. This is a single-number bet, so house odds are 35 to 1 (see Figure 3.1 for house odds). This means that if the player wins, she wins $\$10 \cdot 35 = \350 , and if she loses, she loses her \$10.

Similarly, if a player wanted to bet \$5 that the ball lands either on 13 or 14, he would place \$5 worth of chips on the line separating numbers 13 and 14 on the roulette table. This is a two-numbers bet, so house odds are 17 to 1. This means that if the player wins, he wins $\$5 \cdot 17 = \85 , and if he loses, he loses his \$5.



Bet	House Odds
single number	35 to 1
two numbers (“split”)	17 to 1
three numbers (“street”)	11 to 1
four numbers (“square”)	8 to 1
five numbers (“line”)	6 to 1
six numbers (“line”)	5 to 1
twelve numbers (column or section)	2 to 1
low or high (1 to 18 or 19 to 36, respectively)	1 to 1
even or odd (0 and 00 are neither even nor odd)	1 to 1
red or black	1 to 1

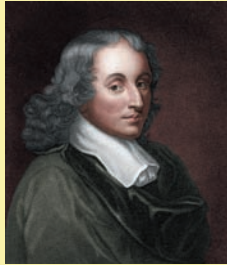
FIGURE 3.1 The roulette table and house odds for the various roulette bets.

HISTORICAL NOTE

BLAISE PASCAL, 1623–1662

As a child, Blaise Pascal showed an early aptitude for science and mathematics, even though he was discouraged from studying in order to protect his poor health. Acute digestive problems and chronic insomnia made his life miserable. Few of Pascal's days were without pain.

At age sixteen, Pascal wrote a paper on geometry that won the respect of the French mathematical community and the jealousy of the prominent French mathematician René Descartes. It has been suggested that the animosity between Descartes and Pascal was in part due to their religious differences. Descartes was a Jesuit, and Pascal was a Jansenist. Although Jesuits and



© Sheila Terry/Photo Researchers

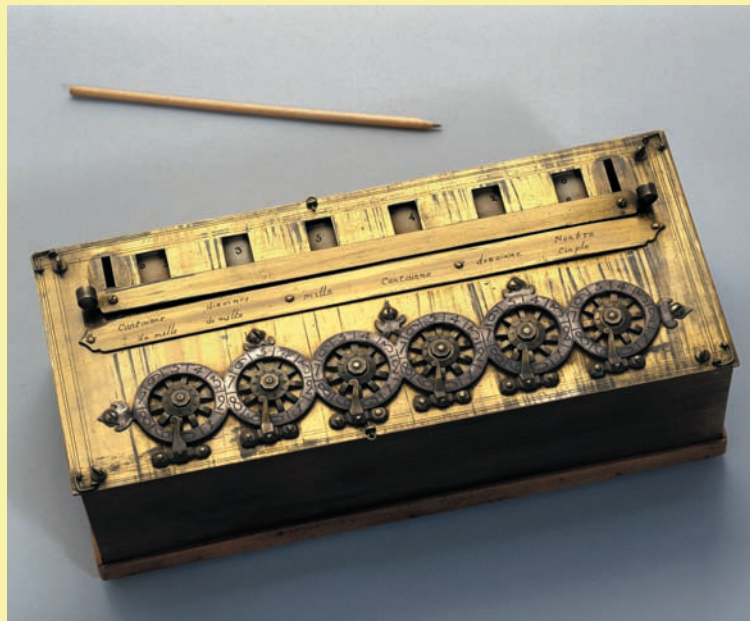
Jansenists were both Roman Catholics, Jesuits believed in free will and supported the sciences, while Jansenists believed in predestination and mysticism and opposed the sciences and the Jesuits.

At age nineteen, to assist his father, a tax administrator, Pascal invented the first calculating machine. Besides co-founding probability theory with Pierre de Fermat, Pascal contributed to the advance of calculus, and his studies in physics culminated with a law on the effects of pressure on fluids that bears his name.

At age thirty-one, after a close escape from death in a carriage accident, Pascal turned his back on mathematics and the sciences and

focused on defending Jansenism against the Jesuits. Pascal came to be the greatest Jansenist, and he aroused a storm with his anti-Jesuit *Provincial Letters*. This work is still famous for its polite irony.

Pascal turned so far from the sciences that he came to believe that reason is inadequate to solve humanity's difficulties or to satisfy our hopes, and he regarded the pursuit of science as a vanity. Even so, he still occasionally succumbed to its lure. Once, to distract himself from pain, he concentrated on a geometry problem. When the pain stopped, he decided that it was a signal from God that he had not sinned by thinking of mathematics instead of his soul. This incident resulted in the only scientific work of his last few years—and his last work. He died later that year, at age thirty-nine.



© SSP/The Image Works

Pascal's calculator, the Pascaline.

HISTORICAL NOTE

GEROLAMO CARDANO, 1501–1576

Gerolamo Cardano is the subject of much disagreement among historians. Some see him as a man of tremendous accomplishments, while others see him as a plagiarist and a liar. All agree that he was a compulsive gambler.

Cardano was trained as a medical doctor, but he was initially denied admission to the College of Physicians of Milan. That denial was ostensibly due to his illegitimate birth, but some suggest that the denial was in fact due to his unsavory reputation as a gambler, since illegitimacy was neither a professional nor a social obstacle in sixteenth-century Italy. His lack of professional success left him with much free time, which he spent gambling and reading. It also resulted in a stay in the poorhouse.

Cardano's luck changed when he obtained a lectureship in mathematics, astronomy, and astrology at the University of Milan. He wrote a



© North Wind/North Wind Picture Archives—All rights reserved.

number of books on mathematics and became famous for publishing a method of solving third-degree equations. Some claim that Cardano's mathematical success

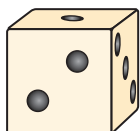
was due not to his own abilities but rather to those of Ludovico Ferrari, a servant of his who went on to become a mathematics professor.

While continuing to teach mathematics, Cardano returned to the practice of medicine. (He was finally allowed to join the College of Physicians, perhaps owing to his success as a mathematician.) Cardano wrote books on medicine and the natural sciences that were well thought of. He became one of the most highly regarded physicians in Europe and counted many prominent people among his patients. He designed a tactile system, somewhat like braille, for the blind. He also designed an undercarriage suspension device that was later adapted as a universal joint

for automobiles and that is still called a *cardan* in Europe.

Cardano's investment in gambling was enormous. He not only wagered (and lost) a great deal of money but also spent considerable time and effort calculating probabilities and devising strategies. His *Book on Games of Chance* contains the first correctly calculated theoretical probabilities.

Cardano's autobiography, *The Book of My Life*, reveals a unique personality. He admitted that he loved talking about himself, his accomplishments, and his illnesses and diseases. He frequently wrote of injuries done him by others and followed these complaints with gleeful accounts of his detractors' deaths. Chapter titles include "Concerning my friends and patrons," "Calumny, defamations, and treachery of my unjust accusers," "Gambling and dicing," "Religion and piety," "The disasters of my sons," "Successes in my practice," "Things absolutely supernatural," and "Things of worth which I have achieved in various studies."



A die. The other three faces have four, five, and six spots.

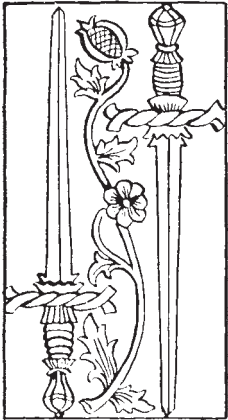
Dice and Craps

Dice have been cast since the beginning of time, for both divination and gambling purposes. The earliest **die** (singular of *dice*) was an animal bone, usually a knucklebone or foot bone. The Romans were avid dice players. The Roman emperor Claudius I wrote a book titled *How to Win at Dice*. During the Middle Ages, dicing schools and guilds of dicers were quite popular among the knights and ladies.

Hazard, an ancestor of the dice game craps, is an English game that was supposedly invented by the Crusaders in an attempt to ward off boredom during long, drawn-out sieges. It became quite popular in England and France in the nineteenth century. The English called a throw of 2, 3, or 12 *crabs*, and it is believed that *craps* is a French mispronunciation of that term. The game came to America with the French colonization of New Orleans and spread up the Mississippi.

Cards

The invention of playing cards has been credited to the Indians, the Arabs, the Egyptians, and the Chinese. During the Crusades, Arabs endured lengthy sieges by



Brooks/Cole

A sixteenth-century 2 of swords.



Brooks/Cole

A sixteenth-century king of clubs.

playing card games. Their European foes acquired the cards and introduced them to their homelands. Cards, like dice, were used for divination as well as gambling. In fact, the modern deck is derived from the Tarot deck, which is composed of four suits plus twenty-two *atouts* that are not part of any suit. Each suit represents a class of medieval society: swords represent the nobility; coins, the merchants; batons or clubs, the peasants; and cups or chalices, the church. These suits are still used in regular playing cards in southern Europe. The Tarot deck also includes a joker and, in each suit, a king, a queen, a knight, a knave, and ten numbered cards.

Around 1500, the French dropped the knights and *atouts* from the deck and changed the suits from swords, coins, clubs, and chalices to *piques* (soldiers' pikes), *carreaux* (diamond-shaped building tiles), *trèfles* (clover leaf-shaped trefoils), and *coeurs* (hearts). In sixteenth-century Spain, *piques* were called *espados*, from which we get our term *spades*. Our diamonds are so named because of the shape of the *carreaux*. Clubs was an original Tarot suit, and hearts is a translation of *coeurs*.

The pictures on the cards were portraits of actual people. In fourteenth-century Europe, the kings were Charlemagne (hearts), the biblical David (spades), Julius Caesar (diamonds), and Alexander the Great (clubs); the queens included Helen of Troy (hearts), Pallas Athena (spades), and the biblical Rachel (diamonds). Others honored as “queen for a day” included Joan of Arc, Elizabeth I, and Elizabeth of York, wife of Henry VII. Jacks were usually famous warriors, including Sir Lancelot (clubs) and Roland, Charlemagne's nephew (diamonds).

A modern deck of cards contains fifty-two cards (thirteen in each of four suits). The four suits are hearts, diamonds, clubs, and spades (♥, ♦, ♣, ♠). Hearts and diamonds are red, and clubs and spades are black. Each suit consists of cards labeled 2 through 10, followed by jack, queen, king, and ace. **Face cards** are the jack, queen, and king; and **picture cards** are the jack, queen, king, and ace.

Two of the most popular card games are poker and blackjack. Poker's ancestor was a Persian game called *dsands*, which became popular in eighteenth-century Paris. It was transformed into a game called *poque*, which spread to America via the French colony in New Orleans. *Poker* is an American mispronunciation of the word *poque*.

The origins of blackjack (also known as *vingt-et-un* or twenty-one) are unknown. The game is called twenty-one because high odds are paid if a player's first two cards total 21 points (an ace counts as either 1 or 11, and the ten, jack, queen, and king each count as 10). A special bonus used to be paid if those two cards were a black jack and a black ace, hence the name *blackjack*.



© Kathleen Olson

A modern deck of cards.

3.1 EXERCISES

1. The Chevalier de Méré generally made money betting that he could roll at least one 6 in four rolls of a single die. Roll a single die four times, and record the number of times a 6 comes up. Repeat this ten times. If you had made the Chevalier de Méré's bet (at \$10 per game), would you have won or lost money? How much?
 2. The Chevalier de Méré generally lost money betting that he could roll at least one pair of 6's in twenty-four rolls of a pair of dice. Roll a pair of dice twenty-four times, and record the number of times a 6 comes up. Repeat this five times. If you had made the Chevalier de Méré's bet (at \$10 per game), would you have won or lost money? How much?
 3.
 - a. If you were to flip a pair of coins thirty times, approximately how many times do you think a pair of heads would come up? A pair of tails? One head and one tail?
 - b. Flip a pair of coins thirty times, and record the number of times a pair of heads comes up, the number of times a pair of tails comes up, and the number of times one head and one tail come up. How closely do the results agree with your guess?
 4.
 - a. If you were to flip a single coin twenty times, approximately how many times do you think heads would come up? Tails?
 - b. Flip a single coin twenty times, and record the number of times heads comes up and the number of times tails comes up. How closely do the results agree with your guesses?
 5.
 - a. If you were to roll a single die twenty times, approximately how many times do you think an even number would come up? An odd number?
 - b. Roll a single die twenty times, and record the number of times an even number comes up and the number of times an odd number comes up. How closely do the results agree with your guess?
 6.
 - a. If you were to roll a pair of dice thirty times, approximately how many times do you think the total would be 7? Approximately how many times do you think the total would be 12?
 - b. Roll a single die thirty times, and record the number of times the total is 7 and the number of times the total is 12. How closely do the results agree with your guess?
 7.
 - a. If you were to deal twenty-six cards from a complete deck (without jokers), approximately how many cards do you think would be red? Approximately how many do you think would be aces?
 - b. Deal twenty-six cards from a complete deck (without jokers), and record the number of times a red card is dealt and the number of times an ace is dealt. How closely do the results agree with your guess?
 8.
 - a. If you were to deal twenty-six cards from a complete deck (without jokers), approximately how many cards do you think would be black? Approximately how many do you think would be jacks, queens, or kings?
 - b. Deal twenty-six cards from a complete deck (without jokers), and record the number of times a black card is dealt and the number of times a jack, queen, or king is dealt. How closely do the results agree with your guess?
- In Exercises 9–24, use Figure 3.1 to find the outcome of the bets in roulette, given the results listed.*
- ▶ 9. You bet \$10 on the 25.
 - a. The ball lands on number 25.
 - b. The ball lands on number 14.
 10. You bet \$15 on the 17.
 - a. The ball lands on 18.
 - b. The ball lands on 17.
 11. You bet \$5 on 17-20 split.
 - a. The ball lands on number 17.
 - b. The ball lands on number 20.
 - c. The ball lands on number 32.
 12. You bet \$30 on the 22-23-24 street.
 - a. The ball lands on number 19.
 - b. The ball lands on number 22.
 - c. The ball lands on number 0.
 13. You bet \$20 on the 8-9-11-12 square.
 - a. The ball lands on number 15.
 - b. The ball lands on number 9.
 - c. The ball lands on number 00.
 - ▶ 14. You bet \$100 on the 0-00-1-2-3 line (the only five-number line on the table).
 - a. The ball lands on number 29.
 - b. The ball lands on number 2.
 15. You bet \$10 on the 31-32-33-34-35-36 line.
 - a. The ball lands on number 5.
 - b. The ball lands on number 33.
 - ▶ 16. You bet \$20 on the 13 through 24 section.
 - a. The ball lands on number 00.
 - b. The ball lands on number 15.

17. You bet \$25 on the first column.
 a. The ball lands on number 13.
 b. The ball lands on number 14.
18. You bet \$30 on the low numbers.
 a. The ball lands on number 8.
 b. The ball lands on number 30.
19. You bet \$50 on the odd numbers.
 a. The ball lands on number 00.
 b. The ball lands on number 5.
- ▶ 20. You bet \$20 on the black numbers.
 a. The ball lands on number 11.
 b. The ball lands on number 12.
21. You make a \$20 single-number bet on number 14 and also a \$25 single-number bet on number 15.
 a. The ball lands on number 16.
 b. The ball lands on number 15.
 c. The ball lands on number 14.
22. You bet \$10 on the low numbers and also bet \$20 on the 16-17-19-20 square.
 a. The ball lands on number 16.
 b. The ball lands on number 19.
 c. The ball lands on number 14.
- ▶ 23. You bet \$30 on the 1-2 split and also bet \$15 on the even numbers.
 a. The ball lands on number 1.
 b. The ball lands on number 2.
 c. The ball lands on number 3.
 d. The ball lands on number 4.
- ▶ 24. You bet \$40 on the 1-12 section and also bet \$10 on number 10.
 a. The ball lands on number 7.
 b. The ball lands on number 10.
 c. The ball lands on number 21.
25. How much must you bet on a single number to be able to win at least \$100? (Bets must be in \$1 increments.)
26. How much must you bet on a two-number split to be able to win at least \$200? (Bets must be in \$1 increments.)
- ▶ 27. How much must you bet on a twelve-number column to be able to win at least \$1000? (Bets must be in \$1 increments.)
- ▶ 28. How much must you bet on a four-number square to win at least \$600? (Bets must be in \$1 increments.)
- ▶ 29. a. How many hearts are there in a deck of cards?
 b. What fraction of a deck is hearts?
- ▶ 30. a. How many red cards are there in a deck of cards?
 b. What fraction of a deck is red?
31. a. How many face cards are there in a deck of cards?
 b. What fraction of a deck is face cards?
32. a. How many black cards are there in a deck of cards?
 b. What fraction of a deck is black?
33. a. How many kings are there in a deck of cards?
 b. What fraction of a deck is kings?



Answer the following questions using complete sentences and your own words.

• HISTORY QUESTIONS

34. Who started probability theory? How?
35. Why was probability theory not considered a serious branch of mathematics?
36. Which authors established probability theory as a serious area of interest? What are some of the areas to which these authors applied probability theory?
37. What did Gregor Mendel do with probabilities?
38. Who was Antoine Gombauld, and what was his role in probability theory?
39. Who was Gerolamo Cardano, and what was his role in probability theory?
40. Which games of chance came to America from France via New Orleans?
41. Which implements of gambling were also used for divination?
42. What is the oldest casino game still being played?
43. How were cards introduced to Europe?
44. What is the modern deck of cards derived from?
45. What game was supposed to take the advantage away from dishonest gamblers?

WEB PROJECT

46. Write a research paper on any historical topic referred to in this section or a related topic. Following is a partial list of topics:
- Jacob Bernoulli
 - Gerolamo Cardano
 - Pierre de Fermat
 - Blaise Pascal
 - The Marquis de Laplace
 - John Graunt
 - Gregor Mendel

Some useful links for this web project are listed on the text web site: www.cengage.com/math/johnson

3.2

Basic Terms of Probability

OBJECTIVES

- Learn the basic terminology of probability theory
- Be able to calculate simple probabilities
- Understand how probabilities are used in genetics

Much of the terminology and many of the computations of probability theory have their basis in set theory, because set theory contains the mathematical way of describing collections of objects and the size of those collections.

BASIC PROBABILITY TERMS

experiment: a process by which an observation, or **outcome**, is obtained

sample space: the set S of all possible outcomes of an experiment

event: any subset E of the sample space S

If a single die is rolled, the *experiment* is the rolling of the die. The possible *outcomes* are 1, 2, 3, 4, 5, and 6. The *sample space* (set of all possible outcomes) is $S = \{1, 2, 3, 4, 5, 6\}$. (The term *sample space* really means the same thing as *universal set*; the only distinction between the two ideas is that *sample space* is used only in probability theory, while *universal set* is used in any situation in which sets are used.) There are several possible *events* (subsets of the sample space), including the following:

$$E_1 = \{3\} \quad \text{“a three comes up”}$$

$$E_2 = \{2, 4, 6\} \quad \text{“an even number comes up”}$$

$$E_3 = \{1, 2, 3, 4, 5, 6\} \quad \text{“a number between 1 and 6 inclusive comes up”}$$

Notice that an event is not the same as an outcome. An *event* is a subset of the sample space; an *outcome* is an element of the sample space. “Rolling an odd number” is an event, not an outcome. It is the set $\{1, 3, 5\}$ that is composed of three separate outcomes. Some events are distinguished from outcomes only in that set brackets are used with events and not with outcomes. For example, $\{5\}$ is an event, and 5 is an outcome; either refers to “rolling a five.”

The event E_3 (“a number between 1 and 6 inclusive comes up”) is called a *certain event*, since $E_3 = S$. That is, E_3 is a sure thing. “Getting 17” is an *impossible event*. No outcome in the sample space $S = \{1, 2, 3, 4, 5, 6\}$ would result in a 17, so this event is actually the empty set.

MORE PROBABILITY TERMS

A **certain event** is an event that is equal to the sample space.

An **impossible event** is an event that is equal to the empty set.

THE FAR SIDE® BY GARY LARSON



The Far Side® by Gary Larson © 1990 FarWorks, Inc. All Rights Reserved. The Far Side® and the Larson® signature are registered trademarks of FarWorks, Inc. Used with permission.

Early shell games

An early certain event.

Finding Probabilities and Odds

The **probability** of an event is a measure of the likelihood that the event will occur. If a single die is rolled, the outcomes are equally likely; a 3 is just as likely to come up as any other number. There are six possible outcomes, so a 3 should come up about one out of every six rolls. That is, the probability of event E_1 (“a 3 comes up”) is $\frac{1}{6}$. The 1 in the numerator is the number of elements in $E_1 = \{3\}$. The 6 in the denominator is the number of elements in $S = \{1, 2, 3, 4, 5, 6\}$.

If an experiment’s outcomes are equally likely, then the probability of an event E is the number of outcomes in the event divided by the number of outcomes in the sample space, or $n(E)/n(S)$. (In this chapter, we discuss only experiments with equally likely outcomes.) Probability can be thought of as “success over a total.”

PROBABILITY OF AN EVENT

The **probability** of an event E , denoted by $p(E)$, is

$$p(E) = \frac{n(E)}{n(S)}$$

if the experiment’s outcomes are equally likely.
(Think: Success over total.)

Many people use the words *probability* and *odds* interchangeably. However, the words have different meanings. The **odds** in favor of an event are the number of ways the event can occur compared to the number of ways the event *can fail to occur*, or “success compared to *failure*” (if the experiment’s outcomes are equally likely). The odds of event E_1 (“a 3 comes up”) are 1 to 5 (or 1:5), since a three can come up in one way and can fail to come up in five ways. Similarly, the odds of event E_3 (“a number between 1 and 6 inclusive comes up”) are 6 to 0 (or 6:0), since a number between 1 and 6 inclusive can come up in six ways and can fail to come up in zero ways.

ODDS OF AN EVENT

The **odds** of an event E with equally likely outcomes, denoted by $o(E)$, are given by

$$o(E) = n(E) : n(E')$$

(Think: *Success compared with failure.*)

In addition to the above meaning, the word *odds* can also refer to “house odds,” which has to do with how much you will be paid if you win a bet at a casino. The odds of an event are sometimes called the **true odds** to distinguish them from the house odds.

EXAMPLE 1

FLIPPING A COIN A coin is flipped. Find the following.

- the sample space
- the probability of event E_1 , “getting heads”
- the odds of event E_1 , “getting heads”
- the probability of event E_2 , “getting heads or tails”
- the odds of event E_2 , “getting heads or tails”

SOLUTION

a. Finding the sample space S : The experiment is flipping a coin. The only possible outcomes are heads and tails. The sample space S is the set of all possible outcomes, so $S = \{h, t\}$.

b. Finding the probability of heads:

$$E_1 = \{h\} \text{ (“getting heads”)}$$

$$p(E_1) = \frac{n(E_1)}{n(S)} = \frac{1}{2}$$

This means that one out of every two possible outcomes is a success.

c. Finding the odds of heads:

$$E_1' = \{t\}$$

$$o(E_1) = n(E_1) : n(E_1') = 1 : 1$$

This means that for every one possible success, there is one possible failure.

d. Finding the probability of heads or tails:

$$E_2 = \{h, t\}$$

$$p(E_2) = \frac{n(E_2)}{n(S)} = \frac{2}{2} = \frac{1}{1}$$

This means that every outcome is a success. Notice that E_2 is a certain event.

e. Finding the odds of heads or tails:

$$E_2' = \emptyset$$

$$o(E_2) = n(E_2) : n(E_2') = 2 : 0 = 1 : 0$$

This means that there are no possible failures.

EXAMPLE 2

ROLLING A DIE A die is rolled. Find the following.

- the sample space
- the event “rolling a 5”
- the probability of rolling a 5
- the odds of rolling a 5
- the probability of rolling a number below 5
- the odds of rolling a number below 5

SOLUTION

- The sample space is $S = \{1, 2, 3, 4, 5, 6\}$.
- Event E_1 “rolling a 5” is $E_1 = \{5\}$.
- The probability of E_1 is

$$p(E_1) = \frac{n(E_1)}{n(S)} = \frac{1}{6}$$

This means that one out of every six possible outcomes is a success (that is, a 5).

- $E_1' = \{1, 2, 3, 4, 6\}$. The odds of E_1 are

$$o(E_1) = n(E_1):n(E_1') = 1:5$$

This means that there is one possible success for every five possible failures.

- $E_2 = \{1, 2, 3, 4\}$ (“rolling a number below 5”). The probability of E_2 is

$$p(E_2) = \frac{n(E_2)}{n(S)} = \frac{4}{6} = \frac{2}{3}$$

This means that two out of every three possible outcomes are a success.

- $E_2' = \{5, 6\}$. The odds of E_2 are

$$o(E_2) = n(E_2):n(E_2') = 4:2 = 2:1$$

This means that there are two possible successes for every one possible failure. Notice that odds are reduced in the same manner that a fraction is reduced.

Relative Frequency versus Probability

So far, we have discussed probabilities only in a theoretical way. When we found that the probability of heads was $\frac{1}{2}$, we never actually tossed a coin. It does not always make sense to calculate probabilities theoretically; sometimes they must be found empirically, the way a batting average is calculated. For example, in 8,389 times at bat, Babe Ruth had 2,875 hits. His batting average was $\frac{2,875}{8,389} \approx 0.343$. In other words, his probability of getting a hit was 0.343.

Sometimes a probability can be found either theoretically or empirically. We have already found that the theoretical probability of heads is $\frac{1}{2}$. We could also flip a coin a number of times and calculate (number of heads)/(total number of flips); this can be called the **relative frequency** of heads, to distinguish it from the theoretical probability of heads.

The Law of Large Numbers

Usually, the relative frequency of an outcome is not equal to its probability, but if the number of trials is large, the two tend to be close. If you tossed a coin a couple of times, anything could happen, and the fact that the probability of heads is

$\frac{1}{2}$ would have no impact on the results. However, if you tossed a coin 100 times, you would probably find that the relative frequency of heads was close to $\frac{1}{2}$. If your friend tossed a coin 1,000 times, she would probably find the relative frequency of heads to be even closer to $\frac{1}{2}$ than in your experiment. This relationship between probabilities and relative frequencies is called the **Law of Large Numbers**. Cardano stated this law in his *Book on Games of Chance*, and Bernoulli proved that it must be true.

LAW OF LARGE NUMBERS

If an experiment is repeated a large number of times, the relative frequency of an outcome will tend to be close to the probability of that outcome.

The graph in Figure 3.2 shows the result of a simulated coin toss, using a computer and a random number generator rather than an actual coin. Notice that when the number of tosses is small, the relative frequency achieves values such as 0, 0.67, and 0.71. These values are not that close to the theoretical probability of 0.5. However, when the number of tosses is large, the relative frequency achieves values such as 0.48. These values are very close to the theoretical probability.

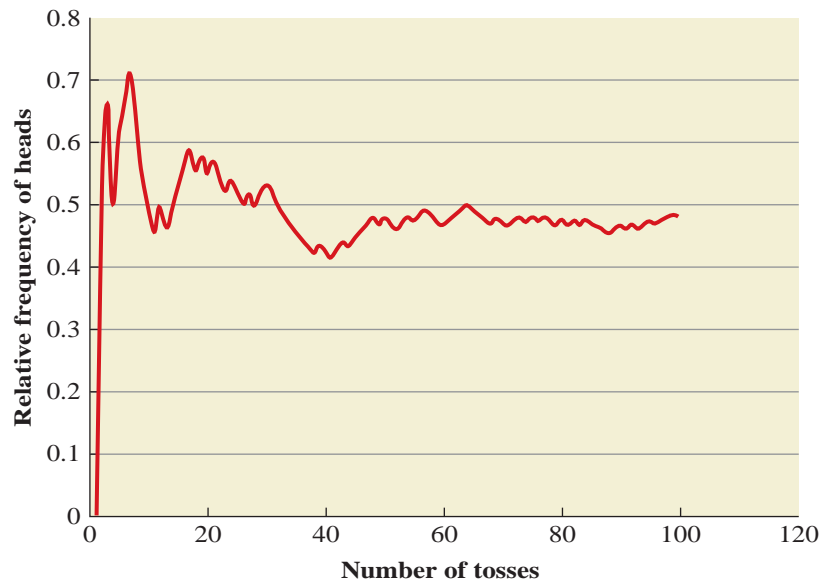


FIGURE 3.2 The relative frequency of heads after 100 simulated coin tosses.

What if we used a real coin, rather than a computer, and we tossed the coin a lot more? Three different mathematicians have performed such an experiment:

- In the eighteenth century, Count Buffon tossed a coin 4,040 times. He obtained 2,048 heads, for a relative frequency of $2,048/4,040 \approx 0.5069$.
- During World War II, the South African mathematician John Kerrich tossed a coin 10,000 times while he was imprisoned in a German concentration camp. He obtained 5,067 heads, for a relative frequency of $5,067/10,000 = 0.5067$.
- In the early twentieth century, the English mathematician Karl Pearson tossed a coin 24,000 times! He obtained 12,012 heads, for a relative frequency of $12,012/24,000 = 0.5005$.

At a casino, probabilities are much more useful to the casino (the “house”) than to an individual gambler, because the house performs the experiment for a

much larger number of trials (in other words, plays the game more often). In fact, the house plays the game so many times that the relative frequencies will be almost exactly the same as the probabilities, with the result that the house is not gambling at all—it knows what is going to happen. Similarly, a gambler with a “system” has to play the game for a long time for the system to be of use.

EXAMPLE 3

FLIPPING A PAIR OF COINS A pair of coins is flipped.

- Find the sample space.
- Find the event E “getting exactly one heads.”
- Find the probability of event E .
- Use the Law of Large Numbers to interpret the probability of event E .

SOLUTION



The outcomes (h, h) , (h, t) , (t, h) , and (t, t) .

- The experiment is the flipping of a pair of coins. One possible outcome is that one coin is heads and the other is tails. A second and *different* outcome is that one coin is tails and the other is heads. These two outcomes seem the same. However, if one coin were painted, it would be easy to tell them apart. Outcomes of the experiment can be described by using ordered pairs in which the first component refers to the first coin and the second component refers to the second coin. The two different ways of getting one heads and one tails are (h, t) and (t, h) .

The sample space, or set of all possible outcomes, is the set $S = \{(h, h), (t, t), (h, t), (t, h)\}$. These outcomes are equally likely.

- The event E “getting exactly one heads” is

$$E = \{(h, t), (t, h)\}$$

- The probability of event E is

$$P(E) = \frac{n(E)}{n(S)} = \frac{2}{4} = \frac{1}{2} = 50\%$$

- According to the Law of Large Numbers, if an experiment is repeated a large number of times, the relative frequency of that outcome will tend to be close to the probability of the outcome. Here, that means that if we were to toss a pair of coins many times, we should expect to get exactly one heads about half (or 50%) of the time. Realize that this is only a prediction, and we might never get exactly one heads.

Mendel's Use of Probabilities

In his experiments with plants, Gregor Mendel pollinated peas until he produced pure-red plants (that is, plants that would produce only red-flowered offspring) and pure-white plants. He then cross-fertilized these pure reds and pure whites and obtained offspring that had only red flowers. This amazed him, because the accepted theory of the day incorrectly predicted that these offspring would all have pink flowers.

He explained this result by postulating that there is a “determiner” that is responsible for flower color. These determiners are now called **genes**. Each plant has two flower color genes, one from each parent. Mendel reasoned that these offspring had to have inherited a red gene from one parent and a white gene from the other. These plants had one red flower gene and one white flower gene, but they had red flowers. That is, the red flower gene is **dominant**, and the white flower gene is **recessive**.

If we use R to stand for the red gene and w to stand for the white gene (with the capital letter indicating dominance), then the results can be described with a **Punnett square** in Figure 3.3 on page 146.

	R	R	← first parent's genes
w	(R, w)	(R, w)	← possible offspring
w	(R, w)	(R, w)	← possible offspring

↑
second parent's genes

FIGURE 3.3 A Punnett square for the first generation.

When the offspring of this experiment were cross-fertilized, Mendel found that approximately three-fourths of the offspring had red flowers and one-fourth had white flowers (see Figure 3.4). Mendel successfully used probability theory to analyze this result.

	R	w	← first parent's genes
R	(R, R)	(w, R)	← possible offspring
w	(R, w)	(w, w)	← possible offspring

↑
second parent's genes

FIGURE 3.4 A Punnett square for the second generation.

Only one of the four possible outcomes, (w, w), results in a white-flowered plant, so event E_1 that the plant has white flowers is $E_1 = \{(w, w)\}$; therefore,

$$p(E_1) = \frac{n(E_1)}{n(S)} = \frac{1}{4}$$

This means that we should expect the actual relative frequency of white-flowered plants to be close to $\frac{1}{4}$. That is, about one-fourth, or 25%, of the second-generation plants should have white flowers (if we have a lot of plants).

Each of the other three outcomes, (R, R), (R, w), and (w, R), results in a red-flowered plant, because red dominates white. The event E_2 that the plant has red flowers is $E_2 = \{(R, R), (R, w), (w, R)\}$; therefore,

$$p(E_2) = \frac{n(E_2)}{n(S)} = \frac{3}{4} = 75\%$$

Thus, we should expect the actual relative frequency of red-flowered plants to be close to $\frac{3}{4}$. That is, about three-fourths, or 75%, of the plants should have red flowers.

Outcomes (R, w) and (w, R) are genetically identical; it does not matter which gene is inherited from which parent. For this reason, geneticists do not use the ordered-pair notation but instead refer to each of these two outcomes as “Rw.” The only difficulty with this convention is that it makes the sample space appear to be $S = \{RR, Rw, ww\}$, which consists of only three elements, when in fact it consists of four elements. This distinction is important; if the sample space consisted of three equally likely elements, then the probability of a red-flowered offspring would be $\frac{2}{3}$ rather than $\frac{3}{4}$. Mendel knew that the sample space had to have four elements, because his cross-fertilization experiments resulted in a relative frequency very close to $\frac{3}{4}$, not $\frac{2}{3}$.

Ronald Fisher, a noted British statistician, used statistics to deduce that Mendel fudged his data. Mendel's relative frequencies were unusually close to the theoretical probabilities, and Fisher found that there was only about a 0.00007 chance of such close agreement. Others have suggested that perhaps Mendel did not willfully change his results but rather continued collecting data until the numbers were in close agreement with his expectations.*

*R. A. Fisher, “Has Mendel's Work Been Rediscovered?” *Annals of Science* 1, 1936, pp. 115–137.

HISTORICAL NOTE

GREGOR JOHANN MENDEL, 1822–1884

Johann Mendel was born to an Austrian peasant family. His interest in botany began on the family farm, where he helped his father graft fruit trees. He studied philosophy, physics, and mathematics at the University Philosophical Institute in Olmütz. He was unsuccessful in finding a job, so he quit school and returned to the farm. Depressed by the prospects of a bleak future, he became ill and stayed at home for a year.

Mendel later returned to Olmütz. After two years of study, he found the pressures of school and work to be too much, and his health again broke down. On the advice of his father and a



professor, he entered the priesthood, even though he did not feel called to serve the church. His name was changed from Johann to Gregor. Relieved of his financial difficulties, he was able to continue his studies. However, his nervous disposition interfered with his pastoral duties, and he was assigned to substitute teaching. He enjoyed this work and was popular with the staff and students, but he failed the examination for certification as a teacher. Ironically, his lowest grades were in biology. The Augustinians then sent him to the University of Vienna, where he became particularly interested in his plant physiology professor's unorthodox belief that new plant varieties can be caused by naturally arising variations. He was also fascinated by his classes in physics, where he was exposed to the physicists' experimental and mathematical approach to their subject.

After further breakdowns and failures, Mendel returned to the monastery and was assigned the low-stress job of keeping the abbey garden. There he combined the experimental and mathematical approach of a physicist with his background in biology and performed a series of experiments designed to determine whether his professor was correct in his beliefs

regarding the role of naturally arising variants in plants.

Mendel studied the transmission of specific traits of the pea plant—such as flower color and stem length—from parent plant to offspring. He pollinated the plants by hand and separated them until he had isolated each trait. For example, in his studies of flower color, he pollinated the plants until he produced pure-red plants (plants that would produce only red-flowered offspring) and pure-white plants.

At the time, the accepted theory of heredity was that of blending. In this view, the characteristics of both parents blend together to form an individual. Mendel reasoned that if the blending theory was correct, the union of a pure-red pea plant and a pure-white pea plant would result in a pink-flowered offspring. However, his experiments showed that such a union consistently resulted in red-flowered offspring.

Mendel crossbred a large number of peas that had different characteristics. In many cases, an offspring would have a characteristic of one of its parents, undiluted by that of the other parent. Mendel concluded that the question of which parent's characteristics would be passed on was a matter of chance, and he successfully used probability theory to estimate the frequency with which characteristics would be passed on. In so doing, Mendel founded modern genetics. Mendel attempted similar experiments with bees, but these experiments were unsuccessful because he was unable to control the mating behavior of the queen bee.

Mendel was ignored when he published his paper "Experimentation in Plant Hybridization." Sixteen years after his death, his work was rediscovered by three European botanists who had reached similar conclusions in plant breeding, and the importance of Mendel's work was finally recognized.



A nineteenth-century drawing illustrating Mendel's pea plants, showing the original cross, the first generation, and the second generation.

Probabilities in Genetics: Inherited Diseases

Cystic fibrosis is an inherited disease characterized by abnormally functioning exocrine glands that secrete a thick mucus, clogging the pancreatic ducts and lung passages. Most patients with cystic fibrosis die of chronic lung disease; until recently, most died in early childhood. This early death made it extremely unlikely that an afflicted person would ever parent a child. Only after the advent of Mendelian genetics did it become clear how a child could inherit the disease from two healthy parents.

In 1989, a team of Canadian and American doctors announced the discovery of the gene that is responsible for most cases of cystic fibrosis. As a result of that discovery, a new therapy for cystic fibrosis is being developed. Researchers splice a therapeutic gene into a cold virus and administer it through an affected person's nose. When the virus infects the lungs, the gene becomes active. It is hoped that this will result in normally functioning cells, without the damaging mucus.

In April 1993, a twenty-three-year-old man with advanced cystic fibrosis became the first patient to receive this therapy. In September 1996, a British team announced that eight volunteers with cystic fibrosis had received this therapy; six were temporarily cured of the disease's debilitating symptoms. In March 1999, another British team announced a new therapy that involves administering the therapeutic gene through an aerosol spray. Thanks to other advances in treatment, in 2006, infants born in the United States with cystic fibrosis had a life expectancy of twenty-seven years.

Cystic fibrosis occurs in about 1 out of every 2,000 births in the Caucasian population and only in about 1 in 250,000 births in the non-Caucasian population. It is one of the most common inherited diseases in North America. One in 25 Americans carries a single gene for cystic fibrosis. Children who inherit two such genes develop the disease; that is, cystic fibrosis is recessive.

There are tests that can be used to determine whether a person carries the gene. However, they are not accurate enough to use for the general population. They are much more accurate with people who have a family history of cystic fibrosis, so The American College of Obstetricians and Gynecologists recommends testing only for couples with a personal or close family history of cystic fibrosis.

EXAMPLE 4

PROBABILITIES AND CYSTIC FIBROSIS Each of two prospective parents carries one cystic fibrosis gene.

- Find the probability that their child would have cystic fibrosis.
- Find the probability that their child would be free of symptoms.
- Find the probability that their child would be free of symptoms but could pass the cystic fibrosis gene on to his or her own child.

SOLUTION

We will denote the recessive cystic fibrosis gene with the letter c and the normal disease-free gene with an N . Each parent is Nc and therefore does not have the disease. Figure 3.5 shows the Punnett square for the child.

	N	c
N	(N, N)	(c, N)
c	(N, c)	(c, c)

FIGURE 3.5 A Punnett square for Example 4.

- a. Cystic fibrosis is recessive, so only the (c, c) child will have the disease. The probability of such an event is $1/4$.
- b. The (N, N), (c, N), and (N, c) children will be free of symptoms. The probability of this event is

$$\begin{aligned} p(\text{healthy}) &= p((N, N)) + p((c, N)) + p((N, c)) \\ &= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4} \end{aligned}$$

- c. The (c, N) and (N, c) children would never suffer from any symptoms but could pass the cystic fibrosis gene on to their own children. The probability of this event is

$$p((c, N)) + p((N, c)) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

In Example 4, the Nc child is called a **carrier** because that child would never suffer from any symptoms but could pass the cystic fibrosis gene on to his or her own child. Both of the parents were carriers.

Sickle-cell anemia is an inherited disease characterized by a tendency of the red blood cells to become distorted and deprived of oxygen. Although it varies in severity, the disease can be fatal in early childhood. More often, patients have a shortened life span and chronic organ damage.

Newborns are now routinely screened for sickle-cell disease. The only true cure is a bone marrow transplant from a sibling without sickle-cell anemia; however, this can cause the patient's death, so it is done only under certain circumstances. Until recently, about 10% of the children with sickle-cell anemia had a stroke before they were twenty-one. But in 2009, it was announced that the rate of these strokes has been cut in half thanks to a new specialized ultrasound scan that identifies the individuals who have a high stroke risk. There are also medications that can decrease the episodes of pain.

Approximately 1 in every 500 black babies is born with sickle-cell anemia, but only 1 in 160,000 nonblack babies has the disease. This disease is **codominant**: A person with two sickle-cell genes will have the disease, while a person with one sickle-cell gene will have a mild, nonfatal anemia called **sickle-cell trait**. Approximately 8–10% of the black population has sickle-cell trait.

Huntington's disease, caused by a dominant gene, is characterized by nerve degeneration causing spasmodic movements and progressive mental deterioration. The symptoms do not usually appear until well after reproductive age has been reached; the disease usually hits people in their forties. Death typically follows 20 years after the onset of the symptoms. No effective treatment is available, but physicians can now assess with certainty whether someone will develop the disease, and they can estimate when the disease will strike. Many of those who are at risk choose not to undergo the test, especially if they have already had children. Folk singer Arlo Guthrie is in this situation; his father, Woody Guthrie, died of Huntington's disease. Woody's wife Marjorie formed the Committee to Combat Huntington's Disease, which has stimulated research, increased public awareness, and provided support for families in many countries.

In August 1999, researchers in Britain, Germany, and the United States discovered what causes brain cells to die in people with Huntington's disease. This discovery may eventually lead to a treatment. In 2008, a new drug that reduces the uncontrollable spasmodic movements was approved.

Tay-Sachs disease is a recessive disease characterized by an abnormal accumulation of certain fat compounds in the spinal cord and brain. Most typically,

HISTORICAL NOTE

NANCY WEXLER

In 1993, scientists working together at six major research centers located most genes that cause Huntington's disease. This discovery will enable people to learn whether they carry a Huntington's gene, and it will allow pregnant women to determine whether their child carries the gene. The discovery could eventually lead to a treatment.

The collaboration of research centers was organized largely by Nancy Wexler, a Columbia University professor of neuropsychology, who is herself at risk for Huntington's disease—her mother died of it in 1978. Dr. Wexler, President of the Hereditary Disease Foundation, has made numerous trips to study and



Courtesy of Dr. Nancy Wexler

aid the people of the Venezuelan village of Lake Maracaibo, many of whom suffer from the disease or are at risk for it. All are related to one woman who died of the disease in the early 1800s. Wexler took blood and tissue samples and gave neurological and psychoneurological tests to the inhabitants of the village. The samples and test results enabled the researchers to find the single gene that causes Huntington's disease.

In October 1993, Wexler received an Albert Lasker Medical Research Award, a prestigious honor that is often a precursor to a Nobel Prize. The award was given in recognition for her contribution to the international effort

that culminated in the discovery of the Huntington's disease gene. At the awards ceremony, she explained to then first lady Hillary Clinton that her genetic heritage has made her uninsurable—she would lose her health coverage if she switched jobs. She told Mrs. Clinton that more Americans will be in the same situation as more genetic discoveries are made, unless the health care system is reformed. The then first lady incorporated this information into her speech at the awards ceremony: "It is likely that in the next years, every one of us will have a pre-existing condition and will be uninsurable. . . . What will happen as we discover those genes for breast cancer, or prostate cancer, or osteoporosis, or any of the thousands of other conditions that affect us as human beings?"



AP Photo

Woody Guthrie's most famous song is "This Land is Your Land." This folksinger, guitarist, and composer was a friend of Leadbelly, Pete Seeger, and Ramblin' Jack Elliott and exerted a strong influence on Bob Dylan. Guthrie died at the age of fifty-five of Huntington's disease.

a child with Tay-Sachs disease starts to deteriorate mentally and physically at six months of age. After becoming blind, deaf, and unable to swallow, the child becomes paralyzed and dies before the age of four years. There is no effective treatment. The disease occurs once in 3,600 births among Ashkenazi Jews (Jews from central and eastern Europe), Cajuns, and French Canadians but only once in 600,000 births in other populations. Carrier-detection tests and fetal-monitoring tests are available. The successful use of these tests, combined with an aggressive counseling program, has resulted in a decrease of 90% of the incidence of this disease.

Genetic Screening

At this time, there are no conclusive tests that will tell a parent whether he or she is a cystic fibrosis carrier, nor are there conclusive tests that will tell whether a fetus has the disease. A new test resulted from the 1989 discovery of the location of most cystic fibrosis genes, but that test will detect only 85% to 95% of the cystic fibrosis genes, depending on the individual's ethnic background. The extent to which this test will be used has created quite a controversy.

Individuals who have relatives with cystic fibrosis are routinely informed about the availability of the new test. The controversial question is whether a massive genetic screening program should be instituted to identify cystic fibrosis carriers in the general population, regardless of family history. This is an important question, considering that four in five babies with cystic fibrosis are born to couples with no previous family history of the condition.

Opponents of routine screening cite a number of important concerns. The existing test is rather inaccurate; 5% to 15% of the cystic fibrosis carriers would be missed. It is not known how health insurers would use this genetic information—insurance firms could raise rates or refuse to carry people if a screening test indicated a presence of cystic fibrosis. Also, some experts question the adequacy of quality assurance for the diagnostic facilities and for the tests themselves.

Supporters of routine testing say that the individual should be allowed to decide whether to be screened. Failing to inform people denies them the opportunity to make a personal choice about their reproductive future. An individual who is found to be a carrier could choose to avoid conception, to adopt, to use artificial insemination by a donor, or to use prenatal testing to determine whether a fetus is affected—at which point the additional controversy regarding abortion could enter the picture.

The Failures of Genetic Screening

The history of genetic screening programs is not an impressive one. In the 1970s, mass screening of blacks for sickle-cell anemia was instituted. This program caused unwarranted panic; those who were told they had sickle-cell trait feared that they would develop symptoms of the disease and often did not understand the probability that their children would inherit the disease (see Exercises 73 and 74). Some people with sickle-cell trait were denied health insurance and life insurance. See the 1973 Newsweek article on “The Row over Sickle-Cell.”

Text not available due to copyright restrictions

3.2 EXERCISES

In Exercises 1–14, use this information: A jar on your desk contains twelve black, eight red, ten yellow, and five green jellybeans. You pick a jellybean without looking.

1. What is the experiment?
2. What is the sample space?

In Exercises 3–14, find the following. Write each probability as a reduced fraction and as a percent, rounded to the nearest 1%.

3. the probability that it is black
4. the probability that it is green
- ▶ 5. the probability that it is red or yellow
6. the probability that it is red or black
7. the probability that it is not yellow
- ▶ 8. the probability that it is not red
9. the probability that it is white
10. the probability that it is not white
- ▶ 11. the odds of picking a black jellybean
- ▶ 12. the odds of picking a green jellybean
13. the odds of picking a red or yellow jellybean
- ▶ 14. the odds of picking a red or black jellybean

In Exercises 15–28, one card is drawn from a well-shuffled deck of fifty-two cards (no jokers).

15. What is the experiment?
16. What is the sample space?

In Exercises 17–28, (a) find the probability and (b) the odds of drawing the given cards. Also, (c) use the Law of Large Numbers to interpret both the probability and the odds. (You might want to review the makeup of a deck of cards in Section 3.1.)

- ▶ 17. a black card
18. a heart
19. a queen
- ▶ 20. a 2 of clubs
21. a queen of spades
22. a club
- ▶ 23. a card below a 5 (count an ace as high)
24. a card below a 9 (count an ace as high)
25. a card above a 4 (count an ace as high)
26. a card above an 8 (count an ace as high)
27. a face card
- ▶ 28. a picture card

▶ Selected exercises available online at www.webassign.net/brookscole

Age (years)	0–4	5–19	20–44	45–64	65–84	85+	total
Male	10,748	31,549	53,060	38,103	14,601	1,864	149,925
Female	10,258	30,085	51,432	39,955	18,547	3,858	154,135

FIGURE 3.6 2008 U.S. population, in thousands, by age and gender. *Source:* 2008 Census, U.S. Bureau of the Census.

Age	White	Black	American Indian, Eskimo, Aleut	Asian, Pacific Islander
0–4	15,041	2,907	222	1,042
5–19	47,556	9,445	709	3,036
20–44	79,593	14,047	1,005	5,165
45–64	60,139	8,034	497	2,911
65–84	27,426	2,826	166	987
85+	4,467	362	29	110

FIGURE 3.7 2005 U.S. population (projected), in thousands, by age and race. *Source:* Annual Population Estimates by Age Group and Sex, U.S. Bureau of the Census

In Exercises 29–38, (a) find the probability and (b) the odds of winning the given bet in roulette. Also, (c) use the Law of Large Numbers to interpret both the probability and the odds. (You might want to review the description of the game in Section 3.1.)

- 29. the single-number bet
- 30. the two-number bet
- 31. the three-number bet
- 32. the four-number bet
- 33. the five-number bet
- 34. the six-number bet
- 35. the twelve-number bet
- 36. the low-number bet
- ▶ 37. the even-number bet
- ▶ 38. the red-number bet
- ▶ 39. Use the information in Figure 3.6 from the U.S. Census Bureau to answer the following questions.
 - a. Find the probability that in the year 2008, a U.S. resident was female.
 - b. Find the probability that in the year 2008, a U.S. resident was male and between 20 and 44 years of age, inclusive.
- ▶ 40. Use the information in Figure 3.7 from the U.S. Census Bureau to answer the following questions.
 - a. Find the probability that in the year 2005, a U.S. resident was American Indian, Eskimo, or Aleut.
 - b. Find the probability that in the year 2005, a U.S. resident was Asian or Pacific Islander and between 20 and 44 years of age.
- ▶ 41. The dartboard in Figure 3.8 is composed of circles with radii 1 inch, 3 inches, and 5 inches.
 - a. What is the probability that a dart hits the red region if the dart hits the target randomly?
 - b. What is the probability that a dart hits the yellow region if the dart hits the target randomly?
 - c. What is the probability that a dart hits the green region if the dart hits the target randomly?

- d. Use the Law of Large Numbers to interpret the probabilities in parts (a), (b), and (c).

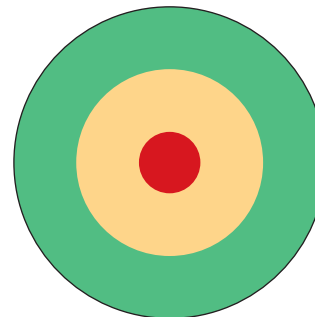


FIGURE 3.8 A dartboard for Exercise 41.

- ▶ 42.
 - a. What is the probability of getting red on the spinner shown in Figure 3.9?
 - b. What is the probability of getting blue?
 - c. Use the Law of Large Numbers to interpret the probabilities in parts (a) and (b).

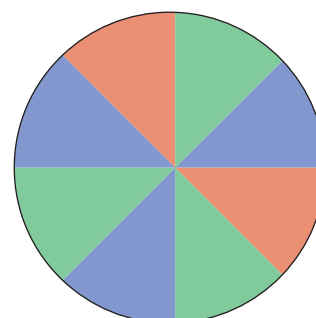


FIGURE 3.9 All sectors are equal in size and shape.

43. Amtrack’s *Empire Service* train starts in Albany, New York, at 6:00 A.M., Mondays through Fridays. It arrives at New York City at 8:25 A.M. If the train breaks down, find the probability that it breaks down in the first half hour of its run. Assume that the train breaks down at random times.
- ▶ 44. Amtrack’s *Downeaster* train leaves Biddeford, Maine, at 6:42 A.M., Mondays through Fridays. It arrives at Boston at 8:50 A.M. If the train breaks down, find the probability that it breaks down in the last 45 minutes of its run. Assume that the train breaks down at random times.
- ▶ 45. If $p(E) = \frac{1}{5}$, find $o(E)$. 46. If $p(E) = \frac{8}{9}$, find $o(E)$.
- ▶ 47. If $o(E) = 3:2$, find $p(E)$. ▶ 48. If $o(E) = 4:7$, find $p(E)$.
- ▶ 49. If $p(E) = \frac{a}{b}$, find $o(E)$.
- In Exercises 50–52, find (a) the probability and (b) the odds of winning the following bets in roulette. In finding the odds, use the formula developed in Exercise 49. Also, (c) use the Law of Large Numbers to interpret both the probability and the odds.*
50. the high-number bet 51. the odd-number bet
- ▶ 52. the black-number bet
53. In June 2009, sportsbook.com gave odds on who would win the 2009 World Series. They gave the New York Yankees 5:2 odds and the New York Mets 7:1 odds.
- a. Are these house odds or true odds? Why?
- b. Convert the odds to probabilities.
- c. Who did they think was more likely to win the World Series: the Yankees or the Mets? Why?
- ▶ 54. In June 2009, linesmaker.com gave odds on who would win the 2009 World Series. They gave the New York Yankees 9:2 odds and the New York Mets 15:1 odds.
- a. Are these house odds or true odds? Why?
- b. Convert the odds to probabilities.
- c. Who did they think was more likely to win the World Series: the Yankees or the Mets? Why?
- d. Use the information in Exercise 53 to determine which source thought it was more probable that the Yankees would win the World Series: sportsbook.com or linesmaker.com.
- In Exercises 55–60, use Figure 3.10.*
- ▶ 55. a. Of the specific causes listed, which is the most likely cause of death in one year?
- b. Which is the most likely cause of death in a lifetime? Justify your answers.
56. a. Of the causes listed, which is the least likely cause of death in one year?

Type of Accident or Injury	Odds of Dying in 1 Year	Lifetime Odds of Dying
All transportation accidents	1:6,121	1:79
Pedestrian transportation accident	1:48,816	1:627
Bicyclist transportation accident	1:319,817	1:4,111
Motorcyclist transportation accident	1:67,588	1:869
Car transportation accident	1:20,331	1:261
Airplane and space transportation accident	1:502,554	1:6,460
All nontransportation accidents	1:4,274	1:55
Falling	1:15,085	1:194
Drowning	1:82,777	1:1,064
Fire	1:92,745	1:1,192
Venomous animals and poisonous plants	1:2,823,877	1:36,297
Lightning	1:6,177,230	1:79,399
Earthquake	1:8,013,704	1:103,004
Storm	1:339,253	1:4,361
Intentional self-harm	1:9,085	1:117
Assault	1:16,360	1:210
Legal execution	1:5,490,872	1:70,577
War	1:10,981,743	1:141,154
Complications of medical care	1:111,763	1:1,437

FIGURE 3.10 The odds of dying due to an accident or injury in the United States in 2005. *Source: The odds of dying from . . . , 2009 edition, www.nsc.org.*

- b. Which is the least likely cause of death in a lifetime? Justify your answers.
57. a. What is the probability that a person will die of a car transportation accident in one year?
b. What is the probability that a person will die of an airplane transportation accident in one year?
Write your answers as fractions.
- ▶ 58. a. What is the probability that a person will die from lightning in a lifetime?
b. What is the probability that a person will die from an earthquake in a lifetime?
Write your answers as fractions.
59. a. Of all of the specific forms of transportation accidents, which has the highest probability of causing death in one year? What is that probability?
b. Which has the lowest probability of causing death in one year? What is that probability?
Write your answers as fractions.
- ▶ 60. a. Of all of the specific forms of nontransportation accidents, which has the highest probability of causing death in a lifetime? What is that probability?
b. Which has the lowest probability of causing death in a lifetime? What is that probability?
Write your answers as fractions.
61. A family has two children. Using b to stand for boy and g for girl in ordered pairs, give each of the following.
a. the sample space
b. the event E that the family has exactly one daughter
c. the event F that the family has at least one daughter
d. the event G that the family has two daughters
e. $p(E)$ f. $p(F)$ g. $p(G)$
h. $o(E)$ i. $o(F)$ j. $o(G)$
(Assume that boys and girls are equally likely.)
62. Two coins are tossed. Using ordered pairs, give the following.
a. the sample space
b. the event E that exactly one is heads
c. the event F that at least one is heads
d. the event G that two are heads
e. $p(E)$ f. $p(F)$ g. $p(G)$
h. $o(E)$ i. $o(F)$ j. $o(G)$
- ▶ 63. A family has three children. Using b to stand for boy and g for girl and using ordered triples such as (b, b, g) , give the following.
a. the sample space
b. the event E that the family has exactly two daughters
c. the event F that the family has at least two daughters
d. the event G that the family has three daughters
e. $p(E)$ f. $p(F)$ g. $p(G)$
h. $o(E)$ i. $o(F)$ j. $o(G)$
(Assume that boys and girls are equally likely.)
- ▶ 64. Three coins are tossed. Using ordered triples, give the following.
a. the sample space
b. the event E that exactly two are heads
c. the event F that at least two are heads
d. the event G that all three are heads
e. $p(E)$ f. $p(F)$ g. $p(G)$
h. $o(E)$ i. $o(F)$ j. $o(G)$
65. A couple plans on having two children.
a. Find the probability of having two girls.
b. Find the probability of having one girl and one boy.
c. Find the probability of having two boys.
d. Which is more likely: having two children of the same sex or two of different sexes? Why?
(Assume that boys and girls are equally likely.)
- ▶ 66. Two coins are tossed.
a. Find the probability that both are heads.
b. Find the probability that one is heads and one is tails.
c. Find the probability that both are tails.
d. Which is more likely: that the two coins match or that they don't match? Why?
67. A couple plans on having three children. Which is more likely: having three children of the same sex or of different sexes? Why? (Assume that boys and girls are equally likely.)
68. Three coins are tossed. Which is more likely: that the three coins match or that they don't match? Why?
69. A pair of dice is rolled. Using ordered pairs, give the following.
a. the sample space
HINT: S has 36 elements, one of which is $(1, 1)$.
b. the event E that the sum is 7
c. the event F that the sum is 11
d. the event G that the roll produces doubles
e. $p(E)$ f. $p(F)$ g. $p(G)$
h. $o(E)$ i. $o(F)$ j. $o(G)$
- ▶ 70. Mendel found that snapdragons have no color dominance; a snapdragon with one red gene and one white gene will have pink flowers. If a pure-red snapdragon is crossed with a pure-white one, find the probability of the following.
a. a red offspring b. a white offspring
c. a pink offspring
71. If two pink snapdragons are crossed (see Exercise 70), find the probability of the following.
a. a red offspring b. a white offspring
c. a pink offspring
- ▶ 72. One parent is a cystic fibrosis carrier, and the other has no cystic fibrosis gene. Find the probability of each of the following.
a. The child would have cystic fibrosis.
b. The child would be a carrier.

- c. The child would not have cystic fibrosis and not be a carrier.
 d. The child would be healthy (i.e., free of symptoms).
73. If carrier-detection tests show that two prospective parents have sickle-cell trait (and are therefore carriers), find the probability of each of the following.
- Their child would have sickle-cell anemia.
 - Their child would have sickle-cell trait.
 - Their child would be healthy (i.e., free of symptoms).
- ▶ 74. If carrier-detection tests show that one prospective parent is a carrier of sickle-cell anemia and the other has no sickle-cell gene, find the probability of each of the following.
- The child would have sickle-cell anemia.
 - The child would have sickle-cell trait.
 - The child would be healthy (i.e., free of symptoms).
75. If carrier-detection tests show that one prospective parent is a carrier of Tay-Sachs and the other has no Tay-Sachs gene, find the probability of each of the following.
- The child would have the disease.
 - The child would be a carrier.
 - The child would be healthy (i.e., free of symptoms).
- ▶ 76. If carrier-detection tests show that both prospective parents are carriers of Tay-Sachs, find the probability of each of the following.
- Their child would have the disease.
 - Their child would be a carrier.
 - Their child would be healthy (i.e., free of symptoms).
77. If a parent started to exhibit the symptoms of Huntington's disease after the birth of his or her child, find the probability of each of the following. (Assume that one parent carries a single gene for Huntington's disease and the other carries no such gene.)
- The child would have the disease.
 - The child would be a carrier.
 - The child would be healthy (i.e., free of symptoms).
81. Consider a “weighted die”—one that has a small weight in its interior. Such a weight would cause the face closest to the weight to come up less frequently and the face farthest from the weight to come up more frequently. Would the probabilities computed in Example 1 still be correct? Why or why not? Would the definition $p(E) = \frac{n(E)}{n(S)}$ still be appropriate? Why or why not?
82. Some dice have spots that are small indentations; other dice have spots that are filled with the same material but of a different color. Which of these two types of dice is not fair? Why? What would be the most likely outcome of rolling this type of die? Why?
HINT: 1 and 6 are on opposite faces, as are 2 and 5, and 3 and 4.
83. In the United States, 52% of the babies are boys, and 48% are girls. Do these percentages contradict an assumption that boys and girls are equally likely? Why?
84. Compare and contrast theoretical probability and relative frequency. Be sure that you discuss both the similarities and differences between the two.
85. Compare and contrast probability and odds. Be sure that you discuss both the similarities and differences between the two.

• HISTORY QUESTIONS

86. What prompted Dr. Nancy Wexler's interest in Huntington's disease?
87. What resulted from Dr. Nancy Wexler's interest in Huntington's disease?

• PROJECTS

88. a. In your opinion, what is the probability that the last digit of a phone number is odd? Justify your answer.
 b. Randomly choose one page from the residential section of your local phone book. Count how many phone numbers on that page have an odd last digit and how many have an even last digit. Then compute (number of phone numbers with odd last digits)/(total number of phone numbers).
 c. Is your answer to part (b) a theoretical probability or a relative frequency? Justify your answer.
 d. How do your answers to parts (a) and (b) compare? Are they exactly the same, approximately the same, or dissimilar? Discuss this comparison, taking into account the ideas of probability theory.
89. a. Flip a coin ten times, and compute the relative frequency of heads.
 b. Repeat the experiment described in part (a) nine more times. Each time, compute the relative frequency of heads. After finishing parts (a) and (b), you will have flipped a coin 100 times, and you will have computed ten different relative frequencies.



Answer the following questions using complete sentences and your own words.

• CONCEPT QUESTIONS

78. Explain how you would find the theoretical probability of rolling an even number on a single die. Explain how you would find the relative frequency of rolling an even number on a single die.
79. Give five examples of events whose probabilities must be found empirically rather than theoretically.
80. Does the theoretical probability of an event remain unchanged from experiment to experiment? Why or why not? Does the relative frequency of an event remain unchanged from experiment to experiment? Why or why not?

- c. Discuss how your ten relative frequencies compare with each other and with the theoretical probability heads. In your discussion, use the ideas discussed under the heading “Relative Frequency versus Probability” in this section. Do not plagiarize; use your own words.
- d. Combine the results of parts (a) and (b), and find the relative frequency of heads for all 100 coin tosses. Discuss how this relative frequency compares with those found in parts (a) and (b) and with the theoretic probability of heads. Be certain to incorporate the Law of Large Numbers into your discussion.
90. a. Roll a single die twelve times, and compute the relative frequency with which you rolled a number below a 3.
- b. Repeat the experiment described in part (a) seven more times. Each time, compute the relative frequency with which you rolled a number below a 3. After finishing parts (a) and (b), you will have rolled a die ninety-six times, and you will have computed eight different relative frequencies.
- c. Discuss how your eight relative frequencies compare with each other and with the theoretical probability rolling a number below a 3. In your discussion, use the ideas discussed under the heading “Relative Frequency versus Probability” in this section. Do not plagiarize; use your own words.
- d. Combine the results of parts (a) and (b), and find the relative frequency of rolling a number below a 3 for all ninety-six die rolls. Discuss how this relative frequency compares with those found in parts (a) and (b) and with the theoretic probability of rolling a number less than a 3. Be certain to incorporate the Law of Large Numbers into your discussion.
91. Stand a penny upright on its edge on a smooth, hard, level surface. Then spin the penny on its edge. To do this, gently place a finger on the top of the penny. Then snap the penny with another finger (and immediately remove the holding finger) so that the penny spins rapidly before falling. Repeat this experiment fifty times, and compute the relative frequency of heads. Discuss whether or not the outcomes of this experiment are equally likely.
92. Stand a penny upright on its edge on a smooth, hard, level surface. Pound the surface with your hand so that the penny falls over. Repeat this experiment fifty times, and compute the relative frequency of heads. Discuss whether or not the outcomes of this experiment are equally likely.



93. In the 1970s, there was a mass screening of blacks for sickle-cell anemia and a mass screening of Jews for Tay-Sachs disease. One of these was a successful program. One was not. Write a research paper on these two programs.

Some useful links for this web project are listed on the text web site:

www.cengage.com/math/johnson

3.3 Basic Rules of Probability

OBJECTIVES

- Understand what type of number a probability can be
- Learn about the relationships between probabilities, unions, and intersections

In this section, we will look at the basic rules of probability theory. The first three rules focus on why a probability can be a number like $\frac{5}{7}$ or 32% but not a number like $\frac{9}{5}$ or -27% . Knowing what type of number a probability can be is key to understanding what a probability means in real life. It also helps in checking your answers.

How Big or Small Can a Probability Be?

No event occurs less than 0% of the time. How could an event occur negative 15% of the time? Also, no event occurs more than 100% of the time. How could an

event occur 125% of the time? Every event must occur between 0% and 100% of the time. For every event E ,

$$0\% \leq p(E) \leq 100\%$$

$$0 \leq p(E) \leq 1$$

converting from percents to decimals

This means that *if you ever get a negative answer or an answer greater than 1 when you calculate a probability, go back and find your error.*

What type of event occurs 100% of the time? That is, what type of event has a probability of 1? Such an event must include *all possible* outcomes; if any possible outcome is left out, the event could not occur 100% of the time. An event that has a probability of 1 must be the sample space, because the sample space is the set of all possible outcomes. As we discussed earlier, such an event is called a *certain event*.

What type of event occurs 0% of the time? That is, what type of event has a probability of 0? Such an event must not include any possible outcome; if a possible outcome is included, the event would not have a probability of 0. An event with a probability of 0 must be the null set. As we discussed earlier, such an event is called an *impossible event*.

PROBABILITY RULES

Rule 1 $p(\emptyset) = 0$ The probability of the null set is 0.

Rule 2 $p(S) = 1$ The probability of the sample space is 1.

Rule 3 $0 \leq p(E) \leq 1$ Probabilities are between 0 and 1 (inclusive).

Probability Rules 1, 2, and 3 can be formally verified as follows:

Rule 1:
$$p(\emptyset) = \frac{n(\emptyset)}{n(S)} = \frac{0}{n(S)} = 0$$

Rule 2:
$$p(S) = \frac{n(S)}{n(S)} = 1$$

Rule 3: E is a subset of S ; therefore,

$$0 \leq n(E) \leq n(S)$$

$$\frac{0}{n(S)} \leq \frac{n(E)}{n(S)} \leq \frac{n(S)}{n(S)} \quad \text{dividing by } n(S)$$

$$0 \leq p(E) \leq 1$$

EXAMPLE 1

ILLUSTRATING RULES 1, 2, AND 3 A single die is rolled once. Find the probability of:

- event E , “a 15 is rolled”
- event F , “a number between 1 and 6 (inclusive) is rolled”
- event G , “a 3 is rolled”

SOLUTION

The sample space is $S = \{1, 2, 3, 4, 5, 6\}$, and $n(S) = 6$.

- a.** It is *wrong* to say that $E = \{15\}$. Remember, we are talking about rolling a single die. Rolling a 15 is not one of the possible outcomes. That is, $15 \notin S$. There are no possible outcomes that result in rolling a 15, so the number of outcomes in event E is $n(E) = 0$. Thus,

$$p(E) = n(E)/n(S) = \frac{0}{6} = 0$$

The number 15 will be rolled 0% of the time. Event E is an impossible event.

Mathematically, we say that event E consists of no possible outcomes, so $E = \emptyset$. This agrees with rule 1, since

$$p(E) = p(\emptyset) = 0$$

b. $F = \{1, 2, 3, 4, 5, 6\}$, so $n(F) = 6$. Thus,

$$p(F) = n(F)/n(S) = \frac{6}{6} = 1$$

A number between 1 and 6 (inclusive) will be rolled 100% of the time. Event F is a certain event.

Mathematically, we say that event F consists of every possible outcome, so $F = S$. This agrees with rule 2, since

$$p(F) = p(S) = 1$$

c. $G = \{3\}$, so $n(G) = 1$. Thus

$$p(G) = n(G)/n(S) = 1/6$$

This agrees with rule 3, since

$$0 \leq p(G) \leq 1$$

Mutually Exclusive Events

Two events that cannot both occur at the same time are called **mutually exclusive**. In other words, E and F are mutually exclusive if and only if $E \cap F = \emptyset$.

EXAMPLE 2

DETERMINING WHETHER TWO EVENTS ARE MUTUALLY EXCLUSIVE A die is rolled. Let E be the event “an even number comes up,” F the event “a number greater than 3 comes up,” and G the event “an odd number comes up.”

- Are E and F mutually exclusive?
- Are E and G mutually exclusive?

SOLUTION

- $E = \{2, 4, 6\}$, $F = \{4, 5, 6\}$, and $E \cap F = \{4, 6\} \neq \emptyset$ (see Figure 3.11). Therefore, E and F are *not* mutually exclusive; the number that comes up could be *both* even and greater than 3. In particular, it could be 4 or 6.
- $E = \{2, 4, 6\}$, $G = \{1, 3, 5\}$, and $E \cap G = \emptyset$. Therefore, E and G are mutually exclusive; the number that comes up could *not* be both even and odd.

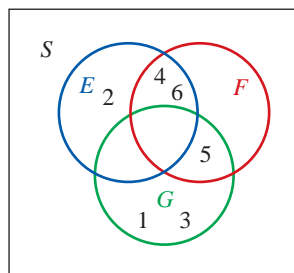


FIGURE 3.11 A Venn diagram for Example 2.

EXAMPLE 3

DETERMINING WHETHER TWO EVENTS ARE MUTUALLY EXCLUSIVE Let M be the event “being a mother,” F the event “being a father,” and D the event “being a daughter.”

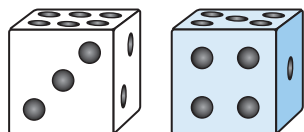
- Are events M and D mutually exclusive?
- Are events M and F mutually exclusive?

SOLUTION

- a. M and D are mutually exclusive if $M \cap D = \emptyset$. $M \cap D$ is the set of all people who are both mothers and daughters, and that set is not empty. A person can be a mother and a daughter at the same time. M and D are not mutually exclusive because being a mother does not exclude being a daughter.
- b. M and F are mutually exclusive if $M \cap F = \emptyset$. $M \cap F$ is the set of all people who are both mothers and fathers, and that set is empty. A person cannot be a mother and a father at the same time. M and F are mutually exclusive because being a mother does exclude the possibility of being a father.

Pair-of-Dice Probabilities

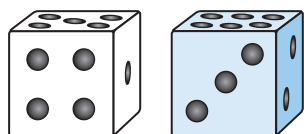
To find probabilities involving the rolling of a pair of dice, we must first determine the sample space. The *sum* can be anything from 2 to 12, but we will not use $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ as the sample space because those outcomes are not equally likely. Compare a sum of 2 with a sum of 7. There is only one way in which the sum can be 2, and that is if each die shows a 1. There are many ways in which the sum can be a 7, including:



The event (3, 4).

- a 1 and a 6
- a 2 and a 5
- a 3 and a 4

What about a 4 and a 3? Is that the same as a 3 and a 4? They certainly *appear* to be the same. But if one die were blue and the other were white, it would be quite easy to tell them apart. So there are other ways in which the sum can be a 7:



The event (4, 3).

- a 4 and a 3
- a 5 and a 2
- a 6 and a 1

Altogether, there are six ways in which the sum can be a 7. There is only one way in which the sum can be a 2. The outcomes “the sum is 2” and “the sum is 7” are not equally likely.

To have equally likely outcomes, we must use outcomes such as “rolling a 4 and a 3” and “rolling a 3 and a 4.” We will use ordered pairs as a way of abbreviating these longer descriptions. So we will denote the outcome “rolling a 4 and a 3” with the ordered pair (4, 3), and we will denote the outcome “rolling a 3 and a 4” with the ordered pair (3, 4). Figure 3.12 lists all possible outcomes and the resulting sums. Notice that $n(S) = 6 \cdot 6 = 36$.

sum	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)
7	8	9	10	11	12	

FIGURE 3.12 Outcomes of rolling two dice.

EXAMPLE 4

FINDING PAIR-OF-DICE PROBABILITIES A pair of dice is rolled. Find the probability of each of the following events.

- a. The sum is 7. b. The sum is greater than 9. c. The sum is even.

SOLUTION

- a. To find the probability that the sum is 7, let D be the event “the sum is 7.” From Figure 3.12, $D = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$, so $n(D) = 6$; therefore,

$$p(D) = \frac{n(D)}{n(S)} = \frac{6}{36} = \frac{1}{6}$$

This means that if we were to roll a pair of dice a large number of times, we should expect to get a sum of 7 approximately one-sixth of the time.



Notice that $p(D) = \frac{1}{6}$ is between 0 and 1, as are all probabilities.

- b. To find the probability that the sum is greater than 9, let E be the event “the sum is greater than 9.”

$E = \{(4, 6), (5, 5), (6, 4), (5, 6), (6, 5), (6, 6)\}$, so $n(E) = 6$; therefore,

$$p(E) = \frac{n(E)}{n(S)} = \frac{6}{36} = \frac{1}{6}$$

This means that if we were to roll a pair of dice a large number of times, we should expect to get a sum greater than 9 approximately one-sixth of the time.

- c. Let F be the event “the sum is even.”

$F = \{(1, 1), (1, 3), (2, 2), (3, 1), \dots, (6, 6)\}$, so $n(F) = 18$ (refer to Figure 3.12); therefore,

$$p(F) = \frac{n(F)}{n(S)} = \frac{18}{36} = \frac{1}{2}$$

This means that if we were to roll a pair of dice a large number of times, we should expect to get an even sum approximately half of the time.



© Scala/Art Resource, NY

A Roman painting on marble of the daughters of Niobe using knucklebones as dice. This painting was found in the ruins of Herculaneum, a city that was destroyed along with Pompeii by the eruption of Vesuvius.

EXAMPLE 5

USING THE CARDINAL NUMBER FORMULAS TO FIND PROBABILITIES

A pair of dice is rolled. Use the Cardinal Number Formulas (where appropriate) and the results of Example 4 to find the probabilities of the following events.

- the sum is not greater than 9.
- the sum is greater than 9 and even.
- the sum is greater than 9 or even.

SOLUTION

- a. We could find the probability that the sum is not greater than 9 by counting, as in Example 4, but the counting would be rather excessive. It is easier to use one of the Cardinal Number Formulas from Chapter 2 on sets. The event “the sum is not greater than 9” is the complement of event E (“the sum is greater than 9”) and can be expressed as E' .

$$\begin{aligned} n(E') &= n(U) - n(E) && \text{Complement Formula} \\ &= n(S) - n(E) && \text{“universal set” and “sample space” represent} \\ &= 36 - 6 = 30 && \text{the same idea.} \\ p(E') &= \frac{n(E')}{n(S)} = \frac{30}{36} = \frac{5}{6} \end{aligned}$$

This means that if we were to roll a pair of dice a large number of times, we should expect to get a sum that’s not greater than 9 approximately five-sixths of the time.

- b. The event “the sum is greater than 9 and even” can be expressed as the event $E \cap F$. $E \cap F = \{(4, 6), (5, 5), (6, 4), (6, 6)\}$, so $n(E \cap F) = 4$; therefore,

$$p(E \cap F) = \frac{n(E \cap F)}{n(S)} = \frac{4}{36} = \frac{1}{9}$$

This means that if we were to roll a pair of dice a large number of times, we should expect to get a sum that’s both greater than 9 and even approximately one-ninth of the time.

- c. Finding the probability that the sum is greater than 9 or even by counting would require an excessive amount of counting. It is easier to use one of the Cardinal Number Formulas from Chapter 2. The event “the sum is greater than 9 or even” can be expressed as the event $E \cup F$.

$$\begin{aligned} n(E \cup F) &= n(E) + n(F) - n(E \cap F) && \text{Union/Intersection Formula} \\ &= 6 + 18 - 4 && \text{from part (e), and Example 4} \\ &= 20 \\ p(E \cup F) &= \frac{n(E \cup F)}{n(S)} = \frac{20}{36} = \frac{5}{9} \end{aligned}$$

This means that if we were to roll a pair of dice a large number of times, we should expect to get a sum that’s either greater than 9 or even approximately five-ninths of the time.



More Probability Rules

In part (b) of Example 4, we found that $p(E) = \frac{1}{6}$, and in part (a) of Example 5, we found that $p(E') = \frac{5}{6}$. Notice that $p(E) + p(E') = \frac{1}{6} + \frac{5}{6} = 1$. This should make sense to you. It just means that if E happens one-sixth of the time, then E' has to happen the other five-sixths of the time. This always happens—for any event E , $p(E) + p(E') = 1$.

As we will see in Exercise 79, the fact that $p(E) + p(E') = 1$ is closely related to the Cardinal Number Formula $n(E) + n(E') = n(S)$. The main difference is that one is expressed in the language of probability theory and the other is expressed in the language of set theory. In fact, the following three rules are all set theory rules (from Chapter 2) rephrased so that they use the language of probability theory rather than the language of set theory.

MORE PROBABILITY RULES

Rule 4	The Union/Intersection Rule	$p(E \cup F) = p(E) + p(F) - p(E \cap F)$
Rule 5	The Mutually Exclusive Rule	$p(E \cup F) = p(E) + p(F), \text{ if } E \text{ and } F \text{ are mutually exclusive}$
Rule 6	The Complement Rule	$p(E) + p(E') = 1 \text{ or, equivalently, } p(E') = 1 - p(E)$

In Example 5, we used some Cardinal Number Formulas from Chapter 2 to avoid excessive counting. Some find it easier to use Probability Rules to calculate probabilities, rather than Cardinal Number Formulas.

EXAMPLE 6

USING RULES 4, 5, AND 6 Use a Probability Rule rather than a Cardinal Number Formula to find:

- the probability that the sum is not greater than 9
- the probability that the sum is greater than 9 or even

SOLUTION

We will use the results of Example 4.

- a. $p(E') = 1 - p(E)$ **the Complement Rule**

$$\begin{aligned} &= 1 - \frac{1}{6} && \text{from Example 4} \\ &= \frac{5}{6} \end{aligned}$$

- b. $p(E \cup F) = p(E) + p(F) - p(E \cap F)$ **the Union/Intersection Rule**
from parts (b) and (c) of Example 4, and part (b) of Example 5

$$\begin{aligned} &= \frac{1}{6} + \frac{1}{2} - \frac{1}{9} && \text{getting a common denominator} \\ &= \frac{3}{18} + \frac{9}{18} - \frac{2}{18} = \frac{5}{9} \end{aligned}$$

Probabilities and Venn Diagrams

Venn diagrams can be used to illustrate probabilities in the same way in which they are used in set theory. In this case, we label each region with its probability rather than its cardinal number.

EXAMPLE 7

USING VENN DIAGRAM Zaptronics manufactures compact discs and their cases for several major labels. A recent sampling of Zaptronics' products has indicated that 5% have a defective case, 3% have a defective disc, and 7% have at least one of the two defects.

- Find the probability that a Zaptronics product has both defects.
- Draw a Venn diagram that shows the probabilities of each of the basic regions.

SOLUTION

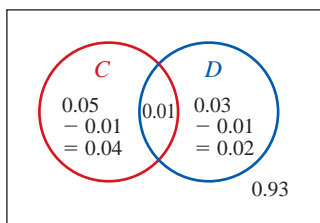


FIGURE 3.13

A Venn diagram for Example 6.

- c. Use the Venn diagram to find the probability that a Zaptronics product has neither defect.
 d. Find the probability that a Zaptronics product has neither defect, using probability rules rather than a Venn diagram.

- a. Let C be the event that the case is defective, and let D be the event that the disc is defective. We are given that $p(C) = 5\% = 0.05$, $p(D) = 3\% = 0.03$, and $p(C \cup D) = 7\% = 0.07$, and we are asked to find $p(C \cap D)$. To do this, substitute into Probability Rule 4.

$$\begin{aligned} p(C \cup D) &= p(C) + p(D) - p(C \cap D) && \text{the Union/Intersection Rule} \\ 0.07 &= 0.05 + 0.03 - p(C \cap D) && \text{substituting} \\ 0.07 &= 0.08 - p(C \cap D) && \text{adding} \\ p(C \cap D) &= 0.01 = 1\% && \text{solving} \end{aligned}$$

This means that 1% of Zaptronics' products have a defective case *and* a defective disc.

- b. The Venn diagram for this is shown in Figure 3.13. The 0.93 probability in the lower right corner is obtained by first finding the sum of the probabilities of the other three basic regions:

$$0.04 + 0.01 + 0.02 = 0.07$$

The rectangle itself corresponds to the sample space, and its probability is 1. Thus, the probability of the outer region, the region outside of the C and D circles, is

$$1 - 0.07 = 0.93$$

- c. The only region that involves no defects is the outer region, the one whose probability is 0.93. This means that 93% of Zaptronics' products have neither defect.
 d. If we are to not use the Venn diagram, we must first rephrase the event “a Zaptronics product has neither defect” as “a product does not have either defect.” The “does not” part of this means *complement*, and the “either defect” means *either C or D*. This means that

- “has either defect” translates to $C \cup D$
- “does not have either defect” translates to $(C \cup D)'$

So we are asked to find $P((C \cup D)')$.

$$\begin{aligned} p((C \cup D)') &= 1 - p(C \cup D) && \text{the Complement Rule} \\ &= 1 - 0.07 && \text{substituting} \\ &= 0.93 = 93\% && \text{subtracting} \end{aligned}$$

This means that 93% of Zaptronics' products are defect-free.

3.3 EXERCISES

In Exercises 1–10, determine whether E and F are mutually exclusive. Write a sentence justifying your answer.

1. E is the event “being a doctor,” and F is the event “being a woman.”
2. E is the event “it is raining,” and F is the event “it is sunny.”
- ▶ 3. E is the event “being single,” and F is the event “being married.”
4. E is the event “having naturally blond hair,” and F is the event “having naturally black hair.”
5. E is the event “having brown hair,” and F is the event “having gray hair.”
6. E is the event “being a plumber,” and F is the event “being a stamp collector.”
7. E is the event “wearing boots,” and F is the event “wearing sandals.”
- ▶ 8. E is the event “wearing shoes,” and F is the event “wearing socks.”
9. If a die is rolled once, E is the event “getting a four,” and F is the event “getting an odd number.”

▶ Selected exercises available online at www.webassign.net/brookscole

- ▶ 10. If a die is rolled once, E is the event “getting a four,” and F is the event “getting an even number.”

In Exercises 11–18, a card is dealt from a complete deck of fifty-two playing cards (no jokers). Use probability rules (when appropriate) to find the probability that the card is as stated. (Count an ace as high.)

11. a. a jack and red b. a jack or red
c. not a red jack
- ▶ 12. a. a jack and a heart b. a jack or a heart
c. not a jack of hearts
- ▶ 13. a. a 10 and a spade b. a 10 or a spade
c. not a 10 of spades
14. a. a 5 and black b. a 5 or black
c. not a black 5
- ▶ 15. a. under a 4
b. above a 9
c. both under a 4 and above a 9
d. either under a 4 or above a 9
- ▶ 16. a. above a jack
b. below a 3
c. both above a jack and below a 3
d. either above a jack or below a 3
17. a. above a 5
b. below a 10
c. both above a 5 and below a 10
d. either above a 5 or below a 10
- ▶ 18. a. above a 7
b. below a queen
c. both above a 7 and below a queen
d. either above a 7 or below a queen

In Exercises 19–26, use complements to find the probability that a card dealt from a full deck (no jokers) is as stated. (Count an ace as high.)

- ▶ 19. not a queen
- ▶ 20. not a 7
21. not a face card
22. not a heart
- ▶ 23. above a 3
- ▶ 24. below a queen
25. below a jack
26. above a 5
- ▶ 27. If $o(E) = 5:9$, find $o(E')$.
- ▶ 28. If $o(E) = 1:6$, find $o(E')$.
- ▶ 29. If $p(E) = \frac{2}{7}$, find $o(E)$ and $o(E')$.
- ▶ 30. If $p(E) = \frac{3}{8}$, find $o(E)$ and $o(E')$.
- ▶ 31. If $o(E) = a:b$, find $o(E')$.
32. If $p(E) = \frac{a}{b}$, find $o(E')$.

HINT: Use Exercise 49 from Section 3.2 and Exercise 31 above.

In Exercises 33–38, use Exercise 32 to find the odds that a card dealt from a full deck (no jokers) is as stated.

- ▶ 33. not a king 34. not an 8
35. not a face card 36. not a club
37. above a 4 38. below a king

In Exercises 39–42, use the following information: To determine the effect their salespeople have on purchases, a department store polled 700 shoppers regarding whether or not they made a purchase and whether or not they were pleased with the service they received. Of those who made a purchase, 151 were happy with the service and 133 were not. Of those who made no purchase, 201 were happy with the service and 215 were not. Use probability rules (when appropriate) to find the probability of the event stated.

- ▶ 39. a. A shopper made a purchase.
b. A shopper did not make a purchase.
- ▶ 40. a. A shopper was happy with the service received.
b. A shopper was unhappy with the service received.
41. a. A shopper made a purchase and was happy with the service.
b. A shopper made a purchase or was happy with the service.
42. a. A shopper made no purchase and was unhappy with the service.
b. A shopper made no purchase or was unhappy with the service.

In Exercises 43–46, use the following information: A supermarket polled 1,000 customers regarding the size of their bill. The results are given in Figure 3.14.

Size of Bill	Number of Customers
below \$20.00	208
\$20.00–\$39.99	112
\$40.00–\$59.99	183
\$60.00–\$79.99	177
\$80.00–\$99.99	198
\$100.00 or above	122

FIGURE 3.14 Supermarket bills.

Use probability rules (when appropriate) to find the relative frequency with which a customer's bill is as stated.

- ▶ 43. a. less than \$40.00
b. \$40.00 or more
44. a. less than \$80.00
b. \$80.00 or more
45. a. between \$40.00 and \$79.99
b. not between \$40.00 and \$79.99

- ▶ 46. a. between \$20.00 and \$79.99
b. not between \$20.00 and \$79.99

In Exercises 47–54, find the probability that the sum is as stated when a pair of dice is rolled.

- 47. a. 7 b. 9 c. 11
- ▶ 48. a. 2 b. 4 c. 6
- 49. a. 7 or 11 b. 7 or 11 or doubles
- ▶ 50. a. 8 or 10 b. 8 or 10 or doubles
- 51. a. odd and greater than 7
b. odd or greater than 7
- ▶ 52. a. even and less than 5
b. even or less than 5
- ▶ 53. a. even and doubles
b. even or doubles
- ▶ 54. a. odd and doubles
b. odd or doubles

In Exercises 55–58, use the following information: After examining their clients' records, Crashorama Auto Insurance calculated the probabilities in Figure 3.15.

	Miles/Year < 10,000	10,000 ≤ Miles/Year < 20,000	20,000 ≤ Miles/Year
Accident	0.05	0.10	0.3
No accident	0.20	0.15	0.2

FIGURE 3.15 Crashorama's Accident Incidence.

- 55. a. Find the probability that a client drives 20,000 miles/year or more or has an accident.
b. Find the probability that a client drives 20,000 miles/year or more and has an accident.
c. Find the probability that the client does not drive 20,000 miles/year or more and does not have an accident.
- ▶ 56. a. Find the probability that the client drives less than 10,000 miles/year and has an accident.
b. Find the probability that a client drives less than 10,000 miles/year or has an accident.
c. Find the probability that a client does not drive less than 10,000 miles per year and does not have an accident.
- 57. Are the probabilities in Figure 3.15 theoretical probabilities or relative frequencies? Why?
- 58. a. Find the probability that a client drives either less than 10,000 miles/year or 20,000 miles/year or more.
b. Find the probability that a client both has an accident and drives either less than 10,000 miles/year or 20,000 miles/year or more.

In Exercises 59–62, use the following information: The results of CNN's 2008 presidential election poll are given in Figure 3.16.

	Obama	McCain	Other/No Answer
Male Voters	23.0%	22.6%	1.4%
Female Voters	29.7%	22.8%	0.5%

FIGURE 3.16 Exit poll results, by gender. Source: CNN.

- 59. a. Find the probability that a polled voter voted for Obama or is male.
b. Find the probability that a polled voter voted for Obama and is male.
c. Find the probability that a polled voter didn't vote for Obama and is not male.
- ▶ 60. a. Find the probability that a polled voter voted for McCain or is female.
b. Find the probability that a polled voter voted for McCain and is female.
c. Find the probability that a polled voter didn't vote for McCain and is not female.
- 61. Are the probabilities in Figure 3.16 theoretical probabilities or relative frequencies? Why?
- ▶ 62. a. Find the probability that a polled voter didn't vote for Obama.
b. Find the probability that a polled voter is female.
- 63. Fried Foods hosted a group of twenty-five consumers at a tasting session. The consumers were asked to taste two new products and state whether or not they would be interested in buying them. Fifteen consumers said that they would be interested in buying Snackolas, twelve said that they would be interested in buying Chippers, and ten said that they would be interested in buying both.
 - a. What is the probability that one of the tasters is interested in buying at least one of the two new products?
 - b. What is the probability that one of the tasters is not interested in buying Snackolas but is interested in buying Chippers?
- ▶ 64. Ink Inc., a publishing firm, offers its 899 employees a cafeteria approach to benefits, in which employees can enroll in the benefit plan of their choice. Seven hundred thirteen employees have health insurance, 523 have dental insurance, and 489 have both health and dental insurance.
 - a. What is the probability that one of the employees has either health or dental insurance?
 - b. What is the probability that one of the employees has health insurance but not dental insurance?
- 65. ComCorp is considering offering city-wide Wi-Fi connections to the Web. They asked 3,000 customers about their use of the Web at home, and they received

- 1,451 responses. They found that 1,230 customers are currently connecting to the Web with a cable modem, 726 are interested in city-wide Wi-Fi, and 514 are interested in switching from a cable modem to city-wide Wi-Fi.
- What is the probability that one of the respondents is interested in city-wide Wi-Fi and is not currently using a cable modem?
 - What is the probability that one of the respondents is interested in city-wide Wi-Fi and is currently using a cable modem?
- ▶ 66. The Video Emporium rents DVDs and blue-rays only. They surveyed their 1,167 rental receipts for the last two weeks. Eight hundred thirty-two customers rented DVDs, and 692 rented blue-rays.
- What is the probability that a customer rents DVDs only?
 - What is the probability that a customer rents DVDs and blue-rays?
67. *Termiyak Magazine* conducted a poll of its readers, asking them about their telephones. Six hundred ninety-six readers responded. Five hundred seventy-two said they have a cell phone. Six hundred twelve said that they have a land line (a traditional, nonwireless telephone) at their home. Everyone has either a cell phone or a land line.
- What is the probability that one of the readers has only a cell phone?
 - What is the probability that one of the readers has only a land line?
68. Which of the probabilities in Exercises 63–67 are theoretical probabilities and which are relative frequencies? Why?
- ▶ 69. Use probability rules to find the probability that a child will either have Tay-Sachs disease or be a carrier if (a) each parent is a Tay-Sachs carrier, (b) one parent is a Tay-Sachs carrier and the other parent has no Tay-Sachs gene, (c) one parent has Tay-Sachs and the other parent has no Tay-Sachs gene.
70. Use probability rules to find the probability that a child will have either sickle-cell anemia or sickle-cell trait if (a) each parent has sickle-cell trait, (b) one parent has sickle-cell trait and the other parent has no sickle-cell gene, (c) one parent has sickle-cell anemia and the other parent has no sickle-cell gene.
71. Use probability rules to find the probability that a child will neither have Tay-Sachs disease nor be a carrier if (a) each parent is a Tay-Sachs carrier, (b) one parent is a Tay-Sachs carrier and the other parent has no Tay-Sachs gene, (c) one parent has Tay-Sachs and the other parent has no Tay-Sachs gene.
- ▶ 72. Use probability rules to find the probability that a child will have neither sickle-cell anemia nor sickle-cell trait if (a) each parent has sickle-cell trait, (b) one parent has sickle-cell trait and the other parent has no sickle-cell gene, (c) one parent has sickle-cell anemia and the other parent has no sickle-cell gene.
73. Mary is taking two courses, photography and economics. Student records indicate that the probability of passing photography is 0.75, that of failing economics is 0.65, and that of passing at least one of the two courses is 0.85. Find the probability of the following.
- Mary will pass economics.
 - Mary will pass both courses.
 - Mary will fail both courses.
 - Mary will pass exactly one course.
- ▶ 74. Alex is taking two courses; algebra and U.S. history. Student records indicate that the probability of passing algebra is 0.35, that of failing U.S. history is 0.35, and that of passing at least one of the two courses is 0.80. Find the probability of the following.
- Alex will pass history.
 - Alex will pass both courses.
 - Alex will fail both courses.
 - Alex will pass exactly one course.
75. Of all the flashlights in a large shipment, 15% have a defective bulb, 10% have a defective battery, and 5% have both defects. If you purchase one of the flashlights in this shipment, find the probability that it has the following.
- a defective bulb or a defective battery
 - a good bulb or a good battery
 - a good bulb and a good battery
- ▶ 76. Of all the DVDs in a large shipment, 20% have a defective disc, 15% have a defective case, and 10% have both defects. If you purchase one of the DVDs in this shipment, find the probability that it has the following.
- a defective disc or a defective case
 - a good disc or a good case
 - a good disc and a good case
77. Verify the Union/Intersection Rule.
HINT: Divide the Cardinal Number Formula for the Union of Sets from Section 2.1 by $n(S)$.
78. Verify the Mutually Exclusive Rule.
HINT: Start with the Union/Intersection Rule. Then use the fact that E and F are mutually exclusive.
79. Verify the Complement Rule.
HINT: Are E and E' mutually exclusive?



Answer the following questions using complete sentences and your own words.

• CONCEPT QUESTIONS

80. What is the complement of a certain event? Justify your answer.

81. Compare and contrast mutually exclusive events and impossible events. Be sure to discuss both the similarities and the differences between these two concepts.
82. Explain why it is necessary to subtract $p(E \cap F)$ in Probability Rule 4. In other words, explain why Probability Rule 5 is not true for all events.



FRACTIONS ON A GRAPHING CALCULATOR

Some graphing calculators—including the TI-83, TI-84, and Casio—will add, subtract, multiply, and divide fractions and will give answers in reduced fractional form.

Reducing Fractions

The fraction $42/70$ reduces to $3/5$. To do this on your calculator, you must make your screen read “ $42/70 \blacktriangleright$ Frac” (“ $42 _ 70$ ” on a Casio). The way that you do this varies.



- TI-83/84**
- Type $42 \div 70$, but do not press **ENTER**. This causes “ $42/70$ ” to appear on the screen.
 - Press the **MATH** button.
 - Highlight option 1, “ \blacktriangleright Frac.” (Option 1 is automatically highlighted. If we were selecting a different option, we would use the **▲** and **▼** buttons to highlight it.)
 - Press **ENTER**. This causes “ $42/70 \blacktriangleright$ Frac” to appear on the screen.
 - Press **ENTER**. This causes $3/5$ to appear on the screen.

- CASIO**
- Type $42 \frac{a}{b}{c} 70$.
 - Type **EXE**, and your display will read “ 3_5 ,” which should be interpreted as “ $3/5$.”

EXAMPLE 8

Use your calculator to compute

$$\frac{1}{6} + \frac{1}{2} - \frac{1}{9}$$

and give your answer in reduced fractional form.

SOLUTION

On a TI, make your screen read “ $1/6 + 1/2 - 1/9 \blacktriangleright$ Frac” by typing

$$1 \div 6 + 1 \div 2 - 1 \div 9$$

and then inserting the “ \blacktriangleright Frac” command, as described above. Once you press **ENTER**, the screen will read “ $5/9$.”

On a Casio, make your screen read “1┘6 + 1┘2 - 1┘9” by typing

$$1 \left[\frac{\square}{\square} \right] 6 + 1 \left[\frac{\square}{\square} \right] 2 - 1 \left[\frac{\square}{\square} \right] 9$$

Once you press EXE, the screen will read “5┘9,” which should be interpreted as “5/9.”

EXERCISES

In Exercises 83–84, reduce the given fractions to lowest terms, both (a) by hand and (b) with a calculator. Check your work by comparing the two answers.

83. $\frac{18}{33}$ 84. $-\frac{42}{72}$

In Exercises 85–90, perform the indicated operations, and reduce the answers to lowest terms, both (a) by hand and (b) with a calculator. Check your work by comparing the two answers.

85. $\frac{6}{15} \cdot \frac{10}{21}$

86. $\frac{6}{15} \div \frac{10}{21}$

87. $\frac{6}{15} + \frac{10}{21}$

88. $\frac{6}{15} - \frac{10}{21}$

89. $\frac{7}{6} - \frac{5}{7} + \frac{9}{14}$

90. $\frac{-8}{5} - \left(\frac{-3}{28} + \frac{5}{21} \right)$

91. How could you use your calculator to get decimal answers to the above exercises rather than fractional answers?

3.4 Combinatorics and Probability

OBJECTIVES

- Apply the concepts of permutations and combinations to probability calculations
- Use probabilities to analyze games of chance such as the lottery.

You would probably guess that it's pretty unlikely that two or more people in your math class share a birthday. It turns out, however, that it's surprisingly likely. In this section, we'll look at why that is.

Finding a probability involves finding the number of outcomes in an event and the number of outcomes in the sample space. So far, we have used probability rules as an alternative to excessive counting. Another alternative is combinatorics—that

is, permutations, combinations, and the Fundamental Principle of Counting, as covered in Sections 2.3 and 2.4. The flowchart used in Chapter 2 is summarized below.

WHICH COMBINATORICS METHOD?

1. If the selection is done with replacement, use the Fundamental Principle of Counting.
2. If the selection is done without replacement and there is only one category, use:
 - a. permutations if the order of selection does matter:

$${}_n P_r = \frac{n!}{(n-r)!}$$

- b. combinations if the order of selection does not matter:

$${}_n C_r = \frac{n!}{(n-r)!r!}$$

3. If there is more than one category, use the Fundamental Principle of Counting with one box per category. Use (2) above to determine the numbers that go in the boxes.

EXAMPLE 1

SOLUTION

A BIRTHDAY PROBABILITY A group of three people is selected at random. What is the probability that at least two of them will have the same birthday?

We will assume that all birthdays are equally likely, and for the sake of simplicity, we will ignore leap-year day (February 29).

- *The experiment* is to ask three people their birthdays. One possible outcome is (May 1, May 3, August 23).
- *The sample space S* is the set of all possible lists of three birthdays.
- *Finding $n(S)$* by counting the elements in S is impractical, so we will use combinatorics as described in “Which Combinatorics Method” above.
- The selected items are birthdays. They are selected with replacement because people can share the same birthday. This tells us to use *the Fundamental Principle of Counting*:

first person's birthday second person's birthday third person's birthday



- Each birthday can be any one of the 365 days in a year
- $n(S) = \boxed{365} \cdot \boxed{365} \cdot \boxed{365} = 365^3$
- *The event E* is the set of all possible lists of three birthdays in which some birthdays are the same. It is difficult to compute $n(E)$ directly; instead, we will compute $n(E')$ and use the Complement Rule. E' is the set of all possible lists of three birthdays in which no birthdays are the same.
- *Finding $n(E')$* using “Which Combinatorics Method”:
 - The birthdays are selected without replacement, because no birthdays are the same.
 - There is only one category: birthdays.

- The order of selection does matter—(May 1, May 3, August 23) is a different list from (May 3, August 23, May 1)—so use permutations
- $n(E') = {}_{365}P_3$
- Finding $p(E')$ and $p(E)$:

$$p(E') = \frac{n(E')}{n(S)} = \frac{{}_{365}P_3}{365^3} \quad \text{using the above results}$$

$$p(E) = 1 - p(E') \quad \text{the Complement Rule}$$

$$= 1 - \frac{{}_{365}P_3}{365^3} \quad \text{using the above result}$$

$$= 0.008204 \dots$$

$$\approx 0.8\% \quad \text{moving the decimal point left two places}$$

This result is not at all surprising. It means that two or more people in a group of three share a birthday slightly less than 1% of the time. In other words, this situation is extremely unlikely. However, we will see in Exercise 1 that it is quite likely that two or more people in a group of thirty share a birthday.

Lotteries

We're going to look at one of the most common forms of gambling in the United States: the lottery. Many people play the lottery, figuring that "someone's got to win—why not me?" Is this a reasonable approach to lottery games?

EXAMPLE 2

WINNING A LOTTERY Arizona, Connecticut, Missouri, and Tennessee operate 6/44 lotteries. In this game, a gambler selects any six of the numbers from 1 to 44. If his or her six selections match the six winning numbers, the player wins first prize. If his or her selections include five of the winning numbers, the player wins second prize. Find the probability of:

- the event E , winning first prize
- the event F , winning second prize

SOLUTION

a. The sample space S is the set of all possible lottery tickets. That is, it is the set of all possible choices of six numbers selected from the numbers 1 through 44.

- Finding $n(S)$ by counting the elements in S is impractical, so we will use combinatorics as described in "Which Combinatorics Method" above.
- The selected is done without replacement, because you can't select the same lottery number twice.
- Order does not matter, because the gambler can choose the six numbers in any order. Use combinations.
- $n(S) = {}_{44}C_6$
- Finding $n(E)$ is easy, because there is only one winning combination of numbers:

$$n(E) = 1$$

- Finding $p(E)$:

$$p(E) = \frac{n(E)}{n(S)} = \frac{1}{{}_{44}C_6} = \frac{1}{7,059,052} \quad \text{using the above results}$$

This means that only one out of approximately *seven million* combinations is the first-prize-winning combinations. This is an incredibly unlikely event. It is less probable than dying in one year by being hit by lightning (see Figure 3.10 on page 154.). You probably don't worry about dying from lightning, because it essentially doesn't happen. Sure, it happens to somebody sometime, but not to anyone you've ever known. Winning the lottery is similar. The next time that you get tempted to buy a lottery ticket, remember that you should feel less certain of winning first prize than you should of dying from lightning.

- b. To find the probability of event F , winning second prize, we need to find $n(F)$. We already found $n(S)$ in part a:

$$n(S) = {}_{44}C_6$$

- Finding $n(F)$:

- To win second prize, we must select five winning numbers and one losing number. This tells us to use *the Fundamental Principle of Counting*:

winning numbers	losing numbers
<input style="width: 40px; height: 20px;" type="text"/>	<input style="width: 40px; height: 20px;" type="text"/>

- We will use combinations in each box, for the same reasons that we used combinations to find $n(S)$.
- The state selects six winning numbers, and the second-prize-winner must select five of them, so there are ${}_6C_5$ ways of selecting the five winning numbers.
- There are $44 - 6 = 38$ losing numbers, of which the player selects 1, so there are ${}_{38}C_1$ ways of selecting one losing number.

- $n(F) = \boxed{{}_6C_5} \cdot \boxed{{}_{38}C_1}$ **using the above results**

- Finding $p(F)$:

- $$p(F) = \frac{n(F)}{n(S)} = \frac{{}_6C_5 \cdot {}_{38}C_1}{{}_{44}C_6} = 0.000032 \dots$$

How do we make sense of a decimal with so many zeros in front of it? One way is to round it off at the first nonzero digit and convert the result to a fraction:

$$p(F) = 0.00003 \dots \quad \text{rounding to the first nonzero digit}$$

$$= 3/100,000 \quad \text{converting to a fraction}$$

This means that if you buy a lot of 6/44 lottery tickets, you will win second prize approximately three times out of every 100,000 times you play. It also means that in any given game, there are about three second-prize ticket for every 100,000 ticket purchases.



To check your work in Example 2 part (b), notice that in the event, there is a distinction between two categories (winning numbers and losing numbers); in the sample space, there is no such distinction. Thus, the numerator of

$$p(E) = \frac{{}_6C_5 \cdot {}_{38}C_1}{{}_{44}C_6}$$

has two parts (one for each category), and the denominator has one part. Also, the numbers in front of the C s add correctly (6 winning numbers + 38 losing numbers = 44 total numbers to choose from), and the numbers after the C s add correctly (5 winning numbers + 1 losing number = 6 total numbers to select).

HOW TO WRITE A PROBABILITY

A probability can be written as a fraction, a percentage, or a decimal. Our goal is to write a probability in a form that is intuitively understandable.

- If $n(E)$ and $n(S)$ are both small numbers, such as $n(E) = 3$ and $n(S) = 12$, then write $p(E)$ as a **reduced fraction**:

$$p(E) = \frac{n(E)}{n(S)} = \frac{3}{12} = \frac{1}{4}$$

It's intuitively understandable to say that something happens about one out of every four times.

- If $n(E)$ and $n(S)$ are not small numbers and $p(E)$ is a decimal with at most two zeros after the decimal point, then write $p(E)$ as a **percentage**, as in Example 1:

$$\begin{aligned} p(E) &= 0.008204 \dots && \text{two zeros after the decimal point} \\ &\approx 0.8\% && \text{writing as a percentage} \end{aligned}$$

A probability of $0.008204 \dots$ is not easily understandable. But it is understandable to say that something happens slightly less than 1% of the time—that is, fewer than one out of every one hundred times.

- If $n(E)$ and $n(S)$ are not small numbers and $p(E)$ is a decimal with three or more zeros after the decimal point, then round off at the first nonzero digit and convert the result to a fraction, as in Example 2:

$$\begin{aligned} p(F) &= \frac{n(F)}{n(S)} = \frac{6 \cdot 38}{7,059,059} = 0.000032298 \dots \\ &\approx 0.00003 && \text{rounding at the first nonzero digit} \\ &= 3/100,000 && \text{converting to a fraction} \end{aligned}$$

A probability of $\frac{6 \cdot 38}{7,059,059}$ or $0.000032298 \dots$ or 0.003% is not so easily understood. It's more understandable to say that something happens about three times out of every 100,000 times.

EXAMPLE 3

WINNING POWERBALL Powerball is played in thirty states, the District of Columbia, and the U.S. Virgin Islands. It involves selecting any five of the numbers from 1 to 59, plus a “powerball number,” which is any one of the numbers from 1 to 39. Find the probability of winning first prize.



SOLUTION

The sample space S is the set of all possible lottery tickets. That is, it is the set of all possible choices of five numbers from 1 to 59 plus a powerball number from 1 to 39.

- Finding $n(S)$:
 - The numbers are selected without replacement, because the gambler can't select the same number twice.

- There are two categories—regular numbers and the powerball number—so we use the Fundamental Principle of Counting with two boxes.



- In each box, we will use combinations, because the player can choose the numbers in any order.
- The state selects five of fifty-nine regular numbers, so there are ${}_{59}C_5$ possible selections.
- The state selects one of thirty-nine powerball numbers, so there are ${}_{39}C_1$ possible powerball selections.
- The number of different lottery tickets is

$$n(S) = {}_{59}C_5 \cdot {}_{39}C_1 \quad \text{using the above results}$$

- *Finding $n(E)$* : Only one of these combinations is the winning lottery ticket, so $n(E) = 1$.
- *Finding $p(E)$* :

$$\begin{aligned} p(E) &= \frac{n(E)}{n(S)} = \frac{1}{{}_{59}C_5 \cdot {}_{39}C_1} && \text{using the above results} \\ &= 0.00000000512\dots && \text{more than three zeros after the decimal point} \\ &\approx 0.000000005 && \text{rounding to the first nonzero digit} \\ &= 5/1,000,000,000 && \text{converting to a fraction} \\ &= 1/200,000,000 && \text{reducing} \end{aligned}$$

This means that if you play powerball a lot, you will win first prize about once every 200 million games. It also means that in any given game, there is about one first prize ticket for every 200 million ticket purchases.

Keno

The game of keno is a casino version of the lottery. In this game, the casino has a container filled with balls numbered from 1 to 80. The player buys a keno ticket, with which he or she selects anywhere from 1 to 15 (usually 6, 8, 9, or 10) of those 80 numbers; the player's selections are called "spots." The casino chooses 20 winning numbers, using a mechanical device to ensure a fair game. If a sufficient number of the player's spots are winning numbers, the player receives an appropriate payoff.

EXAMPLE 4

WINNING AT KENO In the game of keno, if eight spots are marked, the player wins if five or more of his or her spots are selected. Find the probability of having five winning spots.

SOLUTION

The sample space S is the set of all ways in which a player can select eight numbers from the eighty numbers in the game.

- *Finding $n(S)$* :
 - Selection is done without replacement.
 - Order doesn't matter, so use combinations.
 - $n(S) = {}_{80}C_8$
- *The event E* is the set of all ways in which an eight-spot player can select five winning numbers and three losing numbers.
- *Finding $n(E)$* :
 - There are two categories—winning numbers and losing numbers—so use the Fundamental Principle of Counting.



TOPIC X THE BUSINESS OF GAMBLING: PROBABILITIES IN THE REAL WORLD

It used to be that legal commercial gambling was not common. Nevada made casino gambling legal in 1931, and for more than thirty years, it was the only place in the United States that had legal commercial gambling. Then in 1964, New Hampshire instituted the first lottery in the United States since 1894. In 1978, New Jersey became the second state to legalize casino gambling. Now, state-sponsored lotteries are common, Native American tribes have casinos in more than half the states, and some states have casinos in selected cities or on riverboats.

It is very likely that you will be exposed to gambling if you have not been already. You should approach gambling with an educated perspective. If you are considering gambling, know what you are up against. The casinos all use probabilities and combinatorics in designing their games to ensure that they make a consistent profit. Learn this mathematics so that they do not take advantage of you.

Public lotteries have a long history in the United States. The settlement of Jamestown was financed in part by an English lottery. George Washington

managed a lottery that paid for a road through the Cumberland Mountains. Benjamin Franklin used lotteries to finance cannons for the Revolutionary War, and John Hancock used lotteries to rebuild Faneuil Hall in Boston. Several universities, including Harvard, Dartmouth, Yale, and Columbia, were partly financed by lotteries. The U.S. Congress operated a lottery to help finance the Revolutionary War.

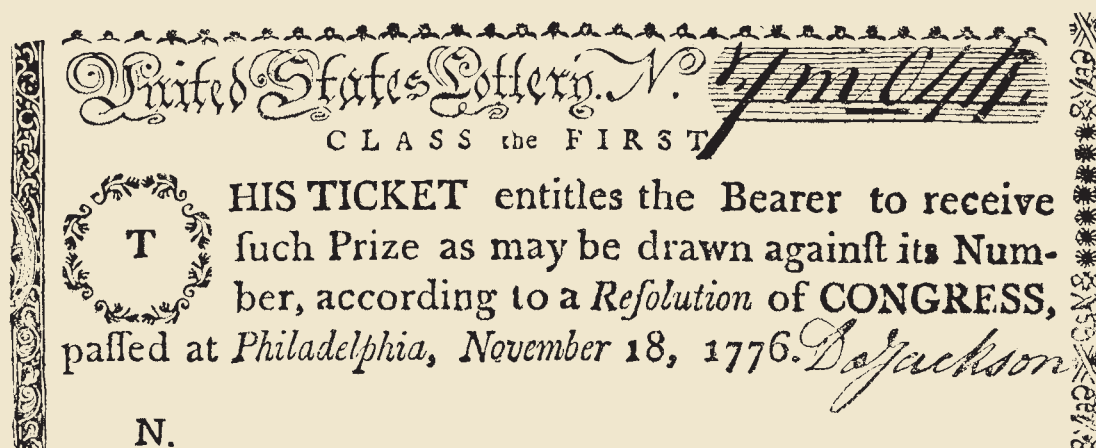
Today, public lotteries are a very big business. In 2008, Americans spent almost \$61 billion on lottery tickets. Lotteries are quite lucrative for the forty-two states that offer them; on the average, 30% of the money went back into government budgets. Fewer than half of the states dedicate the proceeds to education. Frequently, this money goes into the general fund. The states' cuts vary quite a bit. In Oregon in 2008, 54% of the money went to the state, and the remaining 46% went to prizes and administrative costs. In Rhode Island, the state took only 15%.

Most lottery sales come from a relatively small number of people. In Pennsylvania, for example, 29% of the

players accounted for 79% of the spending on the lottery in 2008. However, many people who don't normally play go berserk when the jackpots accumulate, partially because of the amazingly large winnings but also because of a lack of understanding of how unlikely it is that they will actually win. The largest cumulative jackpot was \$390 million, which was split by winners in Georgia and New Jersey in 2007. The largest single-winner jackpot was \$315 million in West Virginia in 2002. The winner opted to take a lump sum payment of \$114 million, instead of receiving twenty years of regular payments that would have added up to \$315 million.

In this section, we will discuss probabilities and gambling. Specifically, we will explore lotteries, keno, and card games. See Examples 2–7, and Exercises 5–22 and 35.

In Section 3.5, we will discuss how much money you can expect to win or lose when gambling. We will continue to discuss lotteries, keno, and card games, but we will also discuss roulette and raffles. See Exercises 1–12, 28–34, and 39–48 in that section.



FAC-SIMILE OF A CONTINENTAL LOTTERY TICKET.

- Use combinations in each category, as with $n(S)$.
- The casino selects twenty winning numbers, from which the gambler is to select five winning spots. There are ${}_{20}C_5$ different ways of doing this.
- The casino selects $80 - 20 = 60$ losing numbers, from which the gambler is to select $8 - 5 = 3$ losing spots. There are ${}_{60}C_3$ different ways of doing this.
- $n(E) = {}_{20}C_5 \cdot {}_{60}C_3$ **using the above results**
- Finding $p(E)$:

$$p(E) = \frac{{}_{20}C_5 \cdot {}_{60}C_3}{{}_{80}C_8} = 0.0183 \dots \quad \text{one zero after the decimal point}$$

$$\approx 1.8\% \quad \text{writing as a percentage}$$

This means that if you play eight-spot keno a lot, you will have five winning spots about 1.8% of the time. It also means that in any given game, about 1.8% of the players will have five winning spots.

Cards

One common form of poker is five-card draw, in which each player is dealt five cards. The order in which the cards are dealt is unimportant, so we compute probabilities with combinations rather than permutations.

EXAMPLE 5

SOLUTION

GETTING FOUR ACES Find the probability of being dealt four aces.

The *sample space* consists of all possible five-card hands that can be dealt from a deck of fifty-two cards

- Finding $n(S)$:
 - Selection is done without replacement.
 - Order does not matter, so use combinations.
 - $n(S) = {}_{52}C_5$
- The event E consists of all possible five-card hands that include four aces and one non-ace.
 - There are two categories—aces and non-aces—so use the Fundamental Principle of Counting.
 - Use combinations as with $n(S)$.
 - The gambler is to be dealt four of four aces. This can happen in ${}_4C_4$ ways.
 - The gambler is to be dealt one of $52 - 4 = 48$ non-aces. This can be done in ${}_{48}C_1$ ways.
 - $n(E) = {}_4C_4 \cdot {}_{48}C_1$
- Finding $p(E)$:

$$p(E) = \frac{{}_4C_4 \cdot {}_{48}C_1}{{}_{52}C_5} = 0.000018 \dots \quad \text{more than two zeros after the decimal point}$$

$$\approx 0.00002 \quad \text{rounding to the first nonzero digit}$$

$$= 2/100,000 \quad \text{rewriting as a fraction}$$

$$= 1/50,000 \quad \text{reducing}$$

This means that if you play cards a lot, you will be dealt four aces about once every 50,000 deals.



In the event, there is a distinction between two categories (aces and non-aces); in the sample space, there is no such distinction. Thus, the numerator of

$$p(E) = \frac{{}^4C_4 \cdot {}_{48}C_1}{{}_{52}C_5}$$

has two parts (one for each category), and the denominator has one part. Also, the numbers in front of the Cs add correctly (4 aces + 48 non-aces = 52 cards to choose from), and the numbers after the Cs add correctly (4 aces + 1 non-ace = 5 cards to select).

EXAMPLE 6

SOLUTION

GETTING FOUR OF A KIND Find the probability of being dealt four of a kind.

The sample space is the same as that in Example 5. The event “being dealt four of a kind” means “being dealt four 2’s or being dealt four 3’s or being dealt four 4’s . . . or being dealt four kings or being dealt four aces.” These latter events (“four 2’s,” “four 3’s,” etc.) are all mutually exclusive, so we can use the Mutually Exclusive Rule.

$$\begin{aligned} p(\text{four of a kind}) &= p(\text{four 2's} \cup \text{four 3's} \cup \cdots \cup \text{four kings} \cup \text{four aces}) \\ &= p(\text{four 2's}) + p(\text{four 3's}) + \cdots + p(\text{four kings}) + p(\text{four aces}) \end{aligned}$$

using the Mutually Exclusive Rule

Furthermore, these probabilities are all the same:

$$p(\text{four 2's}) = p(\text{four 3's}) = \cdots = p(\text{four aces}) = \frac{{}^4C_4 \cdot {}_{48}C_1}{{}_{52}C_5}$$

This means that the probability of being dealt four of a kind is

$$\begin{aligned} p(\text{four of a kind}) &= p(\text{four 2's}) + p(\text{four 3's}) + \cdots + p(\text{four kings}) \\ &\quad + p(\text{four aces}) \\ &= \frac{{}^4C_4 \cdot {}_{48}C_1}{{}_{52}C_5} + \frac{{}^4C_4 \cdot {}_{48}C_1}{{}_{52}C_5} + \cdots + \frac{{}^4C_4 \cdot {}_{48}C_1}{{}_{52}C_5} + \frac{{}^4C_4 \cdot {}_{48}C_1}{{}_{52}C_5} \\ &= 13 \cdot \frac{{}^4C_4 \cdot {}_{48}C_1}{{}_{52}C_5} && \text{there are thirteen denominations} \\ & && \text{(2 through ace)} \\ &= 13 \cdot \frac{1 \cdot 48}{2,598,960} && \text{from Example 4} \\ &= \frac{624}{2,598,960} \\ &= 0.0002400 \dots && \text{three zeros after the decimal point} \\ &\approx 0.0002 && \text{rounding to the first nonzero digit} \\ &= 2/10,000 && \text{converting to a fraction} \\ &= 1/5,000 && \text{reducing} \end{aligned}$$

This means that if you play cards a lot, you will be dealt four of a kind about once every 5,000 deals.

See Example 8 in Section 2.4.

EXAMPLE 7

SOLUTION

GETTING FIVE HEARTS Find the probability of being dealt five hearts.

The sample space is the same as in Example 5. The event consists of all possible five-card hands that include five hearts and no non-hearts. This involves two categories (hearts and non-hearts), so we will use the Fundamental Counting Principle and multiply the number of ways of getting five hearts and the number of ways of getting no non-hearts. There are

$${}_{13}C_5 = \frac{13!}{5! \cdot 8!} = 1,287$$

ways of getting five hearts, and there is

$${}_{39}C_0 = \frac{39!}{0! \cdot 39!} = 1$$

ways of getting no non-hearts. Thus, the probability of being dealt five hearts is

$$p(E) = \frac{{}_{13}C_5 \cdot {}_{39}C_0}{{}_{52}C_5} = \frac{1287 \cdot 1}{2,598,960} \approx 0.000495198 \approx 0.0005 = 1/2000$$



In the event, there is a distinction between two categories (hearts and non-hearts); in the sample space, there is no such distinction. Thus, the numerator of

$$p(E) = \frac{{}_{13}C_5 \cdot {}_{39}C_0}{{}_{52}C_5}$$

has two parts (one for each category), and the denominator has one part. Also, the numbers in front of the Cs add correctly (13 hearts + 39 non-hearts = 52 cards to choose from), and the numbers after the Cs add correctly (5 hearts + 0 non-hearts = 5 cards to select).

Notice that in Example 7, we could argue that since we're selecting only hearts, we can disregard the non-hearts. This would lead to the answer obtained in Example 7:

$$p(E) = \frac{{}_{13}C_5}{{}_{52}C_5} = \frac{1287}{2,598,960} \approx 0.000495198 \approx 0.0005 = 1/2000$$

However, this approach would not allow us to check our work in the manner described above; the numbers in front of the Cs don't add correctly, nor do the numbers after the Cs.

3.4 EXERCISES

- ▶ 1. A group of thirty people is selected at random. What is the probability that at least two of them will have the same birthday?
2. A group of sixty people is selected at random. What is the probability that at least two of them will have the same birthday?
3. How many people would you have to have in a group so that there is a probability of at least 0.5 that at least two of them will have the same birthday?
- ▶ 4. How many people would you have to have in a group so that there is a probability of at least 0.9 that at least two of them will have the same birthday?

▶ Selected exercises available online at www.webassign.net/brookscole

- ▶ 5. In 1990, California switched from a 6/49 lottery to a 6/53 lottery. Later, the state switched again, to a 6/51 lottery.
- Find the probability of winning first prize in a 6/49 lottery.
 - Find the probability of winning first prize in a 6/53 lottery.
 - Find the probability of winning first prize in a 6/51 lottery.
 - How much more probable is it that one will win the 6/49 lottery than the 6/53 lottery?
 - Why do you think California switched from a 6/49 lottery to a 6/53 lottery? And why do you think the state then switched to a 6/51 lottery?
(Answer using complete sentences.)
- ▶ 6. Find the probability of winning second prize—that is, picking five of the six winning numbers—with a 6/53 lottery.
7. Currently, the most common multinumber game is the 5/39 lottery. It is played in California, Georgia, Illinois, Michigan, New York, North Carolina, Ohio, Tennessee, and Washington.



- Find the probability of winning first prize.
 - Find the probability of winning second prize.
- ▶ 8. Currently, the second most common multinumber game is the 6/49 lottery. It is played in Massachusetts, New Jersey, Ohio, Pennsylvania, Washington, and Wisconsin.



- Find the probability of winning first prize.
 - Find the probability of winning second prize.
- ▶ 9. “Cash 5” is a 5/35 lottery. It is played in Arizona, Connecticut, Iowa, Massachusetts and South Dakota.



- Find the probability of winning first prize.
 - Find the probability of winning second prize.
10. The 6/44 lottery is played in Arizona, Connecticut, and New Jersey.
- Find the probability of winning first prize.
 - Find the probability of winning second prize.

11. Games like “Mega Millions” are played in thirty-nine states and the District of Columbia. It involves selecting any five of the numbers from 1 to 56, plus a number from 1 to 46.



Find the probability of winning first prize.

12. “Hot Lotto” is played in Delaware, Idaho, Iowa, Kansas, Minnesota, Montana, New Hampshire, New Mexico, North Dakota, Oklahoma, South Dakota, West Virginia, and the District of Columbia. It involves selecting any five of the numbers from 1 to 39 plus a number from 1 to 19. Find the probability of winning first prize.



13. “Wild Card 2” is played in Idaho, Montana, North Dakota, and South Dakota. It involves selecting any five of the numbers from 1 to 31, plus a card that’s either a jack, queen, king or ace of any one of the four suits. Find the probability of winning first prize.



14. “2 by 2” is played in Kansas, Nebraska, and North Dakota. It involves selecting two red numbers from 1 to 26 and two white numbers from 1 to 26. Find the probability of winning first prize.



15. There is an amazing variety of multinumber lotteries played in the United States. Currently, the following lotteries are played: 4/26, 4/77, 5/30, 5/31, 5/32, 5/34, 5/35, 5/36, 5/37, 5/38, 5/39, 5/40, 5/43, 5/47, 5/50, 6/25, 6/35, 6/39, 6/40, 6/42, 6/43, 6/44, 6/46, 6/47, 6/48, 6/49, 6/52, 6/53, and 6/54. Which is the easiest to win? Which is the hardest to win? Explain your reasoning.

HINT: It isn’t necessary to compute every single probability.

- ▶ 16. In the game of keno, if six spots are marked, the player wins if four or more of his or her spots are selected. Complete the chart in Figure 3.17.

Outcome	Probability
6 winning spots	
5 winning spots	
4 winning spots	
3 winning spots	
fewer than 3 winning spots	

FIGURE 3.17 Chart for Exercise 16.

- ▶ 17. In the game of keno, if eight spots are marked, the player wins if five or more of his or her spots are selected. Complete the chart in Figure 3.18.

Outcome	Probability
8 winning spots	
7 winning spots	
6 winning spots	
5 winning spots	
4 winning spots	
fewer than 4 winning spots	

FIGURE 3.18 Chart for Exercise 17.

18. In the game of keno, if nine spots are marked, the player wins if six or more of his or her spots are selected. Complete the chart in Figure 3.19.

Outcome	Probability
9 winning spots	
8 winning spots	
7 winning spots	
6 winning spots	
5 winning spots	
fewer than 5 winning spots	

FIGURE 3.19 Chart for Exercise 18.

19. “Pick three” games are played in thirty-six states. In this game, the player selects a three-digit number, such as 157. Also, the state selects a three-digit winning number.

- a. How many different three-digit numbers are there? Explain your reasoning.
 - b. If the player opts for “straight play,” she wins if her selecting matches the winning number *in exact order*. For example, if she selected the number 157, she wins only if the winning number is 157. How many different winning straight play numbers are there?
 - c. Find the probability of winning with straight play.
20. If a pick three player (see Exercise 19) opts for “box play,” the player wins if his selection matches the winning number *in any order*. For example, if he selected the number 157, he wins if the winning number is 715 or any other reordering of 157.
- a. How many different winning box play numbers are there if the three digits are different? That is, how many different ways are there to reorder a number such as 157?
 - b. Find the probability of winning with box play if the three digits are different.
 - c. How many different winning box play numbers are there if two digits are the same? That is, how many different ways are there to reorder a number such as 266?
 - d. Find the probability of winning with box play if two digits are the same.
21.
 - a. Find the probability of being dealt five spades when playing five-card draw poker.
 - b. Find the probability of being dealt five cards of the same suit when playing five-card draw poker.
 - c. When you are dealt five cards of the same suit, you have either a *flush* (if the cards are not in sequence) or a *straight flush* (if the cards are in sequence). For each suit, there are ten possible straight flushes (“ace, two, three, four, five,” through “ten, jack, queen, king, ace”). Find the probability of being dealt a straight flush.
 - d. Find the probability of being dealt a flush.
- ▶ 22.
 - a. Find the probability of being dealt an “aces over kings” full house (three aces and two kings).
 - b. Why are there $13 \cdot 12$ different types of full houses?
 - c. Find the probability of being dealt a full house. (Round each answer off to six decimal places.)

You order twelve burritos to go from a Mexican restaurant, five with hot peppers and seven without. However, the restaurant forgot to label them. If you pick three burritos at random, find the probability of each event in Exercises 23–30.

- ▶ 23. All have hot peppers.
24. None has hot peppers.
- ▶ 25. Exactly one has hot peppers.
26. Exactly two have hot peppers.
27. At most one has hot peppers.
28. At least one has hot peppers.

- ▶ 29. At least two have hot peppers.
- ▶ 30. At most two have hot peppers.
31. Two hundred people apply for two jobs. Sixty of the applicants are women.
- If two people are selected at random, what is the probability that both are women?
 - If two people are selected at random, what is the probability that only one is a woman?
 - If two people are selected at random, what is the probability that both are men?
 - If you were an applicant and the two selected people were not of your gender, do you think that the above probabilities would indicate the presence or absence of gender discrimination in the hiring process? Why or why not?
- ▶ 32. Two hundred people apply for three jobs. Sixty of the applicants are women.
- If three people are selected at random, what is the probability that all are women?
 - If three people are selected at random, what is the probability that two are women?
 - If three people are selected at random, what is the probability that one is a woman?
 - If three people are selected at random, what is the probability that none is a woman?
 - If you were an applicant and the three selected people were not of your gender, should the above probabilities have an impact on your situation? Why or why not?
33. In Example 2, $n(E) = 1$ because only one of the 7,059,052 possible lottery tickets is the first prize winner. Use combinations to show that $n(E) = 1$.



Answer the following questions using complete sentences and your own words.

• CONCEPT QUESTIONS

34. Explain why, in Example 6, $p(\text{four } 2\text{'s}) = p(\text{four } 3\text{'s}) = \dots = p(\text{four kings}) = p(\text{four aces})$.
35. Do you think a state lottery is a good thing for the state's citizens? Why or why not? Be certain to include a discussion of both the advantages and disadvantages of a state lottery to its citizens.
36. Why are probabilities for most games of chance calculated with combinations rather than permutations?
37. Suppose a friend or relative of yours regularly spends (and loses) a good deal of money on lotteries. How would you explain to this person why he or she loses so frequently?
38. In Example 2, $n(E) = 1$ because only one of the 7,059,052 possible lottery tickets is the first prize winner. Does this mean that it is impossible for two people to each buy a first-prize-winning lottery ticket? Explain.

• HISTORY QUESTION

39. Are public lotteries relative newcomers to the American scene? Explain.

3.5 Expected Value

OBJECTIVES

- Understand how expected values take both probabilities and winnings into account
- Use expected values to analyze games of chance
- Use expected values to make decisions

Suppose you are playing roulette, concentrating on the \$1 single-number bet. At one point, you were \$10 ahead, but now you are \$14 behind. How much should you expect to win or lose, on the average, if you place the bet many times?

The probability of winning a single-number bet is $\frac{1}{38}$, because there are thirty-eight numbers on the roulette wheel and only one of them is the subject of the bet. This means that if you place the bet a large number of times, it is most likely that

you will win once for every thirty-eight times you place the bet (and lose the other thirty-seven times). When you win, you win \$35, because the house odds are 35 to 1. When you lose, you lose \$1. Your average profit would be

$$\frac{\$35 + 37 \cdot (-\$1)}{38} = \frac{-\$2}{38} \approx -\$0.053$$

per game. This is called the *expected value* of a \$1 single-number bet, because you should expect to lose about a nickel for every dollar you bet if you play the game a long time. If you play a few times, anything could happen—you could win every single bet (though it is not likely). The house makes the bet so many times that it can be certain that its profit will be \$0.053 per dollar bet.

The standard way to find the **expected value** of an experiment is to multiply the value of each outcome of the experiment by the probability of that outcome and add the results. Here the experiment is placing a \$1 single-number bet in roulette. The outcomes are winning the bet and losing the bet. The values of the outcomes are +\$35 (if you win) and -\$1 (if you lose); the probabilities of the outcomes are $\frac{1}{38}$ and $\frac{37}{38}$, respectively (see Figure 3.20). The expected value would then be

$$35 \cdot \frac{1}{38} + (-1) \cdot \frac{37}{38} \approx -\$0.053$$

It is easy to see that this calculation is algebraically equivalent to the calculation done above.

Outcome	Value	Probability
winning	35	$\frac{1}{38}$
losing	-1	$\frac{37}{38}$

FIGURE 3.20 Finding the expected value.

Finding an expected value of a bet is very similar to finding your average test score in a class. Suppose that you are a student in a class in which you have taken four tests. If your scores were 80%, 76%, 90%, and 90%, your average test score would be

$$\frac{80 + 76 + 2 \cdot 90}{4} = 84\%$$

or, equivalently,

$$80 \cdot \frac{1}{4} + 76 \cdot \frac{1}{4} + 90 \cdot \frac{2}{4}$$

The difference between finding an average test score and finding the expected value of a bet is that with the average test score you are summarizing what *has* happened, whereas with a bet you are using probabilities to project what *will* happen.

EXPECTED VALUE

To find the **expected value** (or “long-term average”) of an experiment, multiply the value of each outcome of the experiment by its probability and add the results.



© Bill Bachman/Photo Edit

Roulette, the oldest casino game played today, has been popular since it was introduced to Paris in 1765. Does this game have any good bets?

EXAMPLE 1

COMPUTING AN EXPECTED VALUE By analyzing her sales records, a saleswoman has found that her weekly commissions have the probabilities in Figure 3.21. Find the saleswoman's expected commission.

Commission	0	\$100	\$200	\$300	\$400
Probability	0.05	0.15	0.25	0.45	0.1

FIGURE 3.21 Commission data for Example 1.

SOLUTION

To find the expected commission, we multiply each possible commission by its probability and add the results. Therefore,

$$\begin{aligned} \text{expected commission} &= (0)(0.05) + (100)(0.15) + (200)(0.25) \\ &\quad + (300)(0.45) + (400)(0.1) \\ &= 240 \end{aligned}$$

On the basis of her history, the saleswoman should expect to average \$240 per week in future commissions. Certainly, anything can happen in the future—she could receive a \$700 commission (it's not likely, though, because it has never happened before).

Why the House Wins

Four of the “best” bets that can be made in a casino game of chance are the pass, don’t pass, come, and don’t come bets in craps. They all have the same expected value, $-\$0.014$. In the long run, *there isn’t a single bet in any game of chance with which you can expect to break even, let alone make a profit*. After all, the casinos are out to make money. The expected values for \$1 bets in the more common games are shown in Figure 3.22.

Game	Expected Value of \$1 Bet
baccarat	$-\$0.014$
blackjack	$-\$0.06$ to $+\$0.10$ (varies with strategies)
craps	$-\$0.014$ for pass, don’t pass, come, don’t come bets <i>only</i>
slot machines	$-\$0.13$ to ? (varies)
keno (eight-spot ticket)	$-\$0.29$
average state lottery	$-\$0.48$

FIGURE 3.22 Expected values of common games of chance.

It is possible to achieve a positive expected value in blackjack. To do this, the player must keep a running count of the cards dealt, following a system that assigns a value to each card. Casinos use their pit bosses and video surveillance to watch for gamblers who use this tactic, and casinos will harass or kick such gamblers out when they find them. There is an application available for the iPhone and iPod Touch that will count cards. The casinos do everything in their power to eliminate its use. Some casinos use four decks at once and shuffle frequently to minimize the impact of counting.

Decision Theory

Which is the better bet: a \$1 single-number bet in roulette or a lottery ticket? Each costs \$1. The roulette bet pays \$35, but the lottery ticket might pay several million dollars. Lotteries are successful in part because the possibility of winning a large amount of money distracts people from the fact that winning is extremely unlikely. In Example 2 of Section 3.4, we found that the probability of winning first prize in many state lotteries is $\frac{1}{7,059,052} \approx 0.00000014$. At the beginning of this section, we found that the probability of winning the roulette bet is $\frac{1}{38} \approx 0.03$.

A more informed decision would take into account not only the potential winnings and losses but also their probabilities. The expected value of a bet does just that, since its calculation involves both the value and the probability of each outcome. We found that the expected value of a \$1 single-number bet in roulette is about $-\$0.053$. The expected value of the average state lottery is $-\$0.48$ (see Figure 3.22). The roulette bet is a much better bet than is the lottery. (Of course, there is a third option, which has an even better expected value of \$0.00: not gambling!)

A decision always involves choosing between various alternatives. If you compare the expected values of the alternatives, then you are taking into account

FEATURED IN THE NEWS

VIRGINIA LOTTERY HEDGES ON SYNDICATE'S BIG WIN

RICHMOND, VA.—Virginia lottery officials confirmed yesterday that an Australian gambling syndicate won last month's record \$27 million jackpot after executing a massive block-buying operation that tried to cover all 7 million possible ticket combinations.

But lottery director Kenneth Thorson said the jackpot may not be awarded because the winning ticket may have been bought in violation of lottery rules.

The rules say tickets must be paid for at the same location where they are issued. The Australian syndicate, International Lotto Fund, paid for many of its tickets at the corporate offices of Farm Fresh Inc. grocery stores, rather than at the Farm Fresh store in Chesapeake

where the winning ticket was issued, Thorson said.

"We have to validate who bought the ticket, where the purchase was made and how the purchase was made," Thorson said. "It's just as likely that we will honor the ticket as we won't honor the ticket." He said he may not decide until the end of next week.

Two Australians representing the fund, Joseph Franck and Robert Hans Roos, appeared at lottery headquarters yesterday to claim the prize.

The group succeeded in buying about 5 million of the more than 7 million possible numerical combinations before the February 15 drawing. The tactic is not illegal, although lottery officials announced new rules earlier this week aimed at making such block purchases more difficult.

The Australian fund was started last year and raised about \$13 million from an estimated 2,500 shareholders who each paid a minimum of \$4,000, according to Tim Phillipps of the Australian Securities Commission.

Half the money went for management expenses, much of that to Pacific Financial Resources, a firm controlled by Stefan Mandel, who won fame when he covered all the numbers in a 1986 Sydney lottery. Roos owns 10 percent of Pacific Financial Resources.

Australian Securities Commission officials said last week that the fund is under investigation for possible violations of Australian financial laws.

© 1992, The Washington Post, reprinted with permission.

the alternatives' potential advantages and disadvantages as well as their probabilities. This form of decision making is called **decision theory**.

EXAMPLE 2

USING EXPECTED VALUES TO MAKE A DECISION The saleswoman in Example 1 has been offered a new job that has a fixed weekly salary of \$290. Financially, which is the better job?

SOLUTION

In Example 1, we found that her expected weekly commission was \$240. The new job has a guaranteed weekly salary of \$290. Financial considerations indicate that she should take the new job.

Betting Strategies

One very old betting strategy is to "cover all the numbers." In 1729, the French philosopher and writer François Voltaire organized a group that successfully implemented this strategy to win the Parisian city lottery by buying most if not all of the tickets. Their strategy was successful because, owing to a series of poor financial decisions by the city of Paris, the total value of the prizes was greater than the combined price of all of the tickets! Furthermore, there were not a great number of tickets to buy. This strategy is still being used. (See the above newspaper article on its use in 1992 in Virginia.)

A **martingale** is a gambling strategy in which the gambler doubles his or her bet after each loss. A person using this strategy in roulette, concentrating on the

black numbers bet (which has 1-to-1 house odds), might lose three times before winning. This strategy would result in a net gain, as illustrated in Figure 3.23 below.

This seems to be a great strategy. Sooner or later, the player will win a bet, and because each bet is larger than the player’s total losses, he or she has to come out ahead! We will examine this strategy further in Exercises 47 and 48.

Bet Number	Bet	Result	Total Winnings/Losses
1	\$1	lose	−\$1
2	\$2	lose	−\$3
3	\$4	lose	−\$7
4	\$8	win	+\$1

FIGURE 3.23 Analyzing the martingale strategy.

3.5 EXERCISES

In Exercises 1–10, (a) find the expected value of each \$1 bet in roulette and (b) use the Law of Large Numbers to interpret it.

1. the two-number bet
- ▶ 2. the three-number bet
3. the four-number bet
- ▶ 4. the five-number bet
5. the six-number bet
- ▶ 6. the twelve-number bet
7. the low-number bet
- ▶ 8. the even-number bet
9. the red-number bet
- ▶ 10. the black-number bet
- ▶ 11. Using the expected values obtained in the text and in the preceding odd-numbered exercises, determine a casino’s expected net income from a 24-hour period at a single roulette table if the casino’s total overhead for the table is \$50 per hour and if customers place a total of \$7,000 on single-number bets, \$4,000 on two-number bets, \$4,000 on four-number bets, \$3,000 on six-number bets, \$7,000 on low-number bets, and \$8,000 on red-number bets.
- ▶ 12. Using the expected values obtained in the text and in the preceding even-numbered exercises, determine a casino’s expected net income from a 24-hour period at a single roulette table if the casino’s total overhead for the table is \$50 per hour and if customers place a total of \$8,000 on single-number bets, \$3,000 on three-number bets, \$4,000 on five-number bets, \$4,000 on

twelve-number bets, \$8,000 on even-number bets, and \$9,000 on black-number bets.

13. On the basis of his previous experience, the public librarian at Smallville knows that the number of books checked out by a person visiting the library has the probabilities shown in Figure 3.24.

Number of Books	0	1	2	3	4	5
Probability	0.15	0.35	0.25	0.15	0.05	0.05

FIGURE 3.24 Probabilities for Exercise 13.

Find the expected number of books checked out by a person visiting this library.

- ▶ 14. On the basis of his sale records, a salesman knows that his weekly commissions have the probabilities shown in Figure 3.25.

Commission	0	\$1,000	\$2,000	\$3,000	\$4,000
Probability	0.15	0.2	0.45	0.1	0.1

FIGURE 3.25 Probabilities for Exercise 14.

Find the salesman’s expected commission.

15. Of all workers at a certain factory, the proportions earning certain hourly wages are as shown in Figure 3.26.

▶ Selected exercises available online at www.webassign.net/brookscole

Hourly Wage	\$8.50	\$9.00	\$9.50	\$10.00	\$12.50	\$15.00
Proportion	20%	15%	25%	20%	15%	5%

FIGURE 3.26 Data for Exercise 15.

Find the expected hourly wage that a worker at this factory makes.

- ▶ 16. Of all students at the University of Metropolis, the proportions taking certain numbers of units are as shown in Figure 3.27. Find the expected number of units that a student at U.M. takes.

Units	3	4	5	6	7	8
Proportion	3%	4%	5%	6%	5%	4%

Units	9	10	11	12	13	14
Proportion	8%	12%	13%	13%	15%	12%

FIGURE 3.27 Data for Exercise 16.

17. You have been asked to play a dice game. It works like this:
- If you roll a 1 or 2, you win \$50.
 - If you roll a 3, you lose \$20.
 - If you roll a 4, 5, or 6, you lose \$30.
- Should you play the game? Use expected values to justify your answer.
18. You have been asked to play a dice game. It works like this:
- If you roll a 1, 2, 3, or 4, you win \$50.
 - If you roll a 5 or 6, you lose \$80.
- Should you play the game? Use expected values to justify your answer.
- ▶ 19. You are on a TV show. You have been asked to either play a dice game ten times or accept a \$100 bill. The dice game works like this:
- If you roll a 1 or 2, you win \$50.
 - If you roll a 3, you win \$20.
 - If you roll a 4, 5, or 6, you lose \$30.
- Should you play the game? Use expected values and decision theory to justify your answer.
- ▶ 20. You are on a TV show. You have been asked to either play a dice game five times or accept a \$50 bill. The dice game works like this:
- If you roll a 1, 2, or 3, you win \$50.
 - If you roll a 4 or 5, you lose \$25.
 - If you roll a 6, you lose \$90.
- Should you play the game? Use expected values and decision theory to justify your answer.

21. Show why the calculation at the top of page 182 is algebraically equivalent to the calculation in the middle of the same page.

22. In Example 1, the saleswoman's most likely weekly commission was \$300. With her new job (in Example 2), she will always make \$290 per week. This implies that she would be better off with the old job. Is this reasoning more or less valid than that used in Example 2? Why?

- ▶ 23. Maria just inherited \$10,000. Her bank has a savings account that pays 4.1% interest per year. Some of her friends recommended a new mutual fund, which has been in business for three years. During its first year, the fund went up in value by 10%; during the second year, it went down by 19%; and during its third year, it went up by 14%. Maria is attracted by the mutual fund's potential for relatively high earnings but concerned by the possibility of actually losing some of her inheritance. The bank's rate is low, but it is insured by the federal government. Use decision theory to find the best investment. (Assume that the fund's past behavior predicts its future behavior.)

24. Trang has saved \$8,000. It is currently in a bank savings account that pays 3.9% interest per year. He is considering putting the money into a speculative investment that would either earn 20% in one year if the investment succeeds or lose 18% in one year if it fails. At what probability of success would the speculative investment be the better choice?

- ▶ 25. Erica has her savings in a bank account that pays 4.5% interest per year. She is considering buying stock in a pharmaceuticals company that is developing a cure for cellulite. Her research indicates that she could earn 50% in one year if the cure is successful or lose 60% in one year if it is not. At what probability of success would the pharmaceuticals stock be the better choice?

- ▶ 26. Debra is buying prizes for a game at her school's fundraiser. The game has three levels of prizes, and she has already bought the second and third prizes. She wants the first prize to be nice enough to attract people to the game. The game's manufacturer has supplied her with the probabilities of winning first, second, and third prizes. Tickets cost \$3 each, and she wants the school to profit an average of \$1 per ticket. How much should she spend on each first prize?

Prize	Cost of Prize	Probability
1st	?	.15
2nd	\$1.25	.30
3rd	\$0.75	.45

FIGURE 3.28 Data for Exercise 26.

- ▶ 27. Few students manage to complete their schooling without taking a standardized admissions test such as

- the Scholastic Achievement Test, or S.A.T. (used for admission to college); the Law School Admissions Test, or L.S.A.T.; and the Graduate Record Exam, or G.R.E. (used for admission to graduate school). Sometimes, these multiple-choice tests discourage guessing by subtracting points for wrong answers. In particular, a correct answer will be worth +1 point, and an incorrect answer on a question with five listed answers (a through e) will be worth $-\frac{1}{4}$ point.
- Find the expected value of a random guess.
 - Find the expected value of eliminating one answer and guessing among the remaining four possible answers.
 - Find the expected value of eliminating three answers and guessing between the remaining two possible answers.
 - Use decision theory and your answers to parts (a), (b), and (c) to create a guessing strategy for standardized tests such as the S.A.T.
- ▶ 28. Find the expected value of a \$1 bet in six-spot keno if three winning spots pays \$1 (but you pay \$1 to play, so you actually break even), four winning spots pays \$3 (but you pay \$1 to play, so your profit is \$2), five pays \$100, and six pays \$2,600. (You might want to use the probabilities computed in Exercise 16 of Section 3.4.)
29. Find the expected value of a \$1 bet in eight-spot keno if four winning spots pays \$1 (but you pay \$1 to play, so you actually break even), five winning spots pays \$5 (but you pay \$1 to play, so your profit is \$4), six pays \$100, seven pays \$1,480, and eight pays \$19,000. (You might want to use the probabilities computed in Exercise 17 of Section 3.4.)
- ▶ 30. Find the expected value of a \$1 bet in nine-spot keno if five winning spots pays \$1 (but you pay \$1 to play, so you actually break even), six winning spots pays \$50 (but you pay \$1 to play, so your profit is \$49), seven pays \$390, eight pays \$6,000, and nine pays \$25,000. (You might want to use the probabilities computed in Exercise 18 of Section 3.4.)
31. Arizona's "Cash 4" is a 4/26 lottery. It differs from many other state lotteries in that its payouts are set; they do not vary with sales. If you match all four of the winning numbers, you win \$10,000 (but you pay \$1 to play, so your profit is \$9,999). If you match three of the winning numbers, you win \$25, and if you match two of the winning numbers, you win \$2. Otherwise, you lose your \$1.
- Find the probabilities of winning first prize, second prize, and third prize.
 - Use the results of parts a and the Complement Rule to find the probability of losing.
 - Use the results of parts a and b to find the expected value of Cash 4.
32. New York's "Pick 10" is a 10/80 lottery. Its payouts are set; they do not vary with sales. If you match all ten winning numbers, you win \$500,000 (but you pay \$1 to play, so your profit is \$499,999). If you match nine winning numbers, you win \$6000. If you match eight, seven, or six you win \$300, \$40, or \$10, respectively. If you match no winning numbers, you win \$4. Otherwise, you lose your \$1.
- Find the probabilities of winning first prize, second prize, and third prize.
 - Use the results of part a and the Complement Rule to find the probability of losing.
 - Use the results of parts a and b to find the expected value of Pick 10.
33. Arizona and New York have Pick 3 games (New York's is called "Numbers"), as described in Exercises 19 and 20 of Section 3.4. If the player opts for "straight play" and wins, she wins \$500 (but she pays \$1 to play, so her profit is \$499). If the player opts for "box play" with three different digits and wins, he wins \$80. If the player opts for "box play" with two of the same digits and wins, he wins \$160.



- Use the probabilities from Exercises 19 and 20 of Section 3.4 to find the probability of losing.
 - Use the probabilities from part a and from Exercises 19 and 20 of Section 3.4 to find the expected value of the game.
34. Write a paragraph in which you compare the states' fiscal policies concerning their lotteries with the casinos' fiscal policies concerning their keno games. Assume that the expected value of a \$1 lottery bet described in Exercises 31, 32, and/or 33 is representative of that of the other states' lotteries, and assume that the expected value of a \$1 keno bet described in Exercises 28, 29, and/or 30 is representative of other keno bets.
35. Trustworthy Insurance Co. estimates that a certain home has a 1% chance of burning down in any one year. They calculate that it would cost \$120,000 to rebuild that home. Use expected values to determine the annual insurance premium.
- ▶ 36. Mr. and Mrs. Trump have applied to the Trustworthy Insurance Co. for insurance on Mrs. Trump's diamond tiara. The tiara is valued at \$97,500. Trustworthy estimates that the jewelry has a 2.3% chance of being stolen in any one year. Use expected values to determine the annual insurance premium.
37. The Black Gold Oil Co. is considering drilling either in Jed Clampett's back yard or his front yard. After

thorough testing and analysis, they estimate that there is a 30% chance of striking oil in the back yard and a 40% chance in the front yard. They also estimate that the back yard site would either net \$60 million (if oil is found) or lose \$6 million (if oil is not found), and the front yard site would either net \$40 million or lose \$6 million. Use decision theory to determine where they should drill.

- ▶ 38. If in Exercise 37, Jed Clampett rejected the use of decision theory, where would he drill if he were an optimist? What would he do if he were a pessimist?
- ▶ 39. A community youth group is having a raffle to raise funds. Several community businesses have donated prizes. The prizes and their retail value are listed in Figure 3.29. Each prize will be given away, regardless of the number of raffle tickets sold. Tickets are sold for \$15 each. Determine the expected value of a ticket, and discuss whether it would be to your financial advantage to buy a ticket under the given circumstances.
- 1000 tickets are sold.
 - 2000 tickets are sold.
 - 3000 tickets are sold.

Prize	Retail Value	Number of These Prizes to Be Given Away
new car	\$21,580	1
a cell phone and a one-year subscription	\$940	1
a one-year subscription to an Internet service provider	\$500	2
dinner for two at Spiedini's restaurant	\$100	2
a one-year subscription to the local newspaper	\$180	20

FIGURE 3.29 Data for Exercise 39.

40. The Centerville High School PTA is having a raffle to raise funds. Several community businesses have donated prizes. The prizes and their retail value are listed in Figure 3.30. Each prize will be given away, regardless of the number of raffle tickets sold. Tickets are sold for \$30 each. Determine the expected value of a ticket, and discuss whether it would be to your financial advantage to buy a ticket under the given circumstances.
- 100 tickets are sold.
 - 200 tickets are sold.
 - 300 tickets are sold.

Prize	Retail Value	Number of These Prizes to Be Given Away
one week in a condo in Hawaii, and airfare for two	\$2,575	1
tennis lessons for two	\$500	1
a one-year subscription to an Internet service provider	\$500	2
dinner for two at Haute Stuff restaurant	\$120	2
a one-year subscription at the Centerville Skate Park	\$240	5
a copy of <i>Centerville Cooks</i> , a cookbook containing PTA members' favorite recipes	\$10	20

FIGURE 3.30 Data for Exercise 40.

41. The Central State University Young Republicans Club is having a raffle to raise funds. Several community businesses have donated prizes. The prizes and their retail value are listed in Figure 3.31. Each prize will be given away, regardless of the number of raffle tickets sold. Tickets are sold for \$5 each. Determine the expected value of a ticket, and discuss whether it would be to your financial advantage to buy a ticket under the given circumstances.
- 1000 tickets are sold.
 - 2000 tickets are sold.
 - 3000 tickets are sold.

Prize	Retail Value	Number of These Prizes to Be Given Away
laptop computer	\$2,325	1
MP3 player	\$425	2
20 CDs at Einstein Entertainment	\$320	3
a giant pizza and your choice of beverage at Freddie's Pizza	\$23	4
a one-year subscription to the <i>Young Republican Journal</i>	\$20	15

FIGURE 3.31 Data for Exercise 41.

42. The Southern State University Ecology Club is having a raffle to raise funds. Several community businesses have donated prizes. The prizes and their retail value are listed in Figure 3.32. Each prize will be given away, regardless of the number of raffle tickets sold. Tickets are sold for \$20 each. Determine the expected value of a ticket, and discuss whether it would be to your financial advantage to buy a ticket under the given circumstances.
- 100 tickets are sold.
 - 200 tickets are sold.
 - 300 tickets are sold.

Prize	Retail Value	Number of These Prizes to Be Given Away
one-week ecovacation to Costa Rica, including airfare for two	\$2,990	1
Sierra Designs tent	\$750	1
REI backpack	\$355	2
Jansport daypack	\$75	5
fleece jacket with Ecology Club logo	\$50	10
Ecology Club T-shirt	\$18	20

FIGURE 3.32 Data for Exercise 42.

- ▶ 43. Find the expected value of the International Lotto Fund's application of the "cover all of the numbers" strategy from the newspaper article on page 185. Assume that \$5 million was spent on lottery tickets, that half of the \$13 million raised went for management expenses, and that the balance was never spent. Also assume that Virginia honors the winning ticket.
- ▶ 44. One application of the "cover all the numbers" strategy would be to bet \$1 on every single number in roulette.
- Find the results of this strategy.
 - How could you use the expected value of the \$1 single-number bet ($-\frac{\$2}{38}$) to answer part (a)?
45. The application of the "cover all the numbers" strategy to a modern state lottery would involve the purchase of a large number of tickets.
- How many tickets would have to be purchased if you were in a state that has a 6/49 lottery (the player selects 6 out of 49 numbers)?
 - How much would it cost to purchase these tickets if each costs \$1?
 - If you organized a group of 100 people to purchase these tickets and it takes one minute to purchase each ticket, how many days would it take to purchase the required number of tickets?
- ▶ 46. The application of the "cover all the numbers" strategy to Keno would involve the purchase of a large number of tickets.
- How many tickets would have to be purchased if you were playing eight-spot keno?
 - How much would it cost to purchase these tickets if each costs \$5?
 - If it takes five seconds to purchase one ticket and the average keno game lasts twenty minutes, how many people would it take to purchase the required number of tickets?
47. If you had \$100 and were applying the martingale strategy to the black-number bet in roulette and you started with a \$1 bet, how many successive losses could you afford? How large would your net profit be if you lost each bet except for the last one?
48. If you had \$10,000 and were applying the martingale strategy to the black-number bet in roulette and you started with a \$1 bet, how many successive losses could you afford? How large would your net profit be if you lost each bet except for the last one?



Answer the following questions using complete sentences and your own words.

• CONCEPT QUESTIONS

49. Discuss the meaningfulness of the concept of expected value to three different people: an occasional gambler, a regular gambler, and a casino owner. To whom is the concept most meaningful? To whom is it least meaningful? Why?
50. Discuss the advantages and disadvantages of decision theory. Consider the application of decision theory to a nonrecurring situation and to a recurring situation. Also consider Jed Clampett and the Black Gold Oil Co. in Exercises 37 and 38.

• PROJECTS

51. Design a game of chance. Use probabilities and expected values to set the house odds so that the house will make a profit. However, be certain that your game is not so obviously pro-house that no one would be willing to play. Your project should include the following:
- a complete description of how the game is played
 - a detailed mathematical analysis of the expected value of the game (or of each separate bet in the game, whichever is appropriate)
 - a complete description of the bet(s) and the house odds

3.6

Conditional Probability

OBJECTIVES

- Use conditional probabilities to focus on one or two groups rather than the entire sample space
- Understand the relationship between intersections of events and products of probabilities
- Use tree diagrams to combine probabilities

Public opinion polls, such as those found in newspapers and magazines and on television, frequently categorize the respondents by such groups as sex, age, race, or level of education. This is done so that the reader or listener can make comparisons and observe trends, such as “people over 40 are more likely to support the Social Security system than are people under 40.” The tool that enables us to observe such trends is conditional probability.

Probabilities and Polls

In a newspaper poll concerning violence on television, 600 people were asked, “What is your opinion of the amount of violence on prime-time television—is there too much violence on television?” Their responses are indicated in Figure 3.33.

	Yes	No	Don't Know	Total
Men	162	95	23	280
Women	256	45	19	320
Total	418	140	42	600

FIGURE 3.33 Results of “Violence on Television” poll.

Six hundred people were surveyed in this poll; that is, the sample space consists of 600 responses. Of these, 418 said they thought there was too much violence on television, so the probability of a “yes” response is $\frac{418}{600}$, or about $0.70 = 70\%$. The probability of a “no” response is $\frac{140}{600}$, or about $0.23 = 23\%$.

If we are asked to find the probability that a *woman* responded yes, we do not consider all 600 responses but instead limit the sample space to only the responses from women. See Figure 3.34.

	Yes	No	Don't Know	Total
Women	256	45	19	320

FIGURE 3.34 Women and violence on television.

The probability that a woman responded yes is $\frac{256}{320} = 0.80 = 80\%$.



CBS-TV/TOUCHSTONE/THE KOBAL COLLECTION

Is there too much violence on TV?

Suppose we label the events in the following manner: W is the event that a response is from a woman, M is the event that a response is from a man, Y is the event that a response is yes, and N is the event that a response is no. Then the event that a woman responded yes would be written as

$$Y | W$$

The vertical bar stands for the phrase “given that”; the event $Y | W$ is read “a response is yes, given that the response is from a woman.” The probability of this event is called a *conditional probability*:

$$p(Y | W) = \frac{256}{320} = \frac{4}{5} = 0.80 = 80\%$$

The numerator of this probability, 256, is the number of responses that are yes and are from women; that is, $n(Y \cap W) = 256$. The denominator, 320, is the number of responses that are from women; that is, $n(W) = 320$. A **conditional probability** is a probability whose sample space has been limited to only those outcomes that fulfill a certain condition. Because an event is a subset of the sample space, the event must also fulfill that condition. The numerator of $p(Y | W)$ is 256 rather than 418 even though there were 418 “yes” responses, because many of those 418 responses were made by men; we are interested only in the probability that a woman responded yes.

CONDITIONAL PROBABILITY DEFINITION

The **conditional probability** of event A , given event B , is

$$p(A | B) = \frac{n(A \cap B)}{n(B)}$$

EXAMPLE 1

COMPUTING CONDITIONAL PROBABILITIES Using the data in Figure 3.33, find the following.

- the probability that a response is yes, given that the response is from a man.
- the probability that a response is from a man, given that the response is yes.
- the probability that a response is yes and is from a man.

SOLUTION

- Finding $p(Y | M)$:* We are told to consider only the male responses—that is, to limit our sample space to men. See Figure 3.35.

	Yes	No	Don't Know	Total
Men	162	95	23	280

FIGURE 3.35 Men and violence on television.

$$p(Y | M) = \frac{n(Y \cap M)}{n(M)} = \frac{162}{280} \approx 0.58 = 58\%$$

In other words, approximately 58% of the men responded yes. (Recall that 80% of the women responded yes. This poll indicates that men and women do not have the same opinion regarding violence on television and, in particular, that a woman is more likely to oppose the violence.)

- Finding $p(M | Y)$:* We are told to consider only the “yes” responses shown in Figure 3.36.

	Yes
Men	162
Women	256
Total	418

FIGURE 3.36

Limiting our sample space to “yes” responses.

$$p(M | Y) = \frac{n(M \cap Y)}{n(Y)} = \frac{162}{418} \approx 0.39 = 39\%$$

Therefore, of those who responded yes, approximately 39% were male.

- Finding $p(Y \cap M)$:* This is *not* a conditional probability (there is no vertical bar), so we do *not* limit our sample space.

$$p(Y \cap M) = \frac{n(Y \cap M)}{n(S)} = \frac{162}{600} = 0.27 = 27\%$$

Therefore, of all those polled, 27% were men who responded yes.

Each of the above three probabilities has the *same numerator*, $n(Y \cap M) = 162$. This is the number of responses that are yes and are from men.

But the three probabilities have *different denominators*. This means that we are comparing the group of men who said yes with three different larger groups. See Figure 3.37.

In the probability:	the denominator is:	so we're comparing the group of men who said yes with:
$p(Y M)$	$n(M)$, the number of male responses	all of the men
$p(M Y)$	$n(Y)$, the number of yes responses	all of the yes responses
$p(Y \cap M)$	$n(S)$, the number of responses	all of the people polled

FIGURE 3.37 The impact of the different denominators.

The Product Rule

If two cards are dealt from a full deck (no jokers), how would you find the probability that both are hearts? The probability that the first card is a heart is easy to find—it is $\frac{13}{52}$, because there are fifty-two cards in the deck and thirteen of them are hearts. The probability that the second card is a heart is more difficult to find. There are only fifty-one cards left in the deck (one was already dealt), but how many of these are hearts? The number of hearts left in the deck depends on the first card that was dealt. If it was a heart, then there are twelve hearts left in the deck; if it was not a heart, then there are thirteen hearts left. We could certainly say that the probability that the second card is a heart, *given that the first card was a heart*, is $\frac{12}{51}$.

Therefore, the probability that the first card is a heart is $\frac{13}{52}$, and the probability that the second card is a heart, given that the first was a heart, is $\frac{12}{51}$. How do we put these two probabilities together to find the probability that *both* the first and the second cards are hearts? Should we add them? Subtract them? Multiply them? Divide them?

The answer is obtained by algebraically rewriting the Conditional Probability Definition to obtain what is called the *Product Rule*:

$$\begin{aligned}
 p(A | B) &= \frac{n(A \cap B)}{n(B)} && \text{Conditional Probability Definition} \\
 p(A | B) \cdot n(B) &= n(A \cap B) && \text{multiplying by } n(B) \\
 \frac{p(A | B) \cdot n(B)}{n(S)} &= \frac{n(A \cap B)}{n(S)} && \text{dividing by } n(S) \\
 \frac{p(A | B)}{1} \cdot \frac{n(B)}{n(S)} &= \frac{n(A \cap B)}{n(S)} && \text{since } 1 \cdot n(S) = n(S) \\
 p(A | B) \cdot p(B) &= p(A \cap B) && \text{definition of probability}
 \end{aligned}$$

PRODUCT RULE

For any events A and B , the probability of A and B is

$$p(A \cap B) = p(A | B) \cdot p(B)$$

EXAMPLE 2

SOLUTION

USING THE PRODUCT RULE If two cards are dealt from a full deck, find the probability that both are hearts.

$$\begin{aligned}
 p(A \cap B) &= p(A | B) \cdot p(B) \\
 p(\text{2nd heart and 1st heart}) &= p(\text{2nd heart} | \text{1st heart}) \cdot p(\text{1st heart}) \\
 &= \frac{12}{51} \cdot \frac{13}{52} \\
 &= \frac{4}{17} \cdot \frac{1}{4} \\
 &= \frac{1}{17} \approx 0.06 = 6\%
 \end{aligned}$$

Therefore, there is a 6% probability that both cards are hearts.

Tree Diagrams

Many people find that a *tree diagram* helps them understand problems like the one in Example 2, in which an experiment is performed in stages over time. Figure 3.38 shows the tree diagram for Example 2. The first column gives a list of the possible outcomes of the first stage of the experiment; in Example 2, the first stage is dealing the first card, and its outcomes are “heart,” and “not a heart.” The branches leading to those outcomes represent their probabilities. The second column gives a list of the possible outcomes of the second stage of the experiment; in Example 2, the second stage is dealing the second card. A branch leading from a first-stage outcome to a second-stage outcome is the conditional probability $p(\text{2nd stage outcome} | \text{1st stage outcome})$.

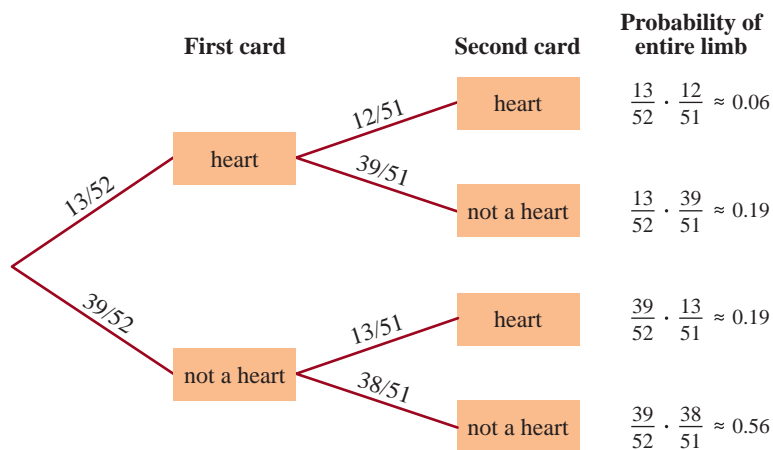


FIGURE 3.38 A tree diagram of dealing two hearts.

Looking at the top pair of branches, we see that the first branch stops at “first card is a heart” and the probability is $p(\text{1st heart}) = \frac{13}{52}$. The second branch starts at “first card is a heart” and stops at “second card is a heart” and gives the conditional probability $p(\text{2nd heart} | \text{1st heart}) = \frac{12}{51}$. The probability that we

were asked to calculate in Example 2, $p(\text{1st heart and 2nd heart})$, is that of the top limb:

$$\begin{aligned} p(\text{1st heart and 2nd heart}) &= p(\text{2nd heart} \mid \text{1st heart}) \cdot p(\text{1st heart}) \\ &= \frac{12}{51} \cdot \frac{13}{52} \end{aligned}$$

(We use the word *limb* to refer to a sequence of branches that starts at the beginning of the tree.) Notice that the sum of the probabilities of the four limbs is 1.00. Because the four limbs are the only four possible outcomes of the experiment, they must add up to 1.

Conditional probabilities always start at their condition, never at the beginning of the tree. For example, $p(\text{2nd heart} \mid \text{1st heart})$ is a conditional probability; its condition is that the first card is a heart. Thus, its branch starts at the box “first card is a heart.” However, $p(\text{1st heart})$ is not a conditional probability, so it starts at the beginning of the tree. Similarly, $p(\text{1st heart and 2nd heart})$ is not a conditional probability, so it too starts at the beginning of the tree. The product rule tells us that

$$p(\text{1st heart and 2nd heart}) = p(\text{2nd heart} \mid \text{1st heart}) \cdot p(\text{1st heart})$$

That is, the product rule tells us to multiply the branches that make up the top horizontal limb. In fact, “Multiply when moving horizontally across a limb” is a restatement of the product rule.

EXAMPLE 3

USING TREE DIAGRAMS AND THE MUTUALLY EXCLUSIVE RULE

Two cards are drawn from a full deck. Use the tree diagram in Figure 3.38 on page 195 to find the probability that the second card is a heart.

SOLUTION

The second card can be a heart if the first card is a heart *or* if it is not. The event “the second card is a heart” is the union of the following two mutually exclusive events:

$$E = \text{1st heart and 2nd heart}$$

$$F = \text{1st not heart and 2nd heart}$$

We previously used the tree diagram to find that

$$p(E) = \frac{13}{52} \cdot \frac{12}{51}$$

Similarly,

$$p(F) = \frac{39}{52} \cdot \frac{13}{51}$$

Thus, we add the probabilities of limbs that result in the second card being a heart:

$$\begin{aligned} p(\text{2nd heart}) &= p(E \cup F) \\ &= p(E) + p(F) && \text{the Mutually Exclusive Rule} \\ &= \frac{13}{52} \cdot \frac{12}{51} + \frac{39}{52} \cdot \frac{13}{51} = 0.25 \end{aligned}$$

In Example 3, the first and third limbs represent the only two ways in which the second card can be a heart. These two limbs represent mutually exclusive

events, so we used Probability Rule 5 [$p(E \cup F) = p(E) + p(F)$] to add their probabilities. In fact, “add when moving vertically from limb to limb” is a good restatement of Probability Rule 5.

TREE DIAGRAM SUMMARY

- Conditional probabilities start at their condition.
- Nonconditional probabilities start at the beginning of the tree.
- Multiply when moving horizontally across a limb.
- Add when moving vertically from limb to limb.

EXAMPLE 4

USING TREE DIAGRAMS, THE PRODUCT RULE, AND THE UNION/INTERSECTION RULE Big Fun Bicycles manufactures its product at two plants, one in Korea and one in Peoria. The Korea plant manufactures 60% of the bicycles; 4% of the Korean bikes are defective; and 5% of the Peorian bikes are defective.

- draw a tree diagram that shows this information.
- use the tree diagram to find the probability that a bike is defective and came from Korea.
- use the tree diagram to find the probability that a bike is defective.
- use the tree diagram to find the probability that a bike is defect-free.

SOLUTION

- First, we need to determine which probabilities have been given and find their complements, as shown in Figure 3.39.

Probabilities Given	Complements of These Probabilities
$p(\text{Korea}) = 60\% = 0.60$	$p(\text{Peoria}) = p(\text{not Korea}) = 1 - 0.60 = 0.40$
$p(\text{defective} \mid \text{Korea}) = 4\% = 0.04$	$p(\text{not defective} \mid \text{Korea}) = 1 - 0.04 = 0.96$
$p(\text{defective} \mid \text{Peoria}) = 5\% = 0.05$	$p(\text{not defective} \mid \text{Peoria}) = 1 - 0.05 = 0.95$

FIGURE 3.39 Probabilities for Example 4.

The first two of these probabilities [$p(\text{Korea})$ and $p(\text{Peoria})$] are not conditional, so they start at the beginning of the tree. The next two probabilities [$p(\text{defective} \mid \text{Korea})$ and $p(\text{not defective} \mid \text{Korea})$] are conditional, so they start at their condition (Korea). Similarly, the last two probabilities are conditional, so they start at their condition (Peoria). This placement of the probabilities yields the tree diagram in Figure 3.40.

- The probability that a bike is defective and came from Korea is a nonconditional probability, so it starts at the beginning of the tree. Do not confuse it with the conditional probability that a bike is defective, *given that* it came from Korea, which starts at its condition (Korea). The former is the limb that goes through “Korea” and stops at “defective”; the latter is one branch of that limb. We use the product rule to multiply when moving horizontally across a limb:

$$\begin{aligned}
 p(\text{defective and Korea}) &= p(\text{defective} \mid \text{Korea}) \cdot p(\text{Korea}) \\
 &= 0.04 \cdot 0.60 = 0.024
 \end{aligned}$$

Product Rule

This means that 2.4% of all of Big Fun’s bikes are defective bikes manufactured in Korea.

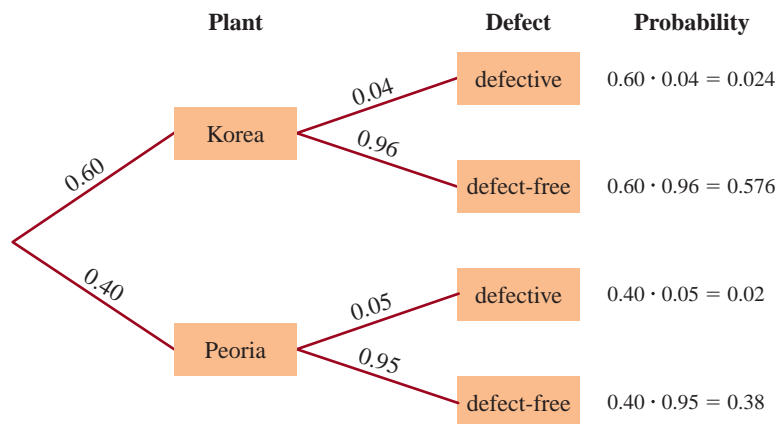


FIGURE 3.40 A tree diagram for Example 4.

c. The event that a bike is defective is the union of two mutually exclusive events:

- The bike is defective and came from Korea.
- The bike is defective and came from Peoria.

These two events are represented by the first and third limbs of the tree. We use the Union/Intersection Rule to add when moving vertically from limb to limb:

$$\begin{aligned}
 p(\text{defective}) &= p(\text{defective and Korea} \cup \text{defective and Peoria}) \\
 &= p(\text{defective and Korea}) + p(\text{defective and Peoria}) \\
 &= 0.024 + 0.02 = 0.044
 \end{aligned}$$

This means that 4.4% of Big Fun’s bicycles are defective.

d. The probability that a bike is defect-free is the complement of part (c).

$$p(\text{defect-free}) = p(\text{not defective}) = 1 - 0.044 = 0.956$$

Alternatively, we can find the sum of all the limbs that stop at “defect-free”:

$$p(\text{defect-free}) = 0.576 + 0.38 = 0.956$$

This means that 95.6% of Big Fun’s bicycles are defect-free.



TOPIC X HIV/AIDS: PROBABILITIES IN THE REAL WORLD

The human immunodeficiency virus (or HIV) is the virus that causes AIDS. The Centers for Disease Control estimates that in 2005, 950,000 Americans had HIV/AIDS. They are of both genders, all ages and sexual orientations. World-wide, nearly half of the 38 million people living with HIV/AIDS are women between 15 and 24 years old.

Most experts agree that the HIV/AIDS epidemic is in its early stages

and that a vaccine is not on the immediate horizon. The only current hope of stemming the infection lies in education and prevention. Education is particularly important, because an infected person can be symptom-free for eight years or more. You may well know many people who are unaware that they are infected, because they have no symptoms.

Health organizations such as the Centers for Disease Control and Prevention

routinely use probabilities to determine whether men or women are more likely to get HIV, which age groups are more likely to develop AIDS, and the different sources of exposure to HIV. See Exercises 49 and 50. In Section 3.7, we will investigate the accuracy of some HIV/AIDS tests. See Exercises 33 and 34 in Section 3.7.

3.6 EXERCISES

In Example 1, we wrote, “the probability that a response is yes, given that it is from a man” as $p(Y | M)$, and we wrote, “the probability that a response is yes and is from a man” as $p(Y \cap M)$. In Exercises 1–4, write the given probabilities in a similar manner. Also, identify which are conditional and which are not conditional.

- ▶ 1. Let H be the event that a job candidate is hired, and let Q be the event that a job candidate is well qualified. Use the symbols H , Q , $|$, and \cap to write the following probabilities.
 - a. The probability that a job candidate is hired given that the candidate is well qualified.
 - b. The probability that a job candidate is hired and the candidate is well qualified.
- 2. Let W be the event that a gambler wins a bet, and let L be the event that a gambler is feeling lucky. Use the symbols W , L , $|$, and \cap to write the following probabilities.
 - a. The probability that a gambler wins a bet and is feeling lucky.
 - b. The probability that a gambler wins a bet given that the gambler is feeling lucky.
- ▶ 3. Let S be the event that a cell phone user switches carriers, and let D be the event that a cell phone user gets dropped a lot. Use the symbols S , D , $|$, and \cap to write the following probabilities.
 - a. The probability that a cell phone user switches carriers given that she gets dropped a lot.
 - b. The probability that a cell phone user switches carriers and gets dropped a lot.
 - c. The probability that a cell phone user gets dropped a lot given that she switches carriers.
- ▶ 4. Let P be the event that a student passes the course, and S be the event that a student studies hard. Use the symbols P , S , $|$, and \cap to write the following probabilities.
 - a. The probability that a student passes the course given that the student studies hard.
 - b. The probability that a student studies hard given that the student passes the course.
 - c. The probability that a student passes the course and studies hard.

In Exercises 5–8, use Figure 3.41.

- ▶ 5. a. Find $p(B | A)$
b. Find $p(B \cap A)$
- 6. a. Find $p(B' | A)$
b. Find $p(B' \cap A)$
- 7. a. Find $p(B | A')$
b. Find $p(B \cap A')$
- ▶ 8. a. Find $p(B' | A')$
b. Find $p(B' \cap A')$

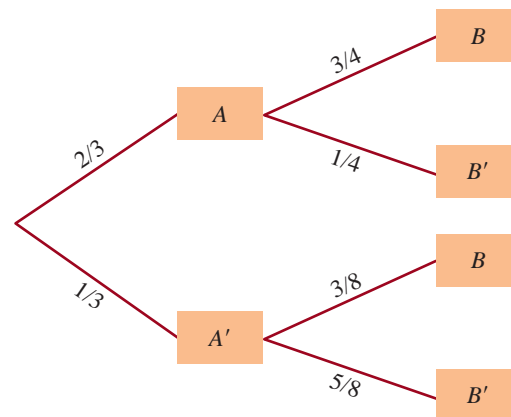


FIGURE 3.41 A tree diagram for Exercises 5–8.

- ▶ 9. Use the data in Figure 3.33 on page 191 to find the given probabilities. Also, write a sentence explaining what each means.
 - a. $p(N)$ b. $p(W)$ c. $p(N | W)$
 - d. $p(W | N)$ e. $p(N \cap W)$ f. $p(W \cap N)$
- 10. Use the data in Figure 3.33 on page 191 to find the given probabilities. Also, write a sentence explaining what each means.
 - a. $p(N)$ b. $p(M)$ c. $p(N | M)$
 - d. $p(M | N)$ e. $p(N \cap M)$ f. $p(M \cap N)$

Use the information in Figure 3.10 on page 154 to answer Exercises 11–14. Round your answers off to the nearest hundredth. Also interpret each of your answers using percentages and everyday English.

- 11. a. Find and interpret the probability that a U.S. resident dies of a pedestrian transportation accident, given that the person dies of a transportation accident in one year.
b. Find and interpret the probability that a U.S. resident dies of a pedestrian transportation accident, given that the person dies of a transportation accident, in a lifetime.
c. Find and interpret the probability that a U.S. resident dies of a pedestrian transportation accident, given that the person dies of a non-transportation accident, in a lifetime.
- 12. a. Find and interpret the probability that a U.S. resident dies from lightning, given that the person dies of a non-transportation accident, in one year.
b. Find and interpret the probability that a U.S. resident dies from lightning, given that the person dies of a non-transportation accident, in a lifetime.
c. Find and interpret the probability that a U.S. resident dies from lightning, given that the person dies of a transportation accident, in a lifetime.

- 13. a. Find and interpret the probability that a U.S. resident dies from an earthquake, given that the person dies from a non-transportation accident, in a lifetime.
- b. Find and interpret the probability that a U.S. resident dies from an earthquake, given that the person dies from a non-transportation accident, in one year.
- c. Find and interpret the probability that a U.S. resident dies from an earthquake, given that the person dies from an external cause, in one year.
- 14. a. Find and interpret the probability that a U.S. resident dies from a motorcyclist transportation accident, given that the person dies from a transportation accident, in a lifetime.
- b. Find and interpret the probability that a U.S. resident dies from a motorcyclist transportation accident, given that the person dies from a transportation accident, in one year.
- c. Find and interpret the probability that a U.S. resident dies from a motorcyclist transportation accident, given that the person dies from an external cause, in one year.

In Exercises 15–18, cards are dealt from a full deck of 52. Find the probabilities of the given events.

15. a. The first card is a club.
- b. The second card is a club, given that the first was a club.
- c. The first and second cards are both clubs.
- d. Draw a tree diagram illustrating this.
- 16. a. The first card is a king.
- b. The second card is a king, given that the first was a king.
- c. The first and second cards are both kings.
- d. Draw a tree diagram illustrating this. (Your diagram need not be a complete tree. It should have all the branches referred to in parts (a), (b), and (c), but it does not need other branches.)
17. a. The first card is a diamond.
- b. The second card is a spade, given that the first was a diamond.
- c. The first card is a diamond and the second is a spade.
- d. Draw a tree diagram illustrating this. (Your diagram need not be a complete tree. It should have all the branches referred to in parts (a), (b), and (c), but it does not need other branches.)
- 18. a. The first card is a jack.
- b. The second card is an ace, given that the first card was a jack.
- c. The first card is a jack and the second is an ace.
- d. Draw a tree diagram illustrating this.

*In Exercises 19 and 20, determine which probability the indicated branch in Figure 3.42 refers to. For example, the branch labeled * refers to the probability $p(A)$.*

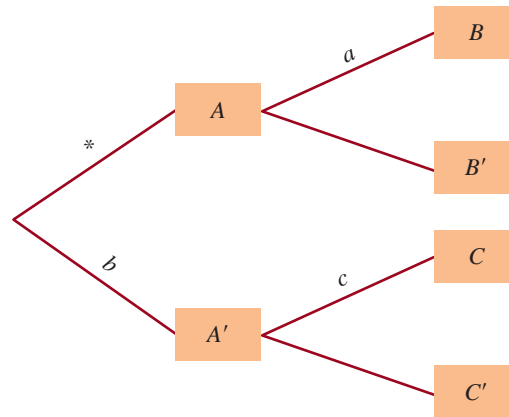


FIGURE 3.42 A tree diagram for Exercises 19 and 20.

19. a. the branch labeled a
- b. the branch labeled b
- c. the branch labeled c
20. a. Should probabilities $(*)$ and (a) be added or multiplied? What rule tells us that? What is the result of combining them?
- b. Should probabilities (b) and (c) be added or multiplied? What rule tells us that? What is the result of combining them?
- c. Should the probabilities that result from parts (a) and (b) of this exercise be added or multiplied? What rule tells us that? What is the result of combining them?

In Exercises 21 and 22, a single die is rolled. Find the probabilities of the given events.

- 21. a. rolling a 6
- b. rolling a 6, given that the number rolled is even
- c. rolling a 6, given that the number rolled is odd
- d. rolling an even number, given that a 6 was rolled
22. a. rolling a 5
- b. rolling a 5, given that the number rolled is even
- c. rolling a 5, given that the number rolled is odd
- d. rolling an odd number, given that a 5 was rolled

In Exercises 23–26, a pair of dice is rolled. Find the probabilities of the given events.

23. a. The sum is 6.
- b. The sum is 6, given that the sum is even.
- c. The sum is 6, given that the sum is odd.
- d. The sum is even, given that the sum is 6.
- 24. a. The sum is 12.
- b. The sum is 12, given that the sum is even.
- c. The sum is 12, given that the sum is odd.
- d. The sum is even, given that the sum was 12.

- ▶ 25. a. The sum is 4.
b. The sum is 4, given that the sum is less than 6.
c. The sum is less than 6, given that the sum is 4.
26. a. The sum is 11.
b. The sum is 11, given that the sum is greater than 10.
c. The sum is greater than 10, given that the sum is 11.
27. A single die is rolled. Determine which of the following events is least likely and which is most likely. Do so without making any calculations. Explain your reasoning.
 E_1 is the event “rolling a 4.”
 E_2 is the event “rolling a 4, given that the number rolled is even.”
 E_3 is the event “rolling a 4, given that the number rolled is odd.”
28. A pair of dice is rolled. Determine which of the following events is least likely and which is most likely. Do so without making any calculations. Explain your reasoning.
 E_1 is the event “rolling a 7.”
 E_2 is the event “rolling a 7, given that the number rolled is even.”
 E_3 is the event “rolling a 7, given that the number rolled is odd.”

In Exercises 29 and 30, use the following information. To determine what effect the salespeople had on purchases, a department store polled 700 shoppers as to whether or not they had made a purchase and whether or not they were pleased with the service. Of those who had made a purchase, 125 were happy with the service and 111 were not. Of those who had made no purchase, 148 were happy with the service and 316 were not.

- ▶ 29. Find the probability that a shopper who was happy with the service had made a purchase (round off to the nearest hundredth). What can you conclude?
30. Find the probability that a shopper who was unhappy with the service had not made a purchase. (Round off to the nearest hundredth.) What can you conclude?

In Exercises 31–34, five cards are dealt from a full deck. Find the probabilities of the given events. (Round off to four decimal places.)

31. All are spades.
32. The fifth is a spade, given that the first four were spades.
33. The last four are spades, given that the first was a spade.
34. All are the same suit.

In Exercises 35–40, round off to the nearest hundredth.

- ▶ 35. If three cards are dealt from a full deck, use a tree diagram to find the probability that exactly two are spades.

36. If three cards are dealt from a full deck, use a tree diagram to find the probability that exactly one is a spade.
37. If three cards are dealt from a full deck, use a tree diagram to find the probability that exactly one is an ace.
38. If three cards are dealt from a full deck, use a tree diagram to find the probability that exactly two are aces.
39. If a pair of dice is rolled three times, use a tree diagram to find the probability that exactly two throws result in 7's.
▶ 40. If a pair of dice is rolled three times, use a tree diagram to find the probability that all three throws result in 7's.

In Exercises 41–44, use the following information: A personal computer manufacturer buys 38% of its chips from Japan and the rest from the United States. Of the Japanese chips, 1.7% are defective, and 1.1% of the American chips are defective.

- ▶ 41. Find the probability that a chip is defective and made in Japan.
42. Find the probability that a chip is defective and made in the United States.
▶ 43. Find the probability that a chip is defective.
▶ 44. Find the probability that a chip is defect-free.
45. The results of CNN's 2008 presidential election poll are given in Figure 3.43.

	Obama	McCain	Other/No Answer
Male Voters	23.0%	22.6%	1.4%
Female Voters	29.7%	22.8%	0.5%

FIGURE 3.43 Exit poll results, by gender. *Source:* CNN.

- a. Find the probability that a voter voted for Obama, given that the voter is male.
b. Find the probability that a voter voted for Obama, given that the voter is female.
c. What observations can you make?
46. Use the information in Figure 3.43 to answer the following questions.
a. Find the probability that a voter is male, given that the voter voted for Obama.
b. Find the probability that a voter is female, given that the voter voted for Obama.
c. What observations can you make?
47. The results of CNN's 2008 presidential election poll are given in Figure 3.44 on page 202.
a. Find that probability that a voter voted for McCain, given that the voter is under 45.
b. Find the probability that a voter voted for McCain, given that the voter is 45 or over.
c. What observations can you make?

Age of Voter	Obama	McCain	Other/No Answer
18–29	11.9%	5.8%	0.4%
30–44	15.1%	13.3%	0.6%
45–64	18.5%	18.1%	0.4%
65 and Older	7.2%	8.5%	0.3%

FIGURE 3.44 Exit poll results, by age. *Source:* CNN.

- ▶ **48.** Use the information in Figure 3.44 to answer the following questions.
 - a. Find the probability that a voter is under 45, given that the voter voted for McCain.
 - b. Find the probability that a voter is 45 or over, given that the voter voted for McCain.
 - c. What observations can you make?
- ▶ **49.** Figure 3.45 gives the estimated number of diagnoses of AIDS among adults and adolescents in the United States by transmission category through the year 2007.

Transmission Category	Estimated Number of AIDS Cases, through 2007		
	Adult and Adolescent Male	Adult and Adolescent Female	Total
Male-to-male sexual contact	487,695	—	487,695
Injection drug use	175,704	80,155	255,859
Male-to-male sexual contact and injection drug use	71,242	—	71,242
High-risk heterosexual contact	63,927	112,230	176,157
Other*	12,108	6,158	18,266

FIGURE 3.45 AIDS sources.

*Includes hemophilia, blood transfusion, perinatal exposure, and risk not reported or not identified. *Source:* Centers for Disease Control and Prevention.

- a. Find the probability that a U.S. adult or adolescent diagnosed with AIDS is male and the probability that a U.S. adult or adolescent diagnosed with AIDS is female.
- b. Find the probability that a male U.S. adult or adolescent diagnosed with AIDS was exposed by injection drug use (possibly combined with male-to-male sexual contact) and the probability that a female U.S. adult or adolescent diagnosed with AIDS was exposed by injection drug use.
- c. Find the probability that a U.S. adult or adolescent diagnosed with AIDS is male and was exposed by injection drug use (possibly combined with

male-to-male sexual contact) and the probability that a U.S. adult or adolescent diagnosed with AIDS is female and was exposed by injection drug use.

- d. Find the probability that a male U.S. adult or adolescent diagnosed with AIDS was exposed by heterosexual contact and the probability that a female U.S. adult or adolescent diagnosed with AIDS was exposed by heterosexual contact.
- e. Find the probability that a U.S. adult or adolescent diagnosed with AIDS is male and was exposed by heterosexual contact and the probability that a U.S. adult or adolescent diagnosed with AIDS is female and was exposed by heterosexual contact.
- f. Explain the difference between parts (b) and (c) and the difference between parts (d) and (e).
- ▶ **50.** Figure 3.46 gives the estimated number of diagnoses of AIDS in the United States by age at the time of diagnosis, through 2007. Also note that in 2007, there were 301,621,157 residents of the United States, according to the U.S. Census Bureau.

Age (Years)	Cumulative Number of AIDS Cases
Under 13	9,209
13–14	1,169
15–24	44,264
25–34	322,370
35–44	396,851
45–54	176,304
55–64	52,409
65 or older	15,853

FIGURE 3.46 AIDS by age. *Source:* Centers for Disease Control and Prevention.

- a. Find the probability that a U.S. resident diagnosed with AIDS was 15 to 24 years old at the time of diagnosis.
- b. Find the probability that a U.S. resident was diagnosed with AIDS and was 15 to 24 years old at the time of diagnosis.
- c. Find the probability that a U.S. resident diagnosed with AIDS was 25 to 34 years old at the time of diagnosis.
- d. Find the probability that a U.S. resident was diagnosed with AIDS and was 25 to 34 years old at the time of diagnosis.
- e. Explain in words the difference between parts (c) and (d).
- 51.** In November 2007, the National Center for Health Statistics published a document entitled “Obesity Among Adults in the United States.” According to that

document, 32% of adult men and 35% of adult women in the United States were obese in 2006. At that time, there were 148 million adult men and 152 million adult women in the country.

Source: National Center for Health Statistics, C. Ogden et al: Obesity Among Adults in the United States.

- Find the probability that an adult man was obese.
 - Find the probability that an adult was an obese man.
 - Find the probability that an adult woman was obese.
 - Find the probability that an adult was an obese woman.
 - Find the probability that an adult was obese.
 - Explain in words the difference between parts (c) and (d).
52. The document mentioned in Exercise 51 also says that 23.2% of adult men and 29.5% of adult women in the United States were overweight.
- Find the probability that an adult man was overweight.
 - Find the probability that an adult was an overweight man.
 - Find the probability that an adult woman was overweight.
 - Find the probability that an adult was an overweight woman.
 - Find the probability that an adult was overweight.
- 53. In 1981 a study on race and the death penalty was released. The data in Figure 3.47 are from that study.

Death penalty imposed?	Victim's race	Defendant's race	Frequency
Yes	White	White	19
Yes	White	Black	11
Yes	Black	White	0
Yes	Black	Black	6
No	White	White	132
No	White	Black	152
No	Black	White	9
No	Black	Black	97

FIGURE 3.47 Race and the death penalty. Source: M. Radelet (1981) "Racial Characteristics and the Imposition of the Death Penalty." American Sociological Review, 46, 918–927.

- Find $p(\text{death penalty imposed} \mid \text{victim white and defendant white})$.
- Find $p(\text{death penalty imposed} \mid \text{victim white and defendant black})$.
- Find $p(\text{death penalty imposed} \mid \text{victim black and defendant white})$.

- Find $p(\text{death penalty imposed} \mid \text{victim black and defendant black})$.
- What can you conclude from parts (a) through (d)?
- Determine what other conditional probabilities would affect this issue, and calculate those probabilities.
- Discuss your results.

- 54. The information in Exercise 53 is rather dated. Use the following information to determine whether things have improved since then. In April 2003, Amnesty International issued a study on race and the death penalty. The following quote is from that study:

"The population of the USA is approximately 75 percent white and 12 percent black. Since 1976, blacks have been six to seven times more likely to be murdered than whites, with the result that blacks and whites are the victims of murder in about equal numbers. Yet, 80 percent of the more than 840 people put to death in the USA since 1976 were convicted of crimes involving white victims, compared to the 13 percent who were convicted of killing blacks. Less than four percent of the executions carried out since 1977 in the USA were for crimes involving Hispanic victims. Hispanics represent about 12 percent of the US population. Between 1993 and 1999, the recorded murder rate for Hispanics was more than 40 percent higher than the national homicide rate."

(Source: Amnesty International. "United States of America: Death by discrimination—the continuing role of race in capital cases." AMR 51/046/2003)

- In the above quote, what conditional probability is 80%? That is, find events A and B such that $p(A \mid B) = 80\%$.
- In the above quote, what conditional probability is 13%? That is, find events C and D such that $p(C \mid D) = 13\%$.
- Use the information in Exercise 53 to compute $p(A \mid B)$ for 1981.
- Use the information in Exercise 53 to compute $p(C \mid D)$ for 1981.
- Have things improved since 1981? Justify your answer.

55. A man and a woman have a child. Both parents have sickle-cell trait. They know that their child does not have sickle-cell anemia because he shows no symptoms, but they are concerned that he might be a carrier. Find the probability that he is a carrier.

- 56. A man and a woman have a child. Both parents are Tay-Sachs carriers. They know that their child does not have Tay-Sachs disease because she shows no symptoms, but they are concerned that she might be a carrier. Find the probability that she is a carrier.

In Exercises 57–60, use the following information. The University of Metropolis requires its students to pass an examination in college-level mathematics before they can graduate. The students

are given three chances to pass the exam; 61% pass it on their first attempt, 63% of those that take it a second time pass it then, and 42% of those that take it a third time pass it then.

- 57. What percentage of the students pass the exam?
- ▶ 58. What percentage of the students are not allowed to graduate because of their performance on the exam?
- 59. What percentage of the students take the exam at least twice?
- ▶ 60. What percent of the students take the test three times?
- 61. In the game of blackjack, if the first two cards dealt to a player are an ace and either a ten, jack, queen, or king, then the player has a blackjack, and he or she wins. Find the probability that a player is dealt a blackjack out of a full deck (no jokers).
- 62. In the game of blackjack, the dealer's first card is dealt face up. If that card is an ace, then the player has the option of "taking insurance." "Insurance" is a side bet. If the dealer has blackjack, the player wins the insurance bet and is paid 2 to 1 odds. If the dealer does not have a blackjack, the player loses the insurance bet. Find the probability that the dealer is dealt a blackjack if his or her first card is an ace.
- 63. Use the data in Figure 3.33 to find the following probabilities, where N is the event "saying no," and W is the event "being a woman":
 - a. $p(N' | W)$ b. $p(N | W')$ c. $p(N' | W')$
 - d. Which event, $N' | W$, $N | W'$, or $N' | W'$, is the complement of the event $N | W$? Why?
- ▶ 64. Use the data in Figure 3.33 to find the following probabilities, where Y is the event "saying yes," and M is the event "being a man.":
 - a. $p(Y | M)$ b. $p(Y | M')$ c. $p(Y' | M')$
 - d. Which event, $Y' | M$, $Y | M'$, or $Y' | M'$, is the complement of the event $Y | M$? Why?
- 65. If A and B are arbitrary events, what is the complement of the event $A | B$?
- 66. Show that $p(A | B) = \frac{p(A \cap B)}{p(B)}$.
HINT: Divide the numerator and denominator of the Conditional Probability Definition by $n(S)$.
- 67. Use Exercise 66 and appropriate answers from Exercise 9 to find $P(N | W)$.
- 68. Use Exercise 66 and appropriate answers from Exercise 10 to find $P(Y | M)$.



Answer the following questions using complete sentences and your own words.

• CONCEPT QUESTIONS

- 69. Which must be true for any events A and B ?
 - $P(A | B)$ is always greater than or equal to $P(A)$.
 - $P(A | B)$ is always less than or equal to $P(A)$.
 - Sometimes $P(A | B)$ is greater than or equal to $P(A)$, and sometimes $P(A | B)$ is less than or equal to $P(A)$, depending on the events A and B .

Answer this without making any calculations. Explain your reasoning.
- 70. Compare and contrast the events A , $A | B$, $B | A$, and $A \cap B$. Be sure to discuss both the similarities and the differences between these events.

WEB PROJECTS

- 71. There are many different blood type systems, but the ABO and Rh systems are the most important systems for blood donation purposes. These two systems generate eight different blood types: $A+$, $A-$, $B+$, $B-$, $AB+$, $AB-$, $O+$, and $O-$.
 - a. For each of these eight blood types, use the web to determine the percentage of U.S. residents that have that blood type. Interpret these percentages as probabilities. Use language like "the probability that a randomly-selected U.S. resident . . ."
 - b. You can always give blood to someone with the same blood type. In some cases, you can give blood to someone with a different blood type, or receive blood from someone with a different blood type. This depends on the donor's blood type and the receiver's blood type. Use the web to complete the chart in Figure 3.48.
 - c. For each of the eight blood types, use the web to determine the percentage of U.S. residents that can donate blood to a person of that blood type. Interpret these percentages as conditional probabilities.
 - d. For each of the eight blood types, use the web to determine the percentage of U.S. residents that can receive blood from a person of that blood type. Interpret these percentages as conditional probabilities.

Blood group	A+	A-	B+	B-	AB+	AB-	O+	O-
Can donate blood to								
Can receive blood from								

FIGURE 3.48 Blood types.

- e. Are the probabilities in parts (a), (c), and (d) theoretical probabilities or relative frequencies? Why?

Some useful links for this web project are listed on the text web site:

www.cengage.com/math/johnson

72. According to the U.S. Department of Transportation's National Highway Traffic Safety Administration, "rollovers are dangerous incidents and have a higher fatality rate than other kinds of crashes. Of the nearly 11 million passenger car, SUV, pickup, and van crashes in 2002, only 3% involved a rollover. However, rollovers accounted for nearly 33% of all deaths from passenger vehicle crashes."

(Source: <http://www.safercar.gov/Rollover/pages/RolloCharFat.htm>)

Use probabilities and the web to investigate rollovers. How likely is a rollover if you are driving a sedan, an SUV, or a van? Which models have the highest probability of a rollover? Which models have the lowest probability? Wherever possible, give specific conditional probabilities.

Some useful links for this web project are listed on the text web site:

www.cengage.com/math/johnson

• PROJECTS

73. In 1973, the University of California at Berkeley admitted 1,494 of 4,321 female applicants for graduate study and 3,738 of 8,442 male applicants.

(Source: P.J. Bickel, E.A. Hammel, and J.W. O'Connell, "Sex Bias in Graduate Admissions: Data from Berkeley," *Science*, vol. 187, 7 February 1975.)

- a. Find the probability that:
- an applicant was admitted
 - an applicant was admitted, given that he was male
 - an applicant was admitted, given that she was female

Discuss whether these data indicate a possible bias against women.

- b. Berkeley's graduate students are admitted by the department to which they apply rather than by a campuswide admissions panel. When $P(\text{admission} |$

male) and $P(\text{admission} | \text{female})$ were computed for each of the school's more than 100 departments, it was found that in four departments, $P(\text{admission} | \text{male})$ was greater than $P(\text{admission} | \text{female})$ by a significant amount and that in six departments, $P(\text{admission} | \text{male})$ was less than $P(\text{admission} | \text{female})$ by a significant amount. Discuss whether this information indicates a possible bias against women and whether it is consistent with that in part (a).

- c. The authors of "Sex Bias in Graduate Admissions: Data from Berkeley" attempt to explain the paradox by discussing an imaginary school with only two departments: "machismatics" and "social warfare." Machismatics admitted 200 of 400 male applicants for graduate study and 100 of 200 female applicants, while social warfare admitted 50 of 150 male applicants for graduate study and 150 of 450 female applicants. For *the school as a whole and for each of the two departments*, find the probability that:
- an applicant was admitted
 - an applicant was admitted, given that he was male
 - an applicant was admitted, given that she was female

Discuss whether these data indicate a possible bias against women.

- d. Explain the paradox illustrated in parts (a)–(c).
e. What conclusions would you make, and what further information would you obtain, if you were an affirmative action officer at Berkeley?

74. a. Use the data in Figure 3.49 to find the probability that:
- a New York City resident died from tuberculosis
 - a Caucasian New York City resident died from tuberculosis
 - a non-Caucasian New York City resident died from tuberculosis
 - a Richmond resident died from tuberculosis
 - a Caucasian Richmond resident died from tuberculosis
 - a non-Caucasian Richmond resident died from tuberculosis
- b. What conclusions would you make, and what further information would you obtain, if you were a public health official?

	New York City		Richmond, Virginia	
	Population	TB deaths	Population	TB deaths
Caucasian	4,675,000	8400	81,000	130
Non-Caucasian	92,000	500	47,000	160

FIGURE 3.49 Tuberculosis deaths by race and location, 1910.

Source: Morris R. Cohen and Ernest Nagel, *An Introduction to Logic and Scientific Method* (New York: Harcourt Brace & Co, 1934.)

3.7

Independence; Trees in Genetics

OBJECTIVES

- Understand the difference between dependent and independent events
- Know the effect of independence on the Product Rule
- Be able to apply tree diagrams in medical and genetic situations

Dependent and Independent Events

Consider the dealing of two cards from a full deck. An observer who saw that the first card was a heart would be better able to predict whether the second card will be a heart than would another observer who did not see the first card. If the first card was a heart, there is one fewer heart in the deck, so it is slightly less likely that the second card will be a heart. In particular,

$$p(\text{2nd heart} \mid \text{1st heart}) = \frac{12}{51} \approx 0.24$$

whereas, as we saw in Example 3 of Section 3.6,

$$p(\text{2nd heart}) = 0.25$$

These two probabilities are different because of the effect the first card drawn has on the second. We say that the two events “first card is a heart” and “second card is a heart” are *dependent*; the result of dealing the second card depends, to some extent, on the result of dealing the first card. In general, two events E and F are **dependent** if $p(E \mid F) \neq p(E)$.

Consider two successive tosses of a single die. An observer who saw that the first toss resulted in a three would be *no better able* to predict whether the second toss will result in a three than another observer who did not observe the first toss. In particular,

$$p(\text{2nd toss is a three}) = \frac{1}{6}$$

and

$$p(\text{2nd toss is a three} \mid \text{1st toss was a three}) = \frac{1}{6}$$

These two probabilities are the same, because the first toss has no effect on the second toss. We say that the two events “first toss is a three” and “second toss is a three” are *independent*; the result of the second toss does *not* depend on the result of the first toss. In general, two events E and F are **independent** if $p(E \mid F) = p(E)$.

INDEPENDENCE/DEPENDENCE DEFINITIONS

Two events E and F are **independent** if $p(E \mid F) = p(E)$.
(Think: Knowing F does not affect E 's probability.)

Two events E and F are **dependent** if $p(E \mid F) \neq p(E)$.
(Think: Knowing F does affect E 's probability.)

Many people have difficulty distinguishing between *independent* and *mutually exclusive*. (Recall that two events E and F are mutually exclusive if $E \cap F = \emptyset$, that is, if one event excludes the other.) This is probably because the relationship between mutually exclusive events and the relationship between independent events both could be described, in a very loose sort of way, by saying that “the two events have nothing to do with each other.” *Never think this way*; mentally replacing “mutually exclusive” or “independent” with “having nothing to do with each other” only obscures the distinction between these two concepts. E and F are independent if knowing that F has occurred *does not* affect the probability that E will occur. E and F are dependent if knowing that F has occurred *does* affect the probability that E will occur. E and F are mutually exclusive if E and F cannot occur simultaneously.

EXAMPLE 1

INDEPENDENT EVENTS AND MUTUALLY EXCLUSIVE EVENTS Let F be the event “a person has freckles,” and let R be the event “a person has red hair.”

- Are F and R independent?
- Are F and R mutually exclusive?

SOLUTION

- F and R are independent if $p(F | R) = p(F)$. With $p(F | R)$, we are given that a person has red hair; with $p(F)$, we are not given that information. Does knowing that a person has red hair affect the probability that the person has freckles? Yes, it does; $p(F | R) > p(F)$. Therefore, F and R are not independent; they are dependent.
- F and R are mutually exclusive if $F \cap R = \emptyset$. Many people have both freckles and red hair, so $F \cap R \neq \emptyset$, and F and R are not mutually exclusive. In other words, having freckles does not exclude the possibility of having red hair; freckles and red hair can occur simultaneously.

EXAMPLE 2

INDEPENDENT EVENTS AND MUTUALLY EXCLUSIVE EVENTS Let T be the event “a person is tall,” and let R be the event “a person has red hair.”

- Are T and R independent?
- Are T and R mutually exclusive?

SOLUTION

- T and R are independent if $p(T | R) = p(T)$. With $p(T | R)$, we are given that a person has red hair; with $p(T)$, we are not given that information. Does knowing that a person has red hair affect the probability that the person is tall? No, it does not; $p(T | R) = p(T)$, so T and R are independent.
- T and R are mutually exclusive if $T \cap R = \emptyset$. $T \cap R$ is the event “a person is tall and has red hair.” There are tall people who have red hair, so $T \cap R \neq \emptyset$, and T and R are not mutually exclusive. In other words, being tall does not exclude the possibility of having red hair; being tall and having red hair can occur simultaneously.

In Examples 1 and 2, we had to rely on our personal experience in concluding that knowledge that a person has red hair does affect the probability that he or she has freckles and does not affect the probability that he or she is tall. It may be the case that you have seen only one red-haired person and she was short and without freckles. Independence is better determined by computing the appropriate probabilities than by relying on one’s own personal experiences. This is especially crucial in determining the effectiveness of an experimental drug. *Double-blind* experiments, in which neither the patient nor the doctor knows whether the given medication is the experimental drug or an inert substance, are often done to ensure reliable, unbiased results.

Independence is an important tool in determining whether an experimental drug is an effective vaccine. Let D be the event that the experimental drug was

administered to a patient, and let R be the event that the patient recovered. It is hoped that $p(R | D) > p(R)$, that is, that the rate of recovery is greater among those who were given the drug. In this case, R and D are dependent. Independence is also an important tool in determining whether an advertisement effectively promotes a product. An ad is effective if $p(\text{consumer purchases product} | \text{consumer saw ad}) > p(\text{consumer purchases product})$.

EXAMPLE 3

SOLUTION

DETERMINING INDEPENDENCE Use probabilities to determine whether the events “thinking there is too much violence in television” and “being a man” in Example 1 of Section 3.6 are independent.

Two events E and F are independent if $p(E | F) = p(E)$. The events “responding yes to the question on violence in television” and “being a man” are independent if $p(Y | M) = p(Y)$. We need to compute these two probabilities and compare them.

In Example 1 of Section 3.6, we found $p(Y | M) \approx 0.58$. We can use the data from the poll in Figure 3.33 to find $p(Y)$.

$$p(Y) = \frac{418}{600} \approx 0.70$$

$$p(Y | M) \neq p(Y)$$

The events “responding yes to the question on violence in television” and “being a man” are dependent. According to the poll, men are less likely to think that there is too much violence on television.

In Example 3, what should we conclude if we found that $p(Y) = 0.69$ and $p(Y | M) = 0.67$? Should we conclude that $p(Y | M) \neq p(Y)$ and that the events “thinking that there is too much violence on television” and “being a man” are dependent? Or should we conclude that $p(Y | M) \approx p(Y)$ and that the events are (probably) independent? In this particular case, the probabilities are relative frequencies rather than theoretical probabilities, and relative frequencies can vary. A group of 600 people was polled to determine the opinions of the entire viewing public; if the same question was asked of a different group, a somewhat different set of relative frequencies could result. While it would be reasonable to conclude that the events are (probably) independent, it would be more appropriate to include more people in the poll and make a new comparison.

Product Rule for Independent Events

The product rule says that $p(A \cap B) = p(A | B) \cdot p(B)$. If A and B are independent, then $p(A | B) = p(A)$. Combining these two equations, we get the following rule:

PRODUCT RULE FOR INDEPENDENT EVENTS

If A and B are independent events, then the probability of A and B is

$$p(A \cap B) = p(A) \cdot p(B)$$

A common error that is made in computing probabilities is using the formula $p(A \cap B) = p(A) \cdot p(B)$ without verifying that A and B are independent. In fact, the Federal Aviation Administration (FAA) has stated that this is the most frequently encountered error in probabilistic analysis of airplane component failures. If it is not known that A and B are independent, you must use the Product Rule $p(A \cap B) = p(A | B) \cdot p(B)$.

EXAMPLE 4

SOLUTION

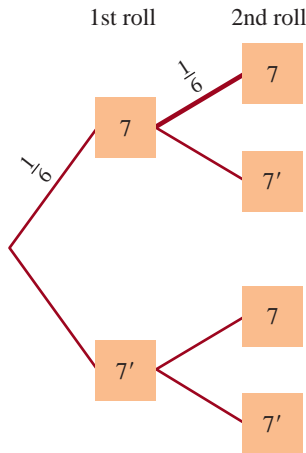


FIGURE 3.50
A tree diagram for Example 4.

USING THE PRODUCT RULE FOR INDEPENDENT EVENTS If a pair of dice is tossed twice, find the probability that each toss results in a seven.

In Example 4 of Section 3.3, we found that the probability of a seven is $\frac{1}{6}$. The two rolls are independent (one roll has no influence on the next), so the probability of a seven is $\frac{1}{6}$ regardless of what might have happened on an earlier roll; we can use the Product Rule for Independent Events:

$$\begin{aligned}
 p(A \cap B) &= p(A) \cdot p(B) \\
 p(\text{1st is 7 and 2nd is 7}) &= p(\text{1st is 7}) \cdot p(\text{2nd is 7}) \\
 &= \frac{1}{6} \cdot \frac{1}{6} \\
 &= \frac{1}{36}
 \end{aligned}$$

See Figure 3.50. The thicker branch of the tree diagram starts at the event “1st roll is 7” and ends at the event “2nd roll is 7,” so it is the conditional probability $p(\text{2nd is 7} \mid \text{1st is 7})$. However, the two rolls are independent, so $p(\text{2nd is 7} \mid \text{1st is 7}) = p(\text{2nd is 7})$. We are free to label this branch as either of these two equivalent probabilities.

Trees in Medicine and Genetics

Usually, medical diagnostic tests are not 100% accurate. A test might indicate the presence of a disease when the patient is in fact healthy (this is called a **false positive**), or it might indicate the absence of a disease when the patient does in fact have the disease (a **false negative**). Probability trees can be used to determine the probability that a person whose test results were positive actually has the disease.

EXAMPLE 5

SOLUTION

ANALYZING THE EFFECTIVENESS OF DIAGNOSTIC TESTS Medical researchers have recently devised a diagnostic test for “white lung” (an imaginary disease caused by the inhalation of chalk dust). Teachers are particularly susceptible to this disease; studies have shown that half of all teachers are afflicted with it. The test correctly diagnoses the presence of white lung in 99% of the people who have it and correctly diagnoses its absence in 98% of the people who do not have it. Find the probability that a teacher whose test results are positive actually has white lung and the probability that a teacher whose test results are negative does not have white lung.

First, we determine which probabilities have been given and find their complements, as shown in Figure 3.51. We use + to denote the event that a person receives a positive diagnosis and – to denote the event that a person receives a negative diagnosis.

Probabilities Given	Complements of Those Probabilities
$p(\text{ill}) = 0.50$	$p(\text{healthy}) = p(\text{not ill}) = 1 - 0.50 = 0.50$
$p(- \mid \text{healthy}) = 98\% = 0.98$	$p(+ \mid \text{healthy}) = 1 - 0.98 = 0.02$
$p(+ \mid \text{ill}) = 99\% = 0.99$	$p(- \mid \text{ill}) = 1 - 0.99 = 0.01$

FIGURE 3.51 Data from Example 5.

The first two of these probabilities [$p(\text{ill})$ and $p(\text{healthy})$] are not conditional, so they start at the beginning of the tree. The next two probabilities [$p(- \mid \text{healthy})$

and $p(+ | \text{healthy})$ are conditional, so they start at their condition (healthy). Similarly, the last two probabilities are conditional, so they start at their condition (ill). This placement of the probabilities yields the tree diagram in Figure 3.52.

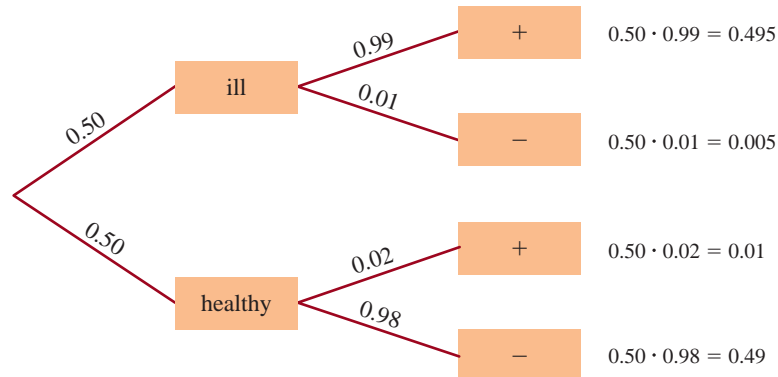


FIGURE 3.52 A tree diagram for Example 5.

The four probabilities to the right of the tree are

$$p(\text{ill and } +) = 0.495$$

$$p(\text{ill and } -) = 0.005$$

$$p(\text{healthy and } +) = 0.01$$

$$p(\text{healthy and } -) = 0.49$$

We are to find the probability that a teacher whose test results are positive actually has white lung; that is, we are to find $p(\text{ill} | +)$. Thus, we are given that the test results are positive, and we need only consider branches that involve positive test results: $p(\text{ill and } +) = 0.495$ and $p(\text{healthy and } +) = 0.01$. The probability that a teacher whose test results are positive actually has white lung is

$$p(\text{ill} | +) = \frac{0.495}{0.495 + 0.01} = 0.9801 \dots \approx 98\%$$

The probability that a teacher whose test results are negative does not have white lung is

$$p(\text{healthy} | -) = \frac{0.49}{0.49 + 0.005} = 0.98989 \dots \approx 99\%$$

The probabilities show that this diagnostic test works well in determining whether a teacher actually has white lung. In the exercises, we will see how well it would work with schoolchildren.

EXAMPLE 6

FINDING THE PROBABILITY THAT YOUR CHILD INHERITS A DISEASE Mr. and Mrs. Smith each had a sibling who died of cystic fibrosis. The Smiths are considering having a child. They have not been tested to determine whether they are carriers. What is the probability that their child would have cystic fibrosis?

SOLUTION

In our previous examples involving trees, we were given probabilities; the conditional or nonconditional status of those probabilities helped us to determine the physical layout of the tree. In this example, we are not given any probabilities. To determine the physical layout of the tree, we have to separate what we know to be true from what is only possible or probable. The tree focuses on what is possible or probable. We know that both Mr. and Mrs. Smith had a sibling who died of cystic fibrosis, and it is possible that their child would inherit that disease. The tree's

branches will represent the series of possible events that could result in the Smith child inheriting cystic fibrosis.

What events must take place if the Smith child is to inherit the disease? First, the grandparents would have to have had cystic fibrosis genes. Next, the Smiths themselves would have to have inherited those genes from their parents. Finally, the Smith child would have to inherit those genes from his or her parents.

Cystic fibrosis is recessive, which means that a person can inherit it only if he or she receives two cystic fibrosis genes, one from each parent. Mr. and Mrs. Smith each had a sibling who had cystic fibrosis, so each of the four grandparents must have been a carrier. They could not have actually had the disease because the Smiths would have known, and we would have been told.

We now know the physical layout of our tree. The four grandparents were definitely carriers. Mr. and Mrs. Smith were possibly carriers. The Smith child will possibly inherit the disease. The first set of branches will deal with Mr. and Mrs. Smith, and the second set will deal with the child.

The Smith child will not have cystic fibrosis unless Mr. and Mrs. Smith are both carriers. Figure 3.53 shows the Punnett square for Mr. and Mrs. Smith's possible genetic configuration.

	N	c	← one grandparent's genes
N	NN	cN	← possible offspring
c	Nc	cc	← possible offspring

↑
other grandparent's genes

FIGURE 3.53 A Punnett square for Example 6.

Neither Mr. Smith nor Mrs. Smith has the disease, so we can eliminate the cc possibility. Thus, the probability that Mr. Smith is a carrier is $\frac{2}{3}$, as is the probability that Mrs. Smith is a carrier. Furthermore, these two events are independent, since the Smiths are (presumably) unrelated.

Using the Product Rule for Independent Events, we have

$$\begin{aligned}
 & p(\text{Mr. S is a carrier and Mrs. S is a carrier}) \\
 &= p(\text{Mr. S is a carrier}) \cdot p(\text{Mrs. S is a carrier}) = \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9}
 \end{aligned}$$

The same Punnett square tells us that the probability that their child will have cystic fibrosis, given that the Smiths are both carriers, is $\frac{1}{4}$. Letting B be the event that both parents are carriers and letting F be the event that the child has cystic fibrosis, we obtain the tree diagram in Figure 3.54.

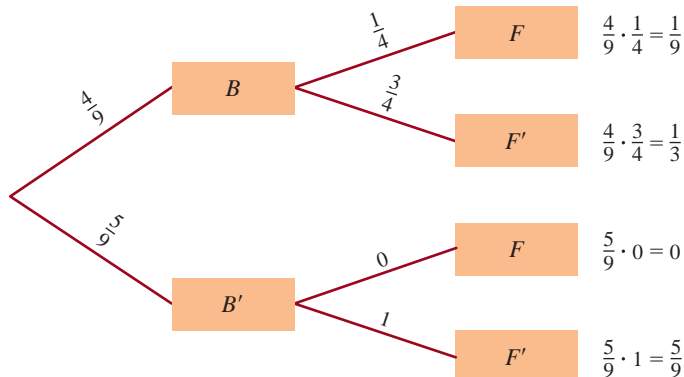


FIGURE 3.54 A tree diagram for Example 6.

The probability that the Smiths' child would have cystic fibrosis is $\frac{1}{9}$.

Notice that the tree in Figure 3.54 could have been drawn differently, as shown in Figure 3.55. It is not necessary to draw a branch going from the B' box to the F box, since the child cannot have cystic fibrosis if both parents are not carriers.

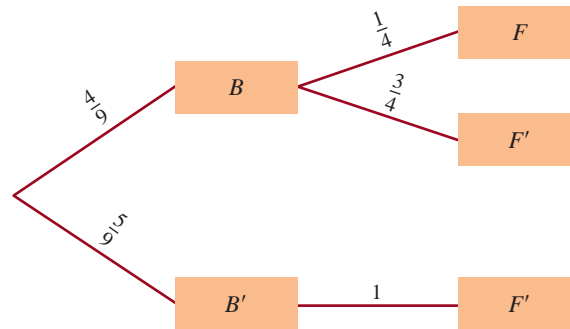


FIGURE 3.55 An alternative to the tree in Figure 3.54.

Hair Color

Like cystic fibrosis, hair color is inherited, but the method by which it is transmitted is more complicated than the method by which an inherited disease is transmitted. This more complicated method of transmission allows for the possibility that a child might have hair that is colored differently from that of other family members.

Hair color is determined by two pairs of genes: one pair that determines placement on a blond/brown/black spectrum and one pair that determines the presence or absence of red pigment. These two pairs of genes are independent.

Melanin is a brown pigment that affects the color of hair (as well as the colors of eyes and skin). The pair of genes that controls the brown hair colors does so by determining the amount of melanin in the hair. This gene has three forms, each traditionally labeled with an M (for melanin): M^{Bd} , or blond (a light melanin deposit); M^{Bw} , or brown (a medium melanin deposit); and M^{Bk} , or black (a heavy melanin deposit). Everyone has two of these genes, and their combination determines the brown aspect of hair color, as illustrated in Figure 3.56.

The hair colors in Figure 3.56 are altered by the presence of red pigment, which is determined by another pair of genes. This gene has two forms: R^- , or no red pigment, and R^+ , or red pigment. Because everyone has two of these genes, there are three possibilities for the amount of red pigment: R^-R^- , R^+R^- , and R^+R^+ . The amount of red pigment in a person's hair is independent of the brownness of his or her hair.

The actual color of a person's hair is determined by the interaction of these two pairs of genes, as shown in Figure 3.57.

Genes	Hair Color
$M^{Bd}M^{Bd}$	blond
$M^{Bd}M^{Bw}$	light brown
$M^{Bd}M^{Bk}$ $M^{Bw}M^{Bw}$	medium brown
$M^{Bw}M^{Bk}$	dark brown
$M^{Bk}M^{Bk}$	black

FIGURE 3.56 Melanin.

Genes	Blond ($M^{Bd}M^{Bd}$)	Light Brown ($M^{Bd}M^{Bw}$)	Medium Brown ($M^{Bd}M^{Bk}$, $M^{Bw}M^{Bw}$)	Dark Brown ($M^{Bw}M^{Bk}$)	Black ($M^{Bk}M^{Bk}$)
R^-R^-	blond	light brown	medium brown	dark brown	black
R^+R^-	strawberry blond	reddish brown	chestnut	shiny dark brown	shiny black
R^+R^+	bright red	dark red	auburn	glossy dark brown	glossy black

FIGURE 3.57 Melanin and red pigment.

EXAMPLE 7

SOLUTION

PREDICTING A CHILD'S HAIR COLOR The Rosses are going to have a child. Mr. Ross has blond hair, and Mrs. Ross has reddish brown hair. Find their child's possible hair colors and the probabilities of each possibility.

The parent with blond hair has genes $M^{Bd}M^{Bd}$ and R^-R^- . The parent with reddish brown hair has genes $M^{Bd}M^{Bw}$ and R^+R^- . We need to use two Punnett squares, one for the brownness of the hair (Figure 3.58) and one for the presence of red pigment (Figure 3.59).

	M^{Bd}	M^{Bd}
M^{Bd}	$M^{Bd}M^{Bd}$	$M^{Bd}M^{Bd}$
M^{Bw}	$M^{Bd}M^{Bw}$	$M^{Bd}M^{Bw}$

FIGURE 3.58 Brownness.

$$p(M^{Bd}M^{Bd}) = \frac{2}{4} = \frac{1}{2}$$

$$p(M^{Bd}M^{Bw}) = \frac{2}{4} = \frac{1}{2}$$

	R^-	R^-
R^+	R^+R^-	R^+R^-
R^-	R^-R^-	R^-R^-

FIGURE 3.59 Red pigment.

$$p(R^+R^-) = \frac{2}{4} = \frac{1}{2}$$

$$p(R^-R^-) = \frac{2}{4} = \frac{1}{2}$$

We will use a tree diagram to determine the possible hair colors and their probabilities (see Figure 3.60).

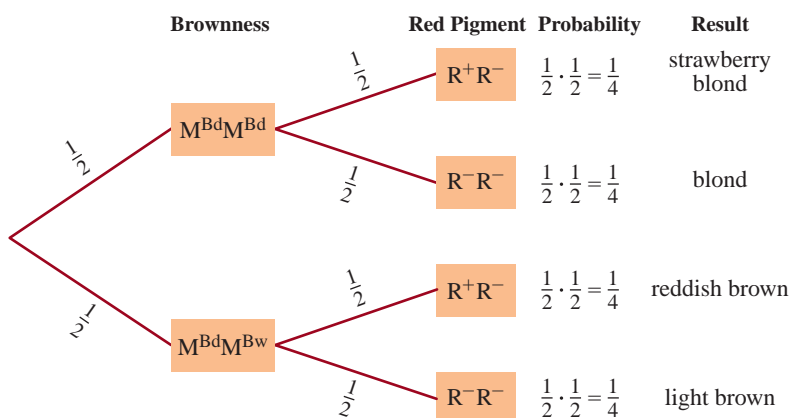


FIGURE 3.60 A tree diagram for Example 7.

The Rosses' child could have strawberry blond, blond, reddish brown, or light brown hair. The probability of each color is $\frac{1}{4}$. Notice that there is a 50% probability that the child will have hair that is colored differently from that of either parent.

3.7 EXERCISES

In Exercises 1–8, use your own personal experience with the events described to determine whether (a) E and F are independent and (b) E and F are mutually exclusive. (Where appropriate, these events are meant to be simultaneous; for example, in Exercise 2, “ E and F ” would mean “it’s simultaneously raining and sunny.”) Write a sentence justifying each of your answers.

- ▶ 1. E is the event “being a doctor,” and F is the event “being a woman.”

2. E is the event “it’s raining,” and F is the event “it’s sunny.”
3. E is the event “being single,” and F is the event “being married.”
4. E is the event “having naturally blond hair,” and F is the event “having naturally black hair.”
5. E is the event “having brown hair,” and F is the event “having gray hair.”

▶ Selected exercises available online at www.webassign.net/brookscole

6. E is the event “being a plumber,” and F is the event “being a stamp collector.”
7. E is the event “wearing shoes,” and F is the event “wearing sandals.”
- ▶ 8. E is the event “wearing shoes,” and F is the event “wearing socks.”

In Exercises 9–10, use probabilities, rather than your own personal experience, to determine whether (a) E and F are independent and (b) E and F are mutually exclusive. (c) Interpret your answers to parts (a) and (b).

- ▶ 9. If a die is rolled once, E is the event “getting a 4,” and F is the event “getting an odd number.”
- ▶ 10. If a die is rolled once, E is the event “getting a 4,” and F is the event “getting an even number.”
11. Determine whether the events “responding yes to the question on violence in television” and “being a woman” in Example 1 of Section 3.6 are independent.
12. Determine whether the events “having a defect” and “being manufactured in Peoria” in Example 4 of Section 3.6 are independent.
13. A single die is rolled once.
- Find the probability of rolling a 5.
 - Find the probability of rolling a 5 given that the number rolled is even.
 - Are the events “rolling a 5” and “rolling an even number” independent? Why?
 - Are the events “rolling a 5” and “rolling an even number” mutually exclusive? Why?
 - Interpret the results of parts (c) and (d).
- ▶ 14. A pair of dice is rolled once.
- Find the probability of rolling a 6.
 - Find the probability of rolling a 6 given that the number rolled is even.
 - Are the events “rolling a 6” and “rolling an even number” independent? Why?
 - Are the events “rolling a 6” and “rolling an even number” mutually exclusive? Why?
 - Interpret the results of parts (c) and (d).
15. A card is dealt from a full deck (no jokers).
- Find the probability of being dealt a jack.
 - Find the probability of being dealt a jack given that you were dealt a red card.
 - Are the events “being dealt a jack” and “being dealt a red card” independent? Why?
 - Are the events “being dealt a jack” and “being dealt a red card” mutually exclusive? Why?
 - Interpret the results of parts (c) and (d).
- ▶ 16. A card is dealt from a full deck (no jokers).
- Find the probability of being dealt a jack.
 - Find the probability of being dealt a jack given that you were dealt a card above a 7 (count aces high).

- Are the events “being dealt a jack” and “being dealt a card above a 7” independent? Why?
- Are the events “being dealt a jack” and “being dealt a card above a 7” mutually exclusive? Why?
- Interpret the results of parts (c) and (d).

In Exercises 17 and 18, use the following information: To determine what effect salespeople had on purchases, a department store polled 700 shoppers as to whether or not they made a purchase and whether or not they were pleased with the service. Of those who made a purchase, 125 were happy with the service and 111 were not. Of those who made no purchase, 148 were happy with the service and 316 were not.

- ▶ 17. Are the events “being happy with the service” and “making a purchase” independent? What conclusion can you make?
18. Are the events “being unhappy with the service” and “not making a purchase” independent? What conclusion can you make?
19. A personal computer manufacturer buys 38% of its chips from Japan and the rest from the United States. Of the Japanese chips, 1.7% are defective, whereas 1.1% of the U.S. chips are defective. Are the events “defective” and “Japanese-made” independent? What conclusion can you draw? (See Exercises 41–44 in Section 3.6.)
20. A skateboard manufacturer buys 23% of its ball bearings from a supplier in Akron, 38% from one in Atlanta, and the rest from one in Los Angeles. Of the ball bearings from Akron, 4% are defective; 6.5% of those from Atlanta are defective; and 8.1% of those from Los Angeles are defective.
- Find the probability that a ball bearing is defective.
 - Are the events “defective” and “from the Los Angeles supplier” independent?
 - Are the events “defective” and “from the Atlanta supplier” independent?
 - What conclusion can you draw?
21. Over the years, a group of nutritionists have observed that their vegetarian clients tend to have fewer health problems. To determine whether their observation is accurate, they collected the following data:
- They had 365 clients.
 - 281 clients are healthy.
 - Of the healthy clients, 189 are vegetarians.
 - Of the unhealthy clients, 36 are vegetarians.
- Use the data to determine whether the events “being a vegetarian” and “being healthy” are independent.
 - Use the data to determine whether the events “being a vegetarian” and “being healthy” are mutually exclusive.
 - Interpret the results of parts (a) and (b).

HINT: Start by organizing the data in a chart, similar to that used in the “Violence on Television” poll in Figure 3.33 in Section 3.6.

- 22. Over the years, a group of exercise physiologists have observed that their clients who exercise solely by running tend to have more ankle problems than do their clients who vary between running and other forms of exercise. To determine whether their observation is accurate, they collected the following data:

- They had 422 clients.
 - 276 clients are have ankle problems.
 - Of the clients with ankle problems, 191 only run.
 - Of the clients without ankle problems, 22 only run.
- a. Use the data to determine whether the events “running only” and “having ankle problems” are independent.
- b. Use the data to determine whether the events “running only” and “having ankle problems” are mutually exclusive.
- c. Interpret the results of parts (a) and (b).

HINT: Start by organizing the data in a chart, similar to that used in the “Violence on Television” poll in Figure 3.33 in Section 3.6.

23. The Venn diagram in Figure 3.61 contains the results of a survey. Event A is “supports proposition 3,” and event B is “lives in Bishop.”

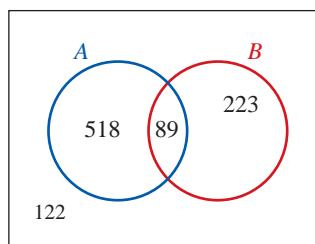


FIGURE 3.61 Venn diagram for Exercise 23.

- a. Are the events “supports proposition 3” and “lives in Bishop” independent?
- b. Are the events “supports proposition 3” and “lives in Bishop” mutually exclusive?
- c. Interpret the results of parts (a) and (b).
- 24. The Venn diagram in Figure 3.62 contains the results of a survey. Event A is “uses Ipana toothpaste,” and event B is “has good dental checkups.”

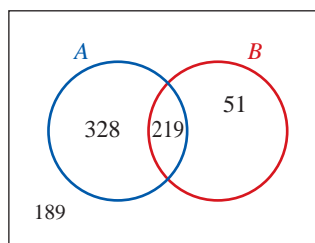


FIGURE 3.62 Venn diagram for Exercise 24.

- a. Are the events “uses Ipana toothpaste” and “has good dental checkups” independent?
- b. Are the events “uses Ipana toothpaste” and “has good dental checkups” mutually exclusive?
- c. Interpret the results of parts (a) and (b).

In Exercises 25–28, you may wish to use Exercise 66 in Section 3.6.

25. Sixty percent of all computers sold last year were HAL computers, 4% of all computer users quit using computers, and 3% of all computer users used HAL computers and then quit using computers. Are the events “using a HAL computer” and “quit using computers” independent? What conclusion can you make?
26. Ten percent of all computers sold last year were Peach computers, 4% of all computer users quit using computers, and 0.3% of all computer users used Peach computers and then quit using computers. Are the events “using a Peach computer” and “quit using computers” independent? What conclusion can you make?
27. Forty percent of the country used SmellSoGood deodorant, 10% of the country quit using deodorant, and 4% of the country used SmellSoGood deodorant and then quit using deodorant. Are the events “used SmellSoGood” and “quit using deodorant” independent? What conclusion can you make?
28. Nationwide, 20% of all TV viewers use VideoLink cable, 8% of all TV viewers switched from cable to a satellite service, and 7% of all TV viewers used VideoLink cable and then switched from cable to a satellite service. Are the events “used VideoLink cable” and “switched from cable to satellite” independent? What conclusion can you make?

- 29. Suppose that the space shuttle has three separate computer control systems: the main system and two backup duplicates of it. The first backup would monitor the main system and kick in if the main system failed. Similarly, the second backup would monitor the first. We can assume that a failure of one system is independent of a failure of another system, since the systems are separate. The probability of failure for any one system on any one mission is known to be 0.01.
- a. Find the probability that the shuttle is left with no computer control system on a mission.
- b. How many backup systems does the space shuttle need if the probability that the shuttle is left with no computer control system on a mission must be $\frac{1}{1 \text{ billion}}$?
30. Use the information in Example 5 to find $p(\text{healthy} \mid +)$, the probability that a teacher receives a false positive.
31. Use the information in Example 5 to find $p(\text{ill} \mid -)$, the probability that a teacher receives a false negative.

- 32. Overwhelmed with their success in diagnosing white lung in teachers, public health officials decided to administer the test to all schoolchildren, even though only one child in 1,000 has contracted the disease. Recall from Example 5 that the test correctly diagnoses the presence of white lung in 99% of the people who have it and correctly diagnoses its absence in 98% of the people who do not have it.
- Find the probability that a schoolchild whose test results are positive actually has white lung.
 - Find the probability that a schoolchild whose test results are negative does not have white lung.
 - Find the probability that a schoolchild whose test results are positive does not have white lung.
 - Find the probability that a schoolchild whose test results are negative actually has white lung.
 - Which of these events is a false positive? Which is a false negative?
 - Which of these probabilities would you be interested in if you or one of your family members tested positive?
 - Discuss the usefulness of this diagnostic test, both for teachers (as in Example 5) and for schoolchildren.
- 33. In 2004, the Centers for Disease Control and Prevention estimated that 1,000,000 of the 287,000,000 residents of the United States are HIV-positive. (HIV is the virus that is believed to cause AIDS.) The SUDS diagnostic test correctly diagnoses the presence of AIDS/HIV 99.9% of the time and correctly diagnoses its absence 99.6% of the time.
- Find the probability that a person whose test results are positive actually has HIV.
 - Find the probability that a person whose test results are negative does not have HIV.
 - Find the probability that a person whose test results are positive does not have HIV.
 - Find the probability that a person whose test results are negative actually has HIV.
 - Which of the probabilities would you be interested in if you or someone close to you tested positive?
 - Which of these events is a false positive? Which is a false negative?
 - Discuss the usefulness of this diagnostic test.
 - It has been proposed that all immigrants to the United States should be tested for HIV before being allowed into the country. Discuss this proposal.
- 34. Assuming that the SUDS test cannot be made more accurate, what changes in the circumstances described in Exercise 33 would increase the usefulness of the SUDS diagnostic test? Give a specific example of this change in circumstances, and demonstrate how it would increase the test's usefulness by computing appropriate probabilities.
35. Compare and contrast the circumstances and the probabilities in Example 5 and in Exercises 32, 33, and 34. Discuss the difficulties in using a diagnostic test when that test is not 100% accurate.
36. Use the information in Example 6 to find the following:
- the probability that the Smiths' child is a cystic fibrosis carrier
 - the probability that the Smiths' child is healthy (i.e., has no symptoms)
 - the probability that the Smiths' child is healthy and not a carrier
- HINT:* You must consider three possibilities: Both of the Smiths are carriers, only one of the Smiths is a carrier, and neither of the Smiths is a carrier.
37.
 - Two first cousins marry. Their mutual grandfather's sister died of cystic fibrosis. They have not been tested to determine whether they are carriers. Find the probability that their child will have cystic fibrosis. (Assume that no other grandparents are carriers.)
 - Two unrelated people marry. Each had a grandparent whose sister died of cystic fibrosis. They have not been tested to determine whether they are carriers. Find the probability that their child will have cystic fibrosis. (Assume that no other grandparents are carriers.)
- 38. It is estimated that one in twenty-five Americans is a cystic fibrosis carrier. Find the probability that a randomly selected American couple's child will have cystic fibrosis (assuming that they are unrelated).
39. In 1989, researchers announced a new carrier-detection test for cystic fibrosis. However, it was discovered in 1990 that the test will detect the presence of the cystic fibrosis gene in only 85% of cystic fibrosis carriers. If two unrelated people are in fact carriers of cystic fibrosis, find the probability that they both test positive. Find the probability that they don't both test positive.
- 40. Ramon del Rosario's mother's father died of Huntington's disease. His mother died in childbirth, before symptoms of the disease would have appeared in her. Find the probability that Ramon will have the disease. (Assume that Ramon's grandfather had one bad gene, and that there was no other source of Huntington's in the family.) (Huntington's disease is discussed on page 149 in Section 3.2.)
41. Albinism is a recessive disorder that blocks the normal production of pigmentation. The typical albino has white hair, white skin, and pink eyes and is visually impaired. Mr. Jones is an albino, and although Ms. Jones is normally pigmented, her brother is an albino. Neither of Ms. Jones' parents is an albino. Find the probability that their child will be an albino.

- ▶ 42. Find the probability that the Joneses' child will not be an albino but will be a carrier. (See Exercise 41.)
43. If the Joneses' first child is an albino, find the probability that their second child will be an albino. (See Exercise 41.)
44. The Donohues are going to have a child. She has shiny black hair, and he has bright red hair. Find their child's possible hair colors and the probabilities of each possibility.
45. The Yorks are going to have a child. She has black hair, and he has dark red hair. Find their child's possible hair colors and the probabilities of each possibility.
- ▶ 46. The Eastwoods are going to have a child. She has chestnut hair ($M^{Bd}M^{Bk}$), and he has dark brown hair. Find their child's possible hair colors and the probabilities of each possibility.
47. The Wilsons are going to have a child. She has strawberry blond hair, and he has shiny dark brown hair. Find their child's possible hair colors and the probabilities of each possibility.
48. The Breuners are going to have a child. She has blond hair, and he has glossy dark brown hair. Find their child's possible hair colors and the probabilities of each possibility.
49. The Landres are going to have a child. She has chestnut hair ($M^{Bd}M^{Bk}$), and he has shiny dark brown hair. Find their child's possible hair colors and the probabilities of each possibility.
- ▶ 50. The Hills are going to have a child. She has reddish brown hair, and he has strawberry blond hair. Find their child's possible hair colors and the probabilities of each possibility.
51. Recall from Section 3.1 that Antoine Gombauld, the Chevalier de Méré, had made money over the years by betting with even odds that he could roll at least one 6 in four rolls of a single die. This problem finds the probability of winning that bet and the expected value of the bet.
- Find the probability of rolling a 6 in one roll of one die.
 - Find the probability of not rolling a 6 in one roll of one die.
 - Find the probability of never rolling a 6 in four rolls of a die.
- HINT:* This would mean that the first roll is not a 6 *and* the second roll is not a 6 *and* the third is not a 6 *and* the fourth is not a 6.
- Find the probability of rolling at least one 6 in four rolls of a die.
- HINT:* Use complements.
- Find the expected value of this bet if \$1 is wagered.
52. Recall from Section 3.1 that Antoine Gombauld, the Chevalier de Méré, had lost money over the years by betting with even odds that he could roll at least one pair of 6's in twenty-four rolls of a pair of dice and that he could not understand why. This problem finds the probability of winning that bet and the expected value of the bet.
- Find the probability of rolling a double 6 in one roll of a pair of dice.
 - Find the probability of not rolling a double 6 in one roll of a pair of dice.
 - Find the probability of never rolling a double 6 in twenty-four rolls of a pair of dice.
 - Find the probability of rolling at least one double 6 in twenty-four rolls of a pair of dice.
 - Find the expected value of this bet if \$1 is wagered.
53. Probability theory began when Antoine Gombauld, the Chevalier de Méré, asked his friend Blaise Pascal why he had made money over the years betting that he could roll at least one 6 in four rolls of a single die but he had lost money betting that he could roll at least one pair of 6's in twenty-four rolls of a pair of dice. Use decision theory and the results of Exercises 51 and 52 to answer Gombauld.
54. Use Exercise 66 of Section 3.6 to explain the calculations of $p(\text{ill} \mid +)$ and $p(\text{healthy} \mid -)$ in Example 5 of this section.
- ▶ 55. Dr. Wellby's patient exhibits symptoms associated with acute neural toxemia (an imaginary disease), but the symptoms can have other, innocuous causes. Studies show that only 25% of those who exhibit the symptoms actually have acute neural toxemia (ANT). A diagnostic test correctly diagnoses the presence of ANT in 88% of the persons who have it and correctly diagnoses its absence in 92% of the persons who do not have it. ANT can be successfully treated, but the treatment causes side effects in 2% of the patients. If left untreated, 90% of those with ANT die; the rest recover fully. Dr. Wellby is considering ordering a diagnostic test for her patient and treating the patient if test results are positive, but she is concerned about the treatment's side effects.
- Dr. Wellby could choose to test her patient and administer the treatment if the results are positive. Find the probability that her patient's good health will return under this plan.
 - Dr. Wellby could choose to avoid the treatment's side effects by not administering the treatment. (This also implies not testing the patient.) Find the probability that her patient's good health will return under this plan.
 - Find the probability that the patient's good health will return if he undergoes treatment regardless of the test's outcome.
 - On the basis of the probabilities, should Dr. Wellby order the test and treat the patient if test results are positive?



Answer the following questions using complete sentences and your own words.

• CONCEPT QUESTIONS

56. In Example 5, we are given that $p(- | \text{healthy}) = 98\%$ and $p(+ | \text{ill}) = 99\%$, and we computed that $p(\text{ill} | +) \approx 98\%$ and that $p(\text{healthy} | -) \approx 99\%$. Which of these probabilities would be most important

to a teacher who was diagnosed as having white lung? Why?

57. Are the melanin hair color genes (M^{Bd} , M^{Bw} , and M^{Bk}) dominant, recessive, or codominant? Are the redness hair color genes (R^+ and R^-) dominant, or recessive, or codominant? Why?
58. Compare and contrast the concepts of independence and mutual exclusivity. Be sure to discuss both the similarities and the differences between these two concepts.



TERMS

carrier
certain event
codominant gene
conditional probability
decision theory

dependent events
dominant gene
event
expected value
experiment
false negative
false positive

genes
impossible event
independent events
Law of Large Numbers
mutually exclusive events
odds
outcome

probability
Punnett square
recessive gene
relative frequency
sample space

PROBABILITY RULES

- The probability of the null set is 0: $p(\emptyset) = 0$
- The probability of the sample space is 1: $p(S) = 1$
- Probabilities are between 0 and 1 (inclusive):
 $0 \leq p(S) \leq 1$
- $p(E \cup F) = p(E) + p(F) - p(E \cap F)$
- $p(E \cup F) = p(E) + p(F)$ if E and F are mutually exclusive
- $p(E) + p(E') = 1$

FORMULAS

If outcomes are equally likely, then:

- the **probability** of an event E is $p(E) = n(E)/n(S)$
- the **odds** of an event E are $o(E) = n(E):n(E')$

To find the **expected value** of an experiment, multiply the value of each outcome by its probability and add the results.

The **conditional probability** of A given B is

$$p(A | B) = n(A \cap B)/n(B) \text{ if outcomes are equally likely}$$

The **product rule**:

$$p(A \cap B) = p(A | B) \cdot p(B) \text{ for any two events } A \text{ and } B$$

$$p(A \cap B) = p(A) \cdot p(B) \text{ if } A \text{ and } B \text{ are independent}$$

REVIEW EXERCISES

In Exercises 1–6, a card is dealt from a well-shuffled deck of fifty-two cards.

- Describe the experiment and the sample space.
- Find and interpret the probability and the odds of being dealt a queen.
- Find and interpret the probability and the odds of being dealt a club.
- Find and interpret the probability and the odds of being dealt the queen of clubs.
- Find and interpret the probability and the odds of being dealt a queen or a club.
- Find and interpret the probability and the odds of being dealt something other than a queen.

In Exercises 7–12, three coins are tossed.

- Find the experiment and the sample space.
- Find the event E that exactly two are tails.
- Find the event F that two or more are tails.
- Find and interpret the probability of E and the odds of E .
- Find and interpret the probability of F and the odds of F .
- Find and interpret the probability of E' and the odds of E' .

In Exercises 13–18, a pair of dice is tossed. Find and interpret the probability of rolling each of the following.

- a 7
- an 11
- a 7, an 11, or doubles
- a number that is both odd and greater than 8
- a number that is either odd or greater than 8
- a number that is neither odd nor greater than 8

In Exercises 19–24, three cards are dealt from a deck of fifty-two cards. Find the probability of each of the following.

- All three are hearts.
- Exactly two are hearts.
- At least two are hearts.
- The first is an ace of hearts, the second is a 2 of hearts, and the third is a 3 of hearts.
- The second is a heart, given that the first is a heart.
- The second is a heart, and the first is a heart.

In Exercises 25–30, a pair of dice is rolled three times. Find the probability of each of the following.

- All three are 7's.
- Exactly two are 7's.
- At least two are 7's.

- None are 7's.
- The second roll is a 7, given that the first roll is a 7.
- The second roll is a 7 and the first roll is a 7.

In Exercises 31–32, use the following information. A long-stemmed pea is dominant over a short-stemmed one. A pea with one long-stemmed gene and one short-stemmed gene is crossed with a pea with two short-stemmed genes.

- Find and interpret the probability that the offspring will be long-stemmed.
- Find and interpret the probability that the offspring will be short-stemmed.

In Exercises 33–35, use the following information: Cystic fibrosis is caused by a recessive gene. Two cystic fibrosis carriers produce a child.

- Find the probability that that child will have the disease.
- Find the probability that that child will be a carrier.
- Find the probability that that child will neither have the disease nor be a carrier.

In Exercises 36–38, use the following information: Sickle-cell anemia is caused by a codominant gene. A couple, each of whom has sickle-cell trait, produce a child. (Sickle-cell trait involves having one bad gene. Sickle-cell anemia involves having two bad genes.)

- Find the probability that that child will have the disease.
- Find the probability that that child will have sickle-cell trait.
- Find the probability that that child will have neither sickle-cell disease nor sickle-cell trait.

In Exercises 39–41, use the following information: Huntington's disease is caused by a dominant gene. One parent has Huntington's disease. This parent has a single gene for Huntington's disease.

- Find the probability that that child will have the disease.
- Find the probability that that child will be a carrier.
- Find the probability that that child will neither have the disease nor be a carrier.
- Find the probability of being dealt a pair of 10's (and no other 10's) when playing five-card draw poker.
- Find the probability of being dealt a pair of tens and three jacks when playing five-card draw poker.
- Find the probability of being dealt a pair of tens and a pair of jacks (and no other tens or jacks) when playing five-card draw poker.
- In nine-spot keno, five winning spots breaks even, six winning spots pays \$50, seven pays \$390, eight pays \$6000, and nine pays \$25,000.
 - Find the probability of each of these events.
 - Find the expected value of a \$1 bet.

46. Some \$1 bets in craps (specifically, the pass, don't pass, come, and don't come bets) have expected values of $-\$0.014$. Use decision theory to compare these bets with a \$1 bet in nine-spot keno. (See Exercise 45.)
47. At a certain office, three people make \$6.50 per hour, three make \$7, four make \$8.50, four make \$10, four make \$13.50, and two make \$25. Find the expected hourly wage at that office.

In Exercises 48–55, use the following information: Jock O'Neill, a sportscaster, and Trudy Bell, a member of the state assembly, are both running for governor of the state of Erehwon. A recent telephone poll asked 800 randomly selected voters whom they planned on voting for. The results of this poll are shown in Figure 3.63.

	Jock O'Neill	Trudy Bell	Undecided
Urban residents	266	184	22
Rural residents	131	181	16

FIGURE 3.63 Data for Exercises 48–55.

48. Find the probability that an urban resident supports O'Neill and the probability that an urban resident supports Bell.
49. Find the probability that a rural resident supports O'Neill and the probability that a rural resident supports Bell.
50. Find the probability that an O'Neill supporter lives in an urban area and the probability that an O'Neill supporter lives in a rural area.
51. Find the probability that a Bell supporter lives in an urban area and the probability that a Bell supporter lives in a rural area.
52. Where are O'Neill supporters more likely to live? Where are Bell supporters more likely to live?
53. Which candidate do the urban residents tend to prefer? Which candidate do the rural residents tend to prefer?
54. Are the events "supporting O'Neill" and "living in an urban area" independent or dependent? What can you conclude?
55. On the basis of the poll, who is ahead in the gubernatorial race?

Are the following events independent or dependent? Are they mutually exclusive?

56. "It is summer" and "it is sunny." (Use your own personal experience.)
57. "It is summer" and "it is Monday." (Use your own personal experience.)
58. "It is summer" and "it is autumn." (Use your own personal experience.)

59. "The first card dealt is an ace" and "the second card dealt is an ace." (Do not use personal experience.)
60. "The first roll of the dice results in a 7" and "the second roll results in a 7" (Do not use personal experience.)

In Exercises 61–66, use the following information: Gregor's Garden Corner buys 40% of their plants from the Green Growery and the balance from Herb's Herbs. Twenty percent of the plants from the Green Growery must be returned, and 10% of those from Herb's Herbs must be returned.

61. What percent of all plants are returned?
62. What percent of all plants are not returned?
63. What percent of all plants are from the Green Growery and are returned?
64. What percent of the returned plants are from the Green Growery?
65. What percent of the returned plants are from Herb's Herbs?
66. Are the events "a plant must be returned" and "a plant was from Herb's Herbs" independent? What conclusion can you make?

In Exercises 67–73, use the following information: The Nissota Automobile Company buys emergency flashers from two different manufacturers, one in Arkansas and one in Nevada. Thirty-nine percent of its turn-signal indicators are purchased from the Arkansas manufacturer, and the rest are purchased from the Nevada manufacturer. Two percent of the Arkansas turn-signal indicators are defective, and 1.7% of the Nevada indicators are defective.

67. What percent of the indicators are defective and made in Arkansas?
68. What percentage of the indicators are defective and made in Nevada?
69. What percentage of the indicators are defective?
70. What percentage of the indicators are not defective?
71. What percentage of the defective indicators are made in Arkansas?
72. What percentage of the defective indicators are made in Nevada?
73. Are the events "made in Arkansas" and "a defective" independent? What conclusion can you make?



Answer the following questions with complete sentences and your own words.

• CONCEPT QUESTIONS

74. What is a conditional probability?
75. Give two events that are independent, and explain why they are independent.
76. Give two events that are dependent, and explain why they are dependent.
77. Give two events that are mutually exclusive, and explain why they are mutually exclusive.

78. Give two events that are not mutually exclusive, and explain why they are not mutually exclusive.
 79. Why are probabilities always between 0 and 1, inclusive?
 80. Give an example of a permutation and a similar example of a combination.
 81. Give an example of two events that are mutually exclusive, and explain why they are mutually exclusive.
 82. Give an example of two events that are not mutually exclusive, and explain why they are not mutually exclusive.
 83. How is set theory used in probability theory?
- **HISTORY**
84. What two mathematicians invented probability theory? Why?
 85. Why was probability theory not considered to be a serious branch of mathematics? What changed that viewpoint? Give a specific example of something that helped change that viewpoint.
 86. What role did the following people play in the development of probability theory?
Jacob Bernoulli
Gerolamo Cardano
Pierre de Fermat
Antoine Gombauld, the Chevalier de Méré
John Graunt
Pierre-Simon, the Marquis de Laplace
Gregor Mendel
Blaise Pascal

STATISTICS

4



© Image copyright Sai Yeung Chan, 2009. Used under license from Shutterstock.com

Statistics are everywhere. The news, whether reported in a newspaper, on television, through the Internet, or over the radio, includes statistics of every kind. When shopping for a new car, you will certainly examine the statistics (average miles per gallon, acceleration times, braking distances, and so on) of the various makes and models you are considering. Statistics abound in government studies, and the interpretation of these statistics affects us all. Industry is driven by statistics; they are essential to the direction of quality control, marketing research, productivity, and many other factors. Sporting events are laden with statistics concerning the past performance of the teams and players.

A person who understands the nature of statistics is equipped to see beyond short-term and individual perspectives.

continued

WHAT WE WILL DO IN THIS CHAPTER

WE'LL EXPLORE DIFFERENT TYPES OF STATISTICS, BOTH DESCRIPTIVE AND INFERENCEAL:

- Once data have been collected from a sample, how should they be organized and presented?
- Once they have been organized, how should data be summarized?
- Once data have been summarized, how are conclusions drawn and predictions made?

WE'LL ANALYZE AND EXPLORE DISTRIBUTIONS OF DATA THAT EXHIBIT SPECIFIC TRENDS OR PATTERNS:

- It is not uncommon for data to be clustered around a central value. In many instances, this type of pattern can be represented as a "bell-shaped" curve (high in the middle, low at each end).
- All bell curves share common features. What are these features and how can they be applied in specific situations?

continued

WHAT WE WILL DO IN THIS CHAPTER — *continued*

WE'LL ANALYZE AND INTERPRET SURVEYS AND OPINION POLLS:

- Once the opinions of a specific sample of people have been collected, how can they be used to predict the overall opinion of a large population?
- How accurate or reliable are opinion polls?

WE'LL EXAMINE RELATIONSHIPS BETWEEN TWO SETS OF DATA:

- Are the values of one variable related to the values of another variable?
- If two variables are related, how can the relationship be expressed so that predictions can be made?

He or she is also better prepared to deal with those who use statistics in misleading ways. To many people, the word *statistics* conjures up an image of an endless list of facts and figures. Where do statistics come from? What do they mean? In this chapter, you will learn to handle basic statistical problems and expand your knowledge of the meanings, uses, and misuses of statistics.

4.1 Population, Sample, and Data

OBJECTIVES

- Construct a frequency distribution
- Construct a histogram
- Construct a pie chart

The field of **statistics** can be defined as the science of collecting, organizing, and summarizing data in such a way that valid conclusions and meaningful predictions can be drawn from them. The first part of this definition, “collecting, organizing, and summarizing data,” applies to **descriptive statistics**. The second part, “drawing valid conclusions and making meaningful predictions,” describes **inferential statistics**.

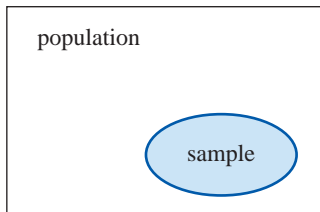


FIGURE 4.1

Population versus sample.



What person who lived in the twentieth century do you admire most? The top ten responses in a Gallup poll taken on Dec. 20–21, 1999, were as follows: (1) Mother Teresa, (2) Martin Luther King, Jr., (3) John F. Kennedy, (4) Albert Einstein, (5) Helen Keller, (6) Franklin D. Roosevelt, (7) Billy Graham, (8) Pope John Paul II, (9) Eleanor Roosevelt, and (10) Winston Churchill.

Population versus Sample

Who will become the next president of the United States? During election years, political analysts spend a lot of time and money trying to determine what percent of the vote each candidate will receive. However, because there are over 175 million registered voters in the United States, it would be virtually impossible to contact each and every one of them and ask, “Whom do you plan on voting for?” Consequently, analysts select a smaller group of people, determine their intended voting patterns, and project their results onto the entire body of all voters.

Because of time and money constraints, it is very common for researchers to study the characteristics of a small group in order to estimate the characteristics of a larger group. In this context, the set of all objects under study is called the **population**, and any subset of the population is called a **sample** (see Figure 4.1).

When we are studying a large population, we might not be able to collect data from every member of the population, so we collect data from a smaller, more manageable sample. Once we have collected these data, we can summarize by calculating various descriptive statistics, such as the average value. Inferential statistics, then, deals with drawing conclusions (hopefully, valid ones!) about the population, based on the descriptive statistics of the sample data.

Sample data are collected and summarized to help us draw conclusions about the population. A good sample is representative of the population from which it was taken. Obviously, if the sample is not representative, the conclusions concerning the population might not be valid. The most difficult aspect of inferential statistics is obtaining a representative sample. Remember that conclusions are only as reliable as the sampling process and that information will usually change from sample to sample.

Frequency Distributions

The first phase of any statistical study is the collection of data. Each element in a set of data is referred to as a **data point**. When data are first collected, the data points might show no apparent patterns or trends. To summarize the data and detect any trends, we must organize the data. This is the second phase of descriptive statistics. The most common way to organize raw data is to create a **frequency distribution**, a table that lists each data point along with the number of times it occurs (its **frequency**).

The composition of a frequency distribution is often easier to see if the frequencies are converted to percents, especially if large amounts of data are being summarized. The **relative frequency** of a data point is the frequency of the data point expressed as a percent of the total number of data points (that is, made *relative* to the total). The relative frequency of a data point is found by dividing its frequency by the total number of data points in the data set. Besides listing the frequency of each data point, a frequency distribution should also contain a column that gives the relative frequencies.

EXAMPLE 1

CREATING A FREQUENCY DISTRIBUTION: SINGLE VALUES While bargaining for their new contract, the employees of 2 Dye 4 Clothing asked their employers to provide daycare service as an employee benefit. Examining the personnel files of the company’s fifty employees, the management recorded

the number of children under six years of age that each employee was caring for. The following results were obtained:

0 2 1 0 3 2 0 1 1 0
 0 1 1 2 4 1 0 1 1 0
 2 1 0 0 3 0 0 1 2 1
 0 0 2 4 1 1 0 1 2 0
 1 1 0 3 5 1 2 1 3 2

Organize the data by creating a frequency distribution.

SOLUTION

First, we list each different number in a column, putting them in order from smallest to largest (or vice versa). Then we use tally marks to count the number of times each data point occurs. The frequency of each data point is shown in the third column of Figure 4.2.

Number of Children under Six	Tally	Frequency	Relative Frequency
0	### ### ###	16	$\frac{16}{50} = 0.32 = 32\%$
1	### ### ###	18	$\frac{18}{50} = 0.36 = 36\%$
2	###	9	$\frac{9}{50} = 0.18 = 18\%$
3		4	$\frac{4}{50} = 0.08 = 8\%$
4		2	$\frac{2}{50} = 0.04 = 4\%$
5		1	$\frac{1}{50} = 0.02 = 2\%$
		$n = 50$	total = 100%

FIGURE 4.2 Frequency distribution of data.

To get the relative frequencies, we divide each frequency by 50 (the total number of data points) and change the resulting decimal to a percent, as shown in the fourth column of Figure 4.2.



Adding the frequencies, we see that there is a total of $n = 50$ data points in the distribution. This is a good way to monitor the tally process.

The raw data have now been organized and summarized. At this point, we can see that about one-third of the employees have no need for child care (32%), while the remaining two-thirds (68%) have at least one child under 6 years of age who would benefit from company-sponsored daycare. The most common trend (that is, the data point with the highest relative frequency for the fifty employees) is having one child (36%).

Grouped Data

When raw data consist of only a few distinct values (for instance, the data in Example 1, which consisted of only the numbers 0, 1, 2, 3, 4, and 5), we can easily organize the data and determine any trends by listing each data point along with its frequency and relative frequency. However, when the raw data consist of many

nonrepeated data points, listing each one separately does not help us to see any trends the data set might contain. In such cases, it is useful to group the data into intervals or classes and then determine the frequency and relative frequency of each group rather than of each data point.

EXAMPLE 2

CREATING A FREQUENCY DISTRIBUTION: GROUPED DATA Keith Reed is an instructor for an acting class offered through a local arts academy. The class is open to anyone who is at least 16 years old. Forty-two people are enrolled; their ages are as follows:

26 16 21 34 45 18 41 38 22
 48 27 22 30 39 62 25 25 38
 29 31 28 20 56 60 24 61 28
 32 33 18 23 27 46 30 34 62
 49 59 19 20 23 24

Organize the data by creating a frequency distribution.

SOLUTION

This example is quite different from Example 1. Example 1 had only six different data values, whereas this example has many. Listing each distinct data point and its frequency might not summarize the data well enough for us to draw conclusions. Instead, we will work with grouped data.

First, we find the largest and smallest values (62 and 16). Subtracting, we find the range of ages to be $62 - 16 = 46$ years. In working with grouped data, it is customary to create between four and eight groups of data points. We arbitrarily choose six groups, the first group beginning at the smallest data point, 16. To find the beginning of the second group (and hence the end of the first group), divide the range by the number of groups, round off this answer to be consistent with the data, and then add the result to the smallest data point:

$$46 \div 6 = 7.6666666 \dots \approx 8 \quad \text{This is the width of each group.}$$

The beginning of the second group is $16 + 8 = 24$, so the first group consists of people from 16 up to (but not including) 24 years of age.

In a similar manner, the second group consists of people from 24 up to (but not including) 32 ($24 + 8 = 32$) years of age. The remaining groups are formed and the ages tallied in the same way. The frequency distribution is shown in Figure 4.3.

$x = \text{Age (years)}$	Tally	Frequency	Relative Frequency
$16 \leq x < 24$	### ###	11	$\frac{11}{42} \approx 26\%$
$24 \leq x < 32$	### ###	13	$\frac{13}{42} \approx 31\%$
$32 \leq x < 40$	###	7	$\frac{7}{42} \approx 17\%$
$40 \leq x < 48$		3	$\frac{3}{42} \approx 7\%$
$48 \leq x < 56$		2	$\frac{2}{42} \approx 5\%$
$56 \leq x < 64$	###	6	$\frac{6}{42} \approx 14\%$
		$n = 42$	total = 100%

FIGURE 4.3 Frequency distribution of grouped data.

Now that the data have been organized, we can observe various trends: Ages from 24 to 32 are most common (31% is the highest relative frequency), and ages from 48 to 56 are least common (5% is the lowest). Also, over half the people enrolled (57%) are from 16 to 32 years old.

When we are working with grouped data, we can choose the groups in any desired fashion. The method used in Example 2 might not be appropriate in all situations. For example, we used the smallest data point as the beginning of the first group, but we could have begun the first group at an even smaller number. The following box gives a general method for constructing a frequency distribution.

CONSTRUCTING A FREQUENCY DISTRIBUTION

1. If the raw data consist of many different values, create intervals and work with grouped data. If not, list each distinct data point. [When working with grouped data, choose from four to eight intervals. Divide the range (high minus low) by the desired number of intervals, round off this answer to be consistent with the data, and then add the result to the lowest data point to find the beginning of the second group.]
2. Tally the number of data points in each interval or the number of times each individual data point occurs.
3. List the frequency of each interval or each individual data point.
4. Find the relative frequency by dividing the frequency of each interval or each individual data point by the total number of data points in the distribution. The resulting decimal can be expressed as a percent.

Histograms

When data are grouped in intervals, they can be depicted by a **histogram**, a bar chart that shows how the data are distributed in each interval. To construct a histogram, mark off the class limits on a horizontal axis. If each interval has equal width, we draw two vertical axes; the axis on the left exhibits the frequency of an interval, and the axis on the right gives the corresponding relative frequency. We then draw a rectangle above each interval; the height of the rectangle corresponds to the number of data points contained in the interval. The vertical scale on the right gives the percentage of data contained in each interval. The histogram depicting the distribution of the ages of the people in Keith Reed's acting class (Example 2) is shown in Figure 4.4.

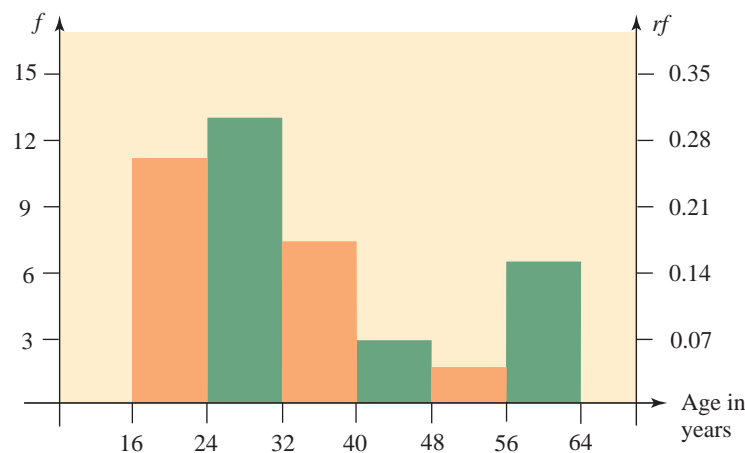


FIGURE 4.4 Ages of the people in Keith Reed's acting class.

What happens if the intervals do not have equal width? For instance, suppose the ages of the people in Keith Reed's acting class are those given in the frequency distribution shown in Figure 4.5.

$x = \text{Age (years)}$	Frequency	Relative Frequency	Class Width
$15 \leq x < 20$	4	$\frac{4}{42} \approx 10\%$	5
$20 \leq x < 25$	11	$\frac{11}{42} \approx 26\%$	5
$25 \leq x < 30$	6	$\frac{6}{42} \approx 14\%$	5
$30 \leq x < 45$	11	$\frac{11}{42} \approx 26\%$	15
$45 \leq x < 65$	10	$\frac{10}{42} \approx 24\%$	20
	$n = 42$	Total = 100%	

FIGURE 4.5 Frequency distribution of age.

With frequency and relative frequency as the vertical scales, the histogram depicting this new distribution is given in Figure 4.6.

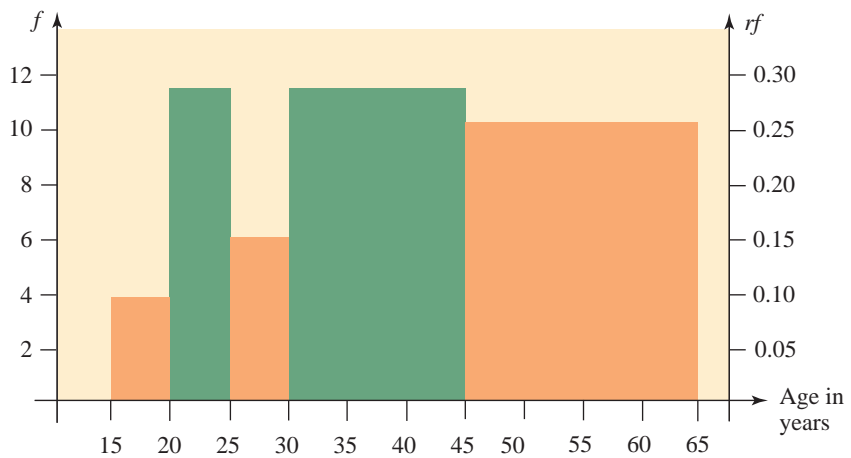


FIGURE 4.6 Why is this histogram misleading?

Does the histogram in Figure 4.6 give a truthful representation of the distribution? No; the rectangle over the interval from 45 to 65 appears to be larger than the rectangle over the interval from 30 to 45, yet the interval from 45 to 65 contains fewer data than the interval from 30 to 45. This is misleading; rather than comparing the heights of the rectangles, our eyes naturally compare the areas of the rectangles. Therefore, to make an accurate comparison, *the areas of the rectangles must correspond to the relative frequencies of the intervals*. This is accomplished by utilizing the **density** of each interval.

Histograms and Relative Frequency Density

Density is a ratio. In science, density is used to determine the concentration of weight in a given volume: $\text{Density} = \text{weight}/\text{volume}$. For example, the density of water is 62.4 pounds per cubic foot. In statistics, density is used to determine the concentration of data in a given interval: $\text{Density} = (\text{percent of total data})/(\text{size of an interval})$. Because relative frequency is a measure of the percentage of data within an interval, we shall calculate the relative frequency density of an interval to determine the concentration of data within the interval.

For example, if the interval $20 \leq x < 25$ contains eleven out of forty-two data points, then the relative frequency density of the interval is

$$rfd = \frac{f/n}{\Delta x} = \frac{11/42}{5} = 0.052380952 \dots$$

DEFINITION OF RELATIVE FREQUENCY DENSITY

Given a set of n data points, if an interval contains f data points, then the **relative frequency density** (rfd) of the interval is

$$rfd = \frac{f/n}{\Delta x}$$

where Δx is the width of the interval.

If a histogram is constructed using relative frequency density as the vertical scale, the area of a rectangle will correspond to the relative frequency of the interval, as shown in Figure 4.7.

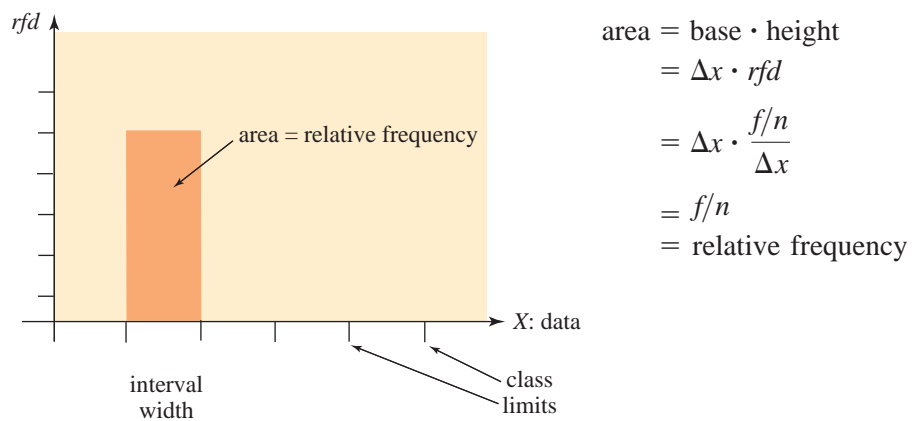


FIGURE 4.7 Area equals relative frequency.

Adding a new column to the frequency distribution given in Figure 4.5, we obtain the relative frequency densities shown in Figure 4.8.

$x = \text{Age (years)}$	Relative Frequency	Class Frequency	Relative Width	Frequency Density
$15 \leq x < 20$	4	$\frac{4}{42} \approx 10\%$	5	$\frac{4}{42} \div 5 \approx 0.020$
$20 \leq x < 25$	11	$\frac{11}{42} \approx 26\%$	5	$\frac{11}{42} \div 5 \approx 0.052$
$25 \leq x < 30$	6	$\frac{6}{42} \approx 14\%$	5	$\frac{6}{42} \div 5 \approx 0.028$
$30 \leq x < 45$	11	$\frac{11}{42} \approx 26\%$	15	$\frac{11}{42} \div 15 \approx 0.017$
$45 \leq x < 65$	10	$\frac{10}{42} \approx 24\%$	20	$\frac{10}{42} \div 20 \approx 0.012$
	$n = 42$	Total = 100%		

FIGURE 4.8 Calculating relative frequency density.

We now construct a histogram using relative frequency density as the vertical scale. The histogram depicting the distribution of the ages of the people in Keith Reed’s acting class (using the frequency distribution in Figure 4.8) is shown in Figure 4.9.

Comparing the histograms in Figures 4.6 and 4.9, we see that using relative frequency density as the vertical scale (rather than frequency) gives a more truthful representation of a distribution when the interval widths are unequal.

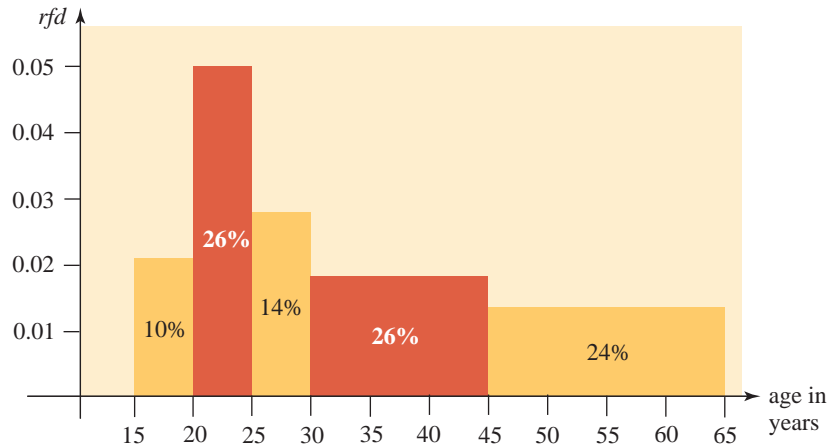


FIGURE 4.9 Ages of the people in Keith Reed’s acting class.

EXAMPLE 3

CONSTRUCTING A HISTOGRAM: GROUPED DATA To study the output of a machine that fills bags with corn chips, a quality control engineer randomly selected and weighed a sample of 200 bags of chips. The frequency distribution in Figure 4.10 summarizes the data. Construct a histogram for the weights of the bags of corn chips.

$x = \text{Weight (ounces)}$	$f = \text{Number of Bags}$
$15.3 \leq x < 15.5$	10
$15.5 \leq x < 15.7$	24
$15.7 \leq x < 15.9$	36
$15.9 \leq x < 16.1$	58
$16.1 \leq x < 16.3$	40
$16.3 \leq x < 16.5$	20
$16.5 \leq x < 16.7$	12

FIGURE 4.10 Weights of bags of corn chips.

SOLUTION

Because each interval has the same width ($\Delta x = 0.2$), we construct a combined frequency and relative frequency histogram. The relative frequencies are given in Figure 4.11.

x	f	$rf = f/n$
$15.3 \leq x < 15.5$	10	0.05
$15.5 \leq x < 15.7$	24	0.12
$15.7 \leq x < 15.9$	36	0.18
$15.9 \leq x < 16.1$	58	0.29
$16.1 \leq x < 16.3$	40	0.20
$16.3 \leq x < 16.5$	20	0.10
$16.5 \leq x < 16.7$	12	0.06
	$n = 200$	Sum = 1.00

FIGURE 4.11 Relative frequencies.

We now draw coordinate axes with appropriate scales and rectangles (Figure 4.12). Notice the (near) symmetry of the histogram. We will study this type of distribution in more detail in Section 4.4.

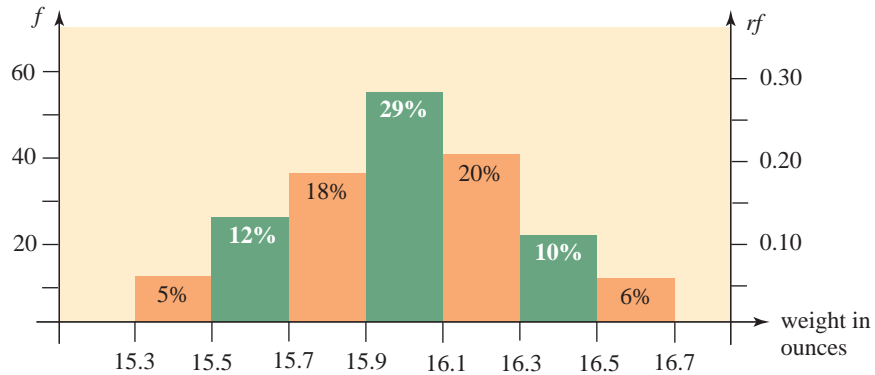


FIGURE 4.12 Weights of bags of corn chips.

Histograms and Single-Valued Classes

The histograms that we have constructed so far have all utilized intervals of grouped data. For instance, the first group of ages in Keith Reed’s acting class was from 16 to 24 years old ($16 \leq x < 24$). However, if a set of data consists of only a few distinct values, it may be advantageous to consider each distinct value to be a “class” of data; that is, we utilize single-valued classes of data.

EXAMPLE 4

CONSTRUCTING A HISTOGRAM: SINGLE VALUES A sample of high school seniors was asked, “How many television sets are in your house?” The frequency distribution in Figure 4.13 summarizes the data. Construct a histogram using single-valued classes of data.

Number of Television Sets	Frequency
0	2
1	13
2	18
3	11
4	5
5	1

FIGURE 4.13 Frequency distribution.

SOLUTION

Rather than using intervals of grouped data, we use a single value to represent each class. Because each class has the same width ($\Delta x = 1$), we construct a combined frequency and relative frequency histogram. The relative frequencies are given in Figure 4.14.

We now draw coordinate axes with appropriate scales and rectangles. In working with single valued classes of data, it is common to write the single value at the midpoint of the base of each rectangle as shown in Figure 4.15.

Number of Television Sets	Frequency	Relative Frequency
0	2	$2/50 = 4\%$
1	13	$13/50 = 26\%$
2	18	$18/50 = 36\%$
3	11	$11/50 = 22\%$
4	5	$5/50 = 10\%$
5	1	$1/50 = 2\%$
	$n = 50$	Total = 100%

FIGURE 4.14 Calculating relative frequency.

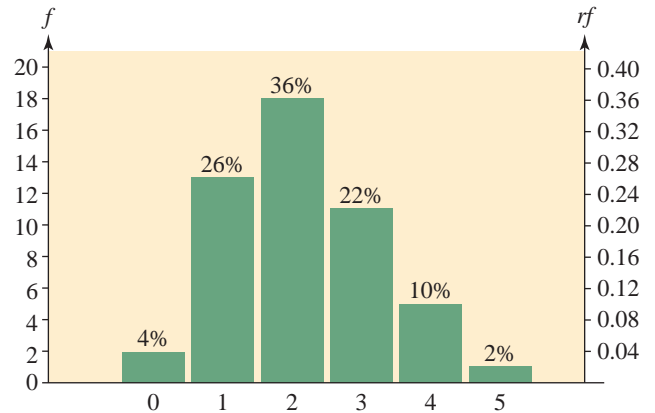


FIGURE 4.15 Number of television sets.

Pie Charts

Many statistical studies involve **categorical data**—that which is grouped according to some common feature or quality. One of the easiest ways to summarize categorical data is through the use of a **pie chart**. A pie chart shows how various categories of a set of data account for certain proportions of the whole. Financial incomes and expenditures are invariably shown as pie charts, as in Figure 4.16.

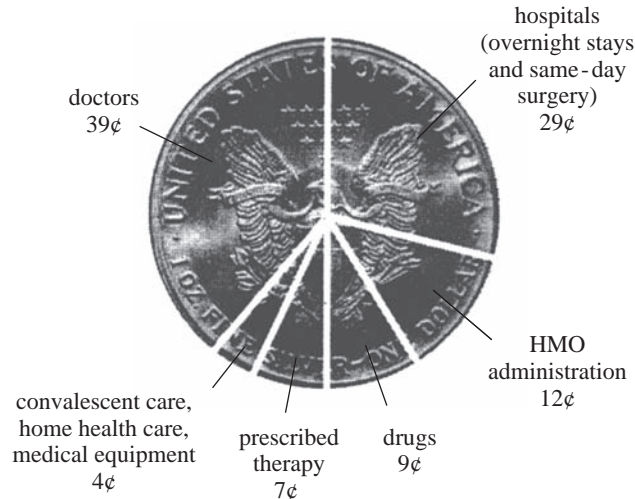


FIGURE 4.16 How a typical medical dollar is spent.

To draw the “slice” of the pie representing the relative frequency (percentage) of the category, the appropriate central angle must be calculated. Since a complete circle comprises 360 degrees, we obtain the required angle by multiplying 360° times the relative frequency of the category.

EXAMPLE 5

CONSTRUCTING A PIE CHART What type of academic degree do you hope to earn? The different types of degrees and the number of each type of degree conferred in the United States during the 2006–07 academic year is given in Figure 4.17. Construct a pie chart to summarize the data.

Type of Degree	Frequency (thousands)
Associate's	728
Bachelor's	1,524
Master's	605
Doctorate	61
Total	2,918

FIGURE 4.17 Academic degrees conferred, 2006–07. *Source:* National Center for Education Statistics.

SOLUTION

Find the relative frequency of each category and multiply it by 360° to determine the appropriate central angle. The necessary calculations are shown in Figure 4.18.

Type of Degree	Frequency (thousands)	Relative Frequency	Central Angle
Associate's	728	$\frac{728}{2,918} \approx 0.249$	$0.249 \times 360^\circ = 89.64^\circ$
Bachelor's	1,524	$\frac{1,524}{2,918} \approx 0.523$	$0.523 \times 360^\circ = 188.28^\circ$
Master's	605	$\frac{605}{2,918} \approx 0.207$	$0.207 \times 360^\circ = 74.52^\circ$
Doctorate	61	$\frac{61}{2,918} \approx 0.021$	$0.021 \times 360^\circ = 7.56^\circ$
	$n = 2,918$	Sum = 1.000	Total = 360°

FIGURE 4.18 Calculating relative frequency and central angles.

Now use a protractor to lay out the angles and draw the “slices.” The name of each category can be written directly on the slice, or, if the names are too long, a legend consisting of various shadings may be used. Each slice of the pie should contain its relative frequency, expressed as a percent. Remember, the whole reason for constructing a pie chart is to convey information visually; pie charts should enable the reader to instantly compare the relative proportions of categorical data. See Figure 4.19.

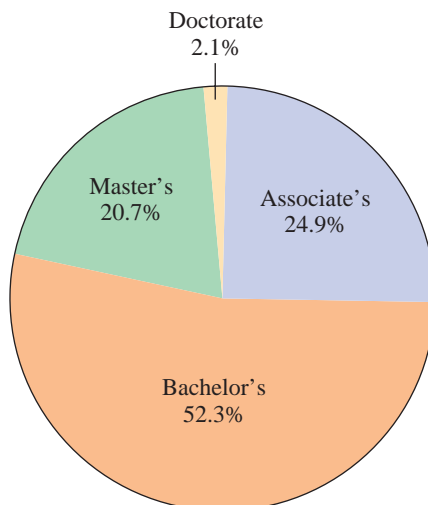


FIGURE 4.19 Academic degrees conferred, 2006–07.

4.1 EXERCISES

1. To study the library habits of students at a local college, thirty randomly selected students were surveyed to determine the number of times they had been to the library during the last week. The following results were obtained:

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

- a. Organize the given data by creating a frequency distribution.
 b. Construct a pie chart to represent the data.
 c. Construct a histogram using single-valued classes of data.

- ▶ 2. To study the eating habits of students at a local college, thirty randomly selected students were surveyed to determine the number of times they had purchased food at the school cafeteria during the last week. The following results were obtained:

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

- a. Organize the given data by creating a frequency distribution.
 b. Construct a pie chart to represent the data.
 c. Construct a histogram using single-valued classes of data.

3. To study the composition of families in Manistee, Michigan, forty randomly selected married couples were surveyed to determine the number of children in each family. The following results were obtained:

2 1 3 0 1 0 2 5 1 2 0 2 2 1
 4 3 1 1 3 4 1 1 0 2 0 0 2 2
 1 0 3 1 1 3 4 2 1 3 0 1

- a. Organize the given data by creating a frequency distribution.
 b. Construct a pie chart to represent the data.
 c. Construct a histogram using single-valued classes of data.

- ▶ 4. To study the spending habits of shoppers in Orlando, Florida, fifty randomly selected shoppers at a mall were surveyed to determine the number of credit cards they carried. The following results were obtained:

2 5 0 4 2 1 0 6 3 5 4 3 4 0
 5 2 5 2 0 2 5 0 2 5 2 5 4 3
 5 6 1 0 6 3 5 3 4 0 5 2 2 5
 2 0 2 0 4 2 1 0

- a. Organize the given data by creating a frequency distribution.
 b. Construct a pie chart to represent the data.
 c. Construct a histogram using single-valued classes of data.

5. The speeds, in miles per hour, of forty randomly monitored cars on Interstate 40 near Winona, Arizona, were as follows:

66 71 76 61 73 78 74 67 80 63 69
 78 66 70 77 60 72 58 65 70 64 75
 80 75 62 67 72 59 74 65 54 69 73
 79 64 68 57 51 68 79

- a. Organize the given data by creating a frequency distribution. (Group the data into six intervals.)
 b. Construct a histogram to represent the data.

- ▶ 6. The weights, in pounds, of thirty-five packages of ground beef at the Cut Above Market were as follows:

1.0 1.9 2.5 1.2 2.0 0.7 1.3 2.4 1.1
 3.3 2.4 0.8 2.3 1.7 1.0 2.8 1.4 3.0
 0.9 1.1 1.4 2.2 1.5 3.2 2.1 2.7 1.8
 1.6 2.3 2.6 1.3 2.9 1.9 1.2 0.5

- a. Organize the given data by creating a frequency distribution. (Group the data into six intervals.)
 b. Construct a histogram to represent the data.

7. To examine the effects of a new registration system, a campus newspaper asked freshmen how long they had to wait in a registration line. The frequency distribution in Figure 4.20 summarizes the responses. Construct a histogram to represent the data.

$x = \text{Time (minutes)}$	Number of Freshmen
$0 \leq x < 10$	32
$10 \leq x < 20$	47
$20 \leq x < 30$	36
$30 \leq x < 40$	22
$40 \leq x < 50$	13
$50 \leq x < 60$	10

FIGURE 4.20 Time waiting in line.

- 8. The frequency distribution shown in Figure 4.21 lists the annual salaries of the managers at Universal Manufacturing of Melonville. Construct a histogram to represent the data.

$x = \text{Salary}$ (thousands of dollars)	Number of Managers
$30 \leq x < 40$	6
$40 \leq x < 50$	12
$50 \leq x < 60$	10
$60 \leq x < 70$	5
$70 \leq x < 80$	7
$80 \leq x < 90$	3

FIGURE 4.21 Annual salaries.

9. The frequency distribution shown in Figure 4.22 lists the hourly wages of the workers at Universal Manufacturing of Melonville. Construct a histogram to represent the data.

$x = \text{Hourly Wage}$ (dollars)	Number of Employees
$8.00 \leq x < 9.50$	21
$9.50 \leq x < 11.00$	35
$11.00 \leq x < 12.50$	42
$12.50 \leq x < 14.00$	27
$14.00 \leq x < 15.50$	18
$15.50 \leq x < 17.00$	9

FIGURE 4.22 Hourly wages.

- 10. To study the output of a machine that fills boxes with cereal, a quality control engineer weighed 150 boxes of Brand X cereal. The frequency distribution in Figure 4.23 summarizes her findings. Construct a histogram to represent the data.

$x = \text{Weight}$ (ounces)	Number of Boxes
$15.3 \leq x < 15.6$	13
$15.6 \leq x < 15.9$	24
$15.9 \leq x < 16.2$	84
$16.2 \leq x < 16.5$	19
$16.5 \leq x < 16.8$	10

FIGURE 4.23 Weights of boxes of cereal.

11. The ages of the nearly 4 million women who gave birth in the United States in 1997 are given in Figure 4.24. Construct a histogram to represent the data.

Age (years)	Number of Women
$15 \leq x < 20$	486,000
$20 \leq x < 25$	948,000
$25 \leq x < 30$	1,075,000
$30 \leq x < 35$	891,000
$35 \leq x < 40$	410,000
$40 \leq x < 45$	77,000
$45 \leq x < 50$	4,000

FIGURE 4.24 Ages of women giving birth in 1997. *Source:* U.S. Bureau of the Census.

- 12. The age composition of the population of the United States in the year 2000 is given in Figure 4.25. Replace the interval “85 and over” with the interval $85 \leq x \leq 100$ and construct a histogram to represent the data.

Age (years)	Number of People (thousands)
$0 < x < 5$	19,176
$5 \leq x < 10$	20,550
$10 \leq x < 15$	20,528
$15 \leq x < 25$	39,184
$25 \leq x < 35$	39,892
$35 \leq x < 45$	44,149
$45 \leq x < 55$	37,678
$55 \leq x < 65$	24,275
$65 \leq x < 85$	30,752
85 and over	4,240

FIGURE 4.25 Age composition of the population of the United States in the year 2000. *Source:* U.S. Bureau of the Census.

In Exercises 13 and 14, use the age composition of the 14,980,000 students enrolled in institutions of higher education in the United States during 2000, as given in Figure 4.26.

13. Using the data in Figure 4.26, replace the interval “35 and over” with the interval $35 \leq x \leq 60$ and construct a histogram to represent the male data.

Age of Males (years)	Number of Students	Age of Females (years)	Number of Students
$14 \leq x < 18$	94,000	$14 \leq x < 18$	78,000
$18 \leq x < 20$	1,551,000	$18 \leq x < 20$	1,907,000
$20 \leq x < 22$	1,420,000	$20 \leq x < 22$	1,597,000
$22 \leq x < 25$	1,091,000	$22 \leq x < 25$	1,305,000
$25 \leq x < 30$	865,000	$25 \leq x < 30$	1,002,000
$30 \leq x < 35$	521,000	$30 \leq x < 35$	664,000
35 and over	997,000	35 and over	1,888,000
Total	6,539,000	Total	8,441,000

FIGURE 4.26 Age composition of students in higher education. *Source:* U.S. National Center for Education Statistics.

- 14. Using the data in Figure 4.26, replace the interval “35 and over” with the interval $35 \leq x \leq 60$ and construct a histogram to represent the female data.
15. The frequency distribution shown in Figure 4.27 lists the ages of 200 randomly selected students who received a bachelor’s degree at State University last year. Where possible, determine what percent of the graduates had the following ages:
- less than 23
 - at least 31
 - at most 20
 - not less than 19
 - at least 19 but less than 27
 - not between 23 and 35

Age (years)	Number of Students
$10 \leq x < 15$	1
$15 \leq x < 19$	4
$19 \leq x < 23$	52
$23 \leq x < 27$	48
$27 \leq x < 31$	31
$31 \leq x < 35$	16
$35 \leq x < 39$	29
39 and over	19

FIGURE 4.27 Age of students.

- 16. The frequency distribution shown in Figure 4.28 lists the number of hours per day a randomly selected sample of teenagers spent watching television. Where possible, determine what percent of the teenagers spent the following number of hours watching television:
- less than 4 hours
 - at least 5 hours

- at least 1 hour
- less than 2 hours
- at least 2 hours but less than 4 hours
- more than 3.5 hours

Hours per Day	Number of Teenagers
$0 \leq x < 1$	18
$1 \leq x < 2$	31
$2 \leq x < 3$	24
$3 \leq x < 4$	38
$4 \leq x < 5$	27
$5 \leq x < 6$	12
$6 \leq x < 7$	15

FIGURE 4.28 Time watching television.

- 17. Figure 4.29 lists the top five reasons given by patients for emergency room visits in 2006. Construct a pie chart to represent the data.

Reason	Number of Patients (thousands)
Stomach pain	8,057
Chest pain	6,392
Fever	4,485
Headache	3,354
Shortness of breath	3,007

FIGURE 4.29 Reasons for emergency room visits, 2006. *Source:* National Center for Health Statistics.

18. Figure 4.30 lists the world's top six countries as tourist destinations in 2007. Construct a pie chart to represent the data.

Country	Number of Arrivals (millions)
France	81.9
Spain	59.2
United States	56.0
China	54.7
Italy	43.7
United Kingdom	30.7

FIGURE 4.30 World's top tourist destinations, 2007. *Source:* World Tourism Organization.

- ▶ 19. Figure 4.31 lists the race of new AIDS cases in the United States in 2005.

Race	Male	Female
White	10,027	1,747
Black	13,048	7,093
Hispanic	5,949	1,714
Asian	389	92
Native American	137	45

FIGURE 4.31 New AIDS cases in the United States, 2005. *Source:* National Center for Health Statistics.

- Construct a pie chart to represent the male data.
 - Construct a pie chart to represent the female data.
 - Compare the results of parts (a) and (b). What conclusions can you make?
 - Construct a pie chart to represent the total data.
- ▶ 20. Figure 4.32 lists the types of accidental deaths in the United States in 2006. Construct a pie chart to represent the data.
- ▶ 21. Figure 4.33 lists some common specialties of physicians in the United States in 2005.
- Construct a pie chart to represent the male data.
 - Construct a pie chart to represent the female data.
 - Compare the results of parts (a) and (b). What conclusions can you make?
 - Construct a pie chart to represent the total data.

Type of Accident	Number of Deaths
Motor vehicle	44,700
Poison	25,300
Falls	21,200
Suffocation	4,100
Drowning	3,800
Fire	2,800
Firearms	680

FIGURE 4.32 Types of accidental deaths in the United States, 2006. *Source:* National Safety Council.

Specialty	Male	Female
Family practice	54,022	26,305
General surgery	32,329	5,173
Internal medicine	104,688	46,245
Obstetric/gynecology	24,801	17,258
Pediatrics	33,515	36,636
Psychiatry	27,213	13,079

FIGURE 4.33 Physicians by Gender and Specialty, 2005. *Source:* American Medical Association.

- ▶ 22. Figure 4.34 lists the major metropolitan areas on intended residence for immigrants admitted to the United States in 2002. Construct a pie chart to represent the data.

Metropolitan Area	Number of Immigrants
Los Angeles/Long Beach, CA	100,397
New York, NY	86,898
Chicago, IL	41,616
Miami, FL	39,712
Washington DC	36,371

FIGURE 4.34 Immigrants and areas of residence, 2002. *Source:* U.S. Immigration and Naturalization Service.



Answer the following questions using complete sentences and your own words.

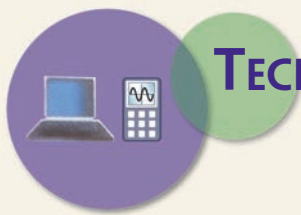
• CONCEPT QUESTIONS

23. Explain the meanings of the terms *population* and *sample*.
24. The cholesterol levels of the 800 residents of Land-o-Lakes, Wisconsin, were recently collected and organized in a frequency distribution. Do these data represent a sample or a population? Explain your answer.
25. Explain the difference between frequency, relative frequency, and relative frequency density. What does each measure?
26. When is relative frequency density used as the vertical scale in constructing a histogram? Why?
27. In some frequency distributions, data are grouped in intervals; in others, they are not.
 - a. When should data be grouped in intervals?
 - b. What are the advantages and disadvantages of using grouped data?



HISTOGRAMS ON A GRAPHING CALCULATOR

In Example 2 of this section, we created a frequency distribution and a histogram for the ages of the students in an acting class. Much of this work can be done on a computer or a graphing calculator.



TECHNOLOGY AND STATISTICAL GRAPHS

Entering the Data

To enter the data from Example 2, do the following.

L1	L2	L3
49		
59		
19		
20		
23		
24		
L1(43)=		

FIGURE 4.35

A TI-83/84's list screen.

ENTERING THE DATA ON A TI-83/84:

- Put the calculator into statistics mode by pressing **STAT**.
- Set the calculator up for entering the data by selecting “Edit” from the “EDIT” menu, and the “list screen” appears, as shown in Figure 4.35. If data already appear in a list (as they do in list L_1 in Figure 4.35), use the arrow buttons to highlight the name of the list (i.e., “ L_1 ” or “xStat”) and press **CLEAR** **ENTER**.
- Enter the students' ages in list L_1 , in any order, using the arrow buttons and the **ENTER** button. When this has been completed, your screen should look similar to that in Figure 4.35. Notice the “ $L_1(43)=$ ” at the bottom of the screen; this indicates that 42 entries have been made, and the calculator is ready to receive the 43rd. This allows you to check whether you have left any entries out.

Note: If some data points frequently recur, you can enter the data points in list L_1 and their frequencies in list L_2 rather than reentering a data point each time it recurs.
- Press **2nd** **QUIT**.

ENTERING THE DATA ON A CASIO:

- Put the calculator into statistics mode by pressing **MENU**, highlighting **STAT**, and pressing **EXE**.

List 1	List 2	List 3	List 4
39	19		
40	20		
41	23		
42	24		
43			

FIGURE 4.36

A Casio's list screen.

- Enter the data in List 1 in any order, using the arrow buttons and the **EXE** button. See Figure 4.36. If data already appear in a list and you want to clear the list, use the arrow keys to move to that list, press **F6** and then **F4** (i.e., **DEL-A**, which stands for “delete all”), and then press **F1** (i.e., **YES**). Notice that the entries are numbered; this allows you to check whether you have left any entries out.

Drawing a Histogram

Once the data have been entered, you can draw a histogram.

DRAWING A HISTOGRAM ON A TI-83/84:

- Press **Y=** and clear any functions that may appear.
- Enter the group boundaries by pressing **WINDOW**, entering the left boundary of the first group as xmin (16 for this problem), the right boundary of the last group plus 1 as xmax ($64 + 1 = 65$ for this problem), and the group width as xscl (8 for this problem). (The calculator will create histograms only with equal group widths.) Enter 0 for ymin, and the largest frequency for ymax. (You may guess; it's easy to change it later if you guess wrong.)
- Set the calculator up to draw a histogram by pressing **2nd** **STAT PLOT** and selecting “Plot 1.” Turn the plot on and select the histogram icon.
- Tell the calculator to put the data entered in list L_1 on the x -axis by selecting “ L_1 ” for “Xlist,” and to consider each entered data point as having a frequency of 1 by selecting “1” for “Freq.”

Note: If some data points frequently recur and you entered their frequencies in list L_2 , then select “ L_2 ” rather than “1” for “Freq” by typing **2nd** **L_2** .

- Draw a histogram by pressing **GRAPH**. If some of the bars are too long or too short for the screen, alter ymin accordingly.
- Press **TRACE** to find out the left and right boundaries and the frequency of the bars, as shown in Figure 4.37. Use the arrow buttons to move from bar to bar.
- Press **2nd** **STAT PLOT**, select “Plot 1” and turn the plot off, or else the histogram will appear on future graphs.

DRAWING A HISTOGRAM ON A CASIO:

- Press **GRPH** (i.e., **F1**).
- Press **SET** (i.e., **F6**).
- Make the resulting screen, which is labeled “StatGraph1,” read as follows:

```
Graph Type   :Hist
Xlist        :List1
Frequency    : 1
```

To make the screen read as described above, do the following:

- Use the down arrow button to scroll down to “Graph Type.”
- Press **F6**.
- Press **HIST** (i.e., **F1**).
- In a similar manner, change “Xlist” and “Frequency” if necessary.

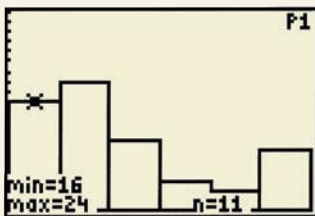


FIGURE 4.37

The first bar's boundaries are 16 and 24; its frequency is 11.

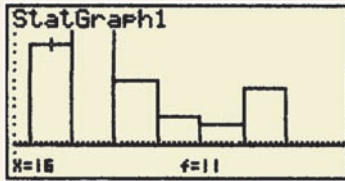
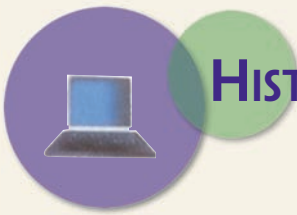


FIGURE 4.38

The first bar's left boundary is 16; its frequency is 11.

- Press **EXE** and return to List1.
- Press **GPH1** (i.e., **F1**) and make the resulting screen, which is labeled “Set Interval,” read as follows:
 - Start: 16 “Start” refers to the beginning of the first group
 - Pitch: 8 “Pitch” refers to the width of each group
- Press **DRAW** (i.e., **F6**), and the calculator will display the histogram.
- Press **SHIFT** and then **TRCE** (i.e., **F1**) to find out the left boundaries and the frequency of the bars. Use the arrow buttons to move from bar to bar. See Figure 4.38.



HISTOGRAMS AND PIE CHARTS ON A COMPUTERIZED SPREADSHEET

A **spreadsheet** is a large piece of paper marked off in rows and columns. Accountants use spreadsheets to organize numerical data and perform computations. A **computerized spreadsheet**, such as Microsoft Excel, is a computer program that mimics the appearance of a paper spreadsheet. It frees the user from performing any computations; instead, it allows the user merely to give instructions on how to perform those computations. The instructions in this subsection were specifically written for Microsoft Excel; however, all computerized spreadsheets work somewhat similarly.

When you start a computerized spreadsheet, you see something that looks like a table waiting to be filled in. The rows are labeled with numbers and the columns with letters, as shown in Figure 4.39.

◇	A	B	C	D
1				
2				
3				
4				
5				

FIGURE 4.39 A blank spreadsheet.

The individual boxes are called **cells**. The cell in column A row 1 is called cell A1; the cell below it is called cell A2, because it is in column A row 2.

A computerized spreadsheet is an ideal tool to use in creating a histogram or a pie chart. We will illustrate this process by preparing both a histogram and a pie chart for the ages of the students in Keith Reed's acting class, as discussed in Example 2.

Entering the Data

1. *Label the columns.* Use the mouse and/or the arrow buttons to move to cell A1, type in “age of student” and press “return” or “enter.” (If there were other data, we could

enter it in other columns. For example, if the students' names were included, we could type "name of student" in cell A1, and "age of student" in cell B1.)

2. *Enter the students' ages in column A.* Move to cell A2, type in "26" and press "return" or "enter." Move to cell A3, type in "16" and press "return" or "enter." In a similar manner, enter all of the ages. You can enter the ages in any order. After you complete this step, your spreadsheet should look like that in Figure 4.40 (except that it should go down a lot further).

	A	B	C	D
1	age of student			
2	26			
3	16			
4	21			
5	34			

FIGURE 4.40 The spreadsheet after entering the students' ages.

3. *Save the spreadsheet.* Use your mouse to select "File" at the very top of the screen. Then pull your mouse down until "Save As" is highlighted, and let go. Your instructor may give you further instructions on where and how to save your spreadsheet.

Preparing the Data for a Chart

	A	B
1	age of student	bin numbers
2	26	23
3	16	31
4	21	39
5	34	47
6	45	55
7	18	63
8	41	

FIGURE 4.41

The spreadsheet after entering the bin numbers.

1. *Enter the group boundaries.* Excel uses "bin numbers" rather than group boundaries. A group's bin number is the highest number that should be included in that group. In Example 2, the first group was $16 \leq x < 24$. The highest age that should be included in this group is 23, so the group's bin number is 23. Be careful; this group's bin number is not 24, since people who are 24 years old should be included in the second group.
 - If necessary, use the up arrow in the upper-right corner of the spreadsheet to scroll to the top 60 of the spreadsheet.
 - Type "bin numbers" in cell B1.
 - Determine each group's bin number, and enter them in column B.

After you complete this step, your spreadsheet should look like that in Figure 4.41. You might need to adjust the width of column A. Click on the right edge of the "A" label at the top of the first column, and move it.

2. *Have Excel determine the frequencies.*
 - Use your mouse to select "Tools" at the very top of the screen. Then pull your mouse down until "Data Analysis" is highlighted, and let go.
 - If "Data Analysis" is not listed under "Tools," then
 - Select "Tools" at the top of the screen, pull down until "Add-Ins" is highlighted, and let go.
 - Select "Analysis ToolPak-VBA."
 - Use your mouse to press the "OK" button.
 - Select "Tools" at the top of the screen, pull down until "Data Analysis" is highlighted, and let go.
 - In the "Data Analysis" box that appears, use your mouse to highlight "Histogram" and press the "OK" button.
 - In the "Histogram" box that appears, use your mouse to click on the white rectangle that follows "Input Range," and then use your mouse to draw a box around all of the ages. (To draw the box, move your mouse to cell A2, press the mouse button, and move the mouse down until all of the entered ages are enclosed in a box.) This should cause "\$A\$2:\$A\$43" to appear in the Input Range rectangle.

- Use your mouse to click on the white rectangle that follows “Bin Range,” and then use your mouse to draw a box around all of the bin numbers. This should cause “\$B\$2:\$B\$7” to appear in the Bin Range rectangle.
- Be certain that there is a dot in the button to the left of “Output Range.” Then click on the white rectangle that follows “Output Range,” and use your mouse to click on cell C1. This should cause “\$C\$1” to appear in the Output Range rectangle.
- Use your mouse to press the “OK” button.

After you complete this step, your spreadsheet should look like that in Figure 4.42.

	A	B	C	D
1	age of student	bin numbers	<i>Bin</i>	<i>Frequency</i>
2	26	23	23	11
3	16	31	31	13
4	21	39	39	7
5	34	47	47	3
6	45	55	55	2
7	18	63	63	6
8	41		More	0
9	38			

FIGURE 4.42 The spreadsheet after Excel determines the frequencies.

3. *Prepare labels for the chart.*

- In column E, list the group boundaries.
- In column F, list the frequencies.

After you complete this step, your spreadsheet should look like that in Figure 4.43 (the first few columns are not shown).

C	D	E	F
<i>Bin</i>	<i>Frequency</i>	group boundaries	frequency
23	11	$16 \leq x < 24$	11
31	13	$24 \leq x < 32$	13
39	7	$32 \leq x < 40$	7
47	3	$40 \leq x < 48$	3
55	2	$48 \leq x < 56$	2
63	6	$56 \leq x < 64$	6
More	0		

FIGURE 4.43 The spreadsheet after preparing labels.

Drawing a Histogram

1. *Use the Chart Wizard to draw a bar chart.*

- Use your mouse to press the “Chart Wizard” button at the top of the spreadsheet. (It looks like a histogram and it might have a magic wand.)
- Select “Column” and then an appropriate style.

- Press the “Next” button.
- Use your mouse to click on the white rectangle that follows “Data range,” and then use your mouse to draw a box around all of the group boundaries and frequencies from step 3 in “Preparing the data for a chart”. This should cause “=Sheet1!\$E\$2:\$F\$7” to appear in the Data range rectangle.
- Press the “Next” button.
- Under “Chart Title” type an appropriate title, such as “Acting class ages.”
- After “Category (X)” type an appropriate title for the x -axis, such as “Students’ ages.”
- After “Value (Y)” type an appropriate title for the y -axis, such as “frequencies.”
- Press “Legend” at the top of the chart options box, and then remove the check mark next to “show legend.”
- Press the “Finish” button and the bar chart will appear. See Figure 4.44.

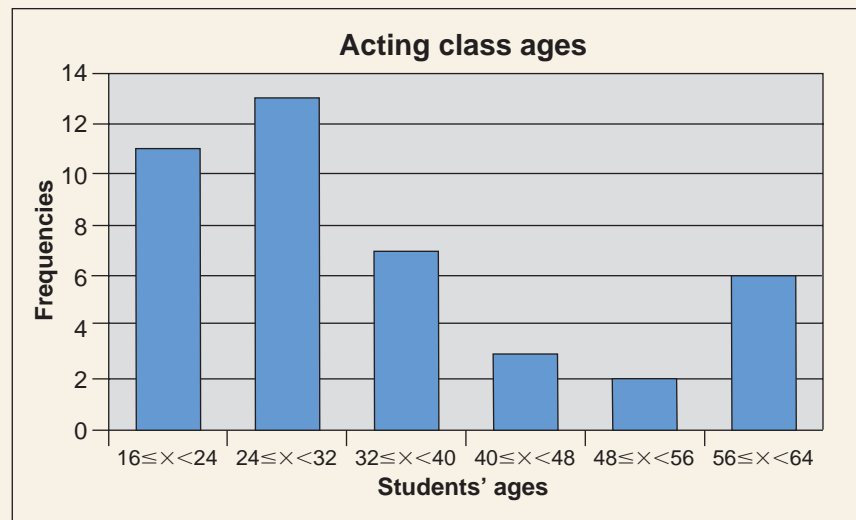


FIGURE 4.44 The bar chart.

2. *Save the spreadsheet.* Use your mouse to select “File” at the very top of the screen. Then pull your mouse down until “Save” is highlighted, and let go.
3. *Convert the bar graph to a histogram.* The graph is not a histogram because of the spaces between the bars.
 - Double click or right click on the bar graph and a “Format Data Series” box will appear.
 - Press the “Options” tab.
 - Remove the spaces between the bars by changing the “Gap width” to 0.
 - Press OK.
 - Save the spreadsheet. See Figure 4.45.
4. *Print the histogram.*
 - Click on the histogram, and it will be surrounded by a thicker border than before.
 - Use your mouse to select “File” at the very top of the screen. Then pull your mouse down until “Print” is highlighted, and let go.
 - Respond appropriately to the “Print” box that appears.

Drawing a Pie Chart

1. *Use the Chart Wizard to draw the pie chart.*
 - Use your mouse to press the “Chart Wizard” button at the top of the spreadsheet. (It looks like a histogram and it might have a magic wand.)

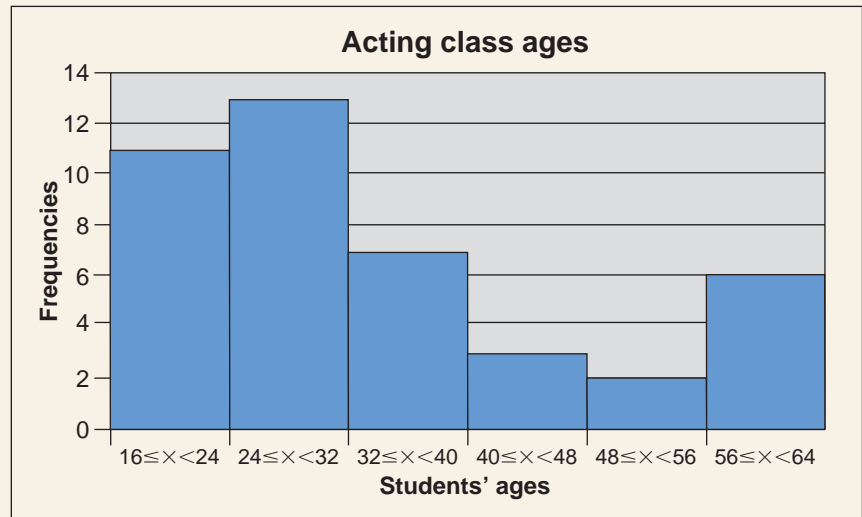


FIGURE 4.45 The histogram.

- Select “Pie” and then an appropriate style.
- Press the “Next” button.
- Use your mouse to click on the white rectangle that follows “Data range,” and then use your mouse to draw a box around all of the group boundaries and frequencies from step 3 in “Preparing the data for a chart.” This should cause “=Sheet1!\$E\$2:\$F\$7” to appear in the Data range rectangle.
- Press the “Next” button.
- Under “Chart Title” type an appropriate title.
- Press the “Finish” button and the pie chart will appear. See Figure 4.46.

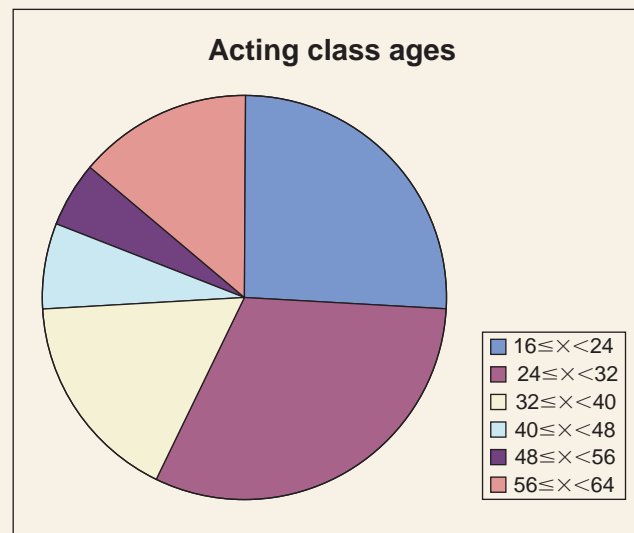


FIGURE 4.46 The pie chart.

2. *Save the spreadsheet.* Use your mouse to select “File” at the very top of the screen. Then pull your mouse down until “Save” is highlighted, and let go.
3. *Print the pie chart.*
 - Click on the chart, and it will be surrounded by a thicker border than before.
 - Use your mouse to select “File” at the very top of the screen. Then pull your mouse down until “Print” is highlighted, and let go.
 - Respond appropriately to the “Print” box that appears.

EXERCISES

Use a graphing calculator or Excel on the following exercises. Answers will vary depending on your choice of groups.

Graphing calculator instructions: By hand, copy a graphing calculator's histogram onto paper for submission.

Excel instructions: Print out a histogram for submission.

28. Do Double Stuf Oreos really contain twice as much “stuf”? In 1999, Marie Revak and Jihan Williams’s “interest was piqued by the unusual spelling of the word stuff and the bold statement ‘twice the filling’ on the package.” So they conducted an experiment in which they weighed the amount of filling in a sample of traditional Oreos and Double Stuf Oreos. The data from that experiment are in Figure 4.47 (weight in grams). An unformatted spreadsheet containing these data can be downloaded from the text web site: www.cengage.com/math/johnson



- a. Create one histogram for Double Stuf Oreos and one for Traditional Oreos. Use the same categories for each histogram.
- b. Compare the results of part (a). What do they tell you about the two cookies?
29. Every year, CNNMoney.com lists “the best places to live.” They compare the “best small towns” on several bases, one of which is the “quality of life.”

See Figure 4.48. An unformatted spreadsheet containing these data can be downloaded from the text web site: www.cengage.com/math/johnson

- a. Find the mean of the air quality indices.
- b. Would the population standard deviation or the sample standard deviation be more appropriate here? Why?
- c. Find the standard deviation from part (b).
- d. What do your results say about the air quality index of all of the best places to live in the United States?
30. Every year, CNNMoney.com lists “the best places to live.” They compare the “best small towns” on several bases, one of which is the “quality of life.” See Figure 4.48. An unformatted spreadsheet containing these data can be downloaded from the text web site: www.cengage.com/math/johnson
- a. Find the mean of the commute times.
- b. Would the population standard deviation or the sample standard deviation be more appropriate here? Why?
- c. Find the standard deviation from part (b).
- d. What do your results say about commuting in all of the best places to live in the United States?
31. Every year, CNNMoney.com lists “the best places to live.” They compare the “best small towns” on several bases, one of which is the “quality of life.” See Figure 4.48. An unformatted spreadsheet containing these data can be downloaded from the text web site: www.cengage.com/math/johnson
- a. Find the mean of the property crime levels.
- b. Would the population standard deviation or the sample standard deviation be more appropriate here? Why?
- c. Find the standard deviation from part (b).
- d. What do your results say about property crime in all of the best places to live in the United States?

Double Stuf Oreos:																		
4.7	6.5	5.5	5.6	5.1	5.3	5.4	5.4	3.5	5.5	6.5	5.9	5.4	4.9	5.6	5.7	5.3	6.9	6.5
6.3	4.8	3.3	6.4	5.0	5.3	5.5	5.0	6.0	5.7	6.3	6.0	6.3	6.1	6.0	5.8	5.8	5.9	6.2
5.9	6.5	6.5	6.1	5.8	6.0	6.2	6.2	6.0	6.8	6.2	5.4	6.6	6.2					
Traditional Oreos:																		
2.9	2.8	2.6	3.5	3.0	2.4	2.7	2.4	2.5	2.2	2.6	2.6	2.9	2.6	2.6	3.1	2.9	2.4	2.8
3.8	3.1	2.9	3.0	2.1	3.8	3.0	3.0	2.8	2.9	2.7	3.2	2.8	3.1	2.7	2.8	2.6	2.6	3.0
2.8	3.5	3.3	3.3	2.8	3.1	2.6	3.5	3.5	3.1	3.1								

FIGURE 4.47 Amount of Oreo filling. Source: Revak, Marie A. and Jihan G. Williams, “Sharing Teaching Ideas: The Double Stuf Dilemma,” *The Mathematics Teacher*, November 1999, Volume 92, Issue 8, page 674.

Rank	City	Air quality index (% days AQI ranked good)	Property crime— incidents per 1000	Median commute time (mins)
1	Louisville, CO	74.0%	15	19.5
2	Chanhasen, MN	N.A.	16	23.1
3	Papillion, NE	92.0%	16	19.3
4	Middleton, WI	76.0%	27	16.3
5	Milton, MA	81.0%	9	27.7
6	Warren, NJ	N.A.	21	28.7
7	Keller, TX	65.0%	14	29.4
8	Peachtree City, GA	84.0%	12	29.3
9	Lake St. Louis, MO	79.0%	16	26.8
10	Mukilteo, WA	89.0%	27	22.3

FIGURE 4.48 The best places to live.

- 32.** Figure 4.49 gives EPA fuel efficiency ratings for 2010 family sedans with an automatic transmission and the smallest available engine. An unformatted spreadsheet containing these data can be downloaded from the text web site: www.cengage.com/math/johnson
- a.** For each of the two manufacturing regions, find the mean and an appropriate standard deviation for city driving, and defend your choice of standard deviations.
- b.** What do the results of part (a) tell you about the two regions?
- 33.** Figure 4.49 gives EPA fuel efficiency ratings for 2010 family sedans with an automatic transmission and the smallest available engine. An unformatted spreadsheet containing these data can be downloaded from the text web site: www.cengage.com/math/johnson
- a.** For each of the two manufacturing regions, find the mean and an appropriate standard deviation for highway driving, and defend your choice of standard deviations.
- b.** What do the results of part (a) tell you about the two regions?

American cars	city mpg	highway mpg	Asian cars	city mpg	highway mpg
Ford Fusion hybrid	41	36	Toyota Prius	51	48
Chevrolet Malibu hybrid	26	34	Nissan Altima hybrid	35	33
Ford Fusion fwd	24	34	Toyota Camry Hybrid	33	34
Mercury Milan	23	34	Kia Forte	27	36
Chevrolet Malibu	22	33	Hyundai Elantra	26	35
Saturn Aura	22	33	Nissan Altima	23	32
Dodge Avenger	21	30	Subaru Legacy awd	23	31
Buick Lacrosse	17	27	Toyota Camry	22	33
			Hyundai Sonata	22	32
			Kia Optima	22	32
			Honda Accord	22	31
			Mazda 6	21	30
			Mitsubishi Galant	21	30
			Hyunda Azera	18	26

FIGURE 4.49 Fuel efficiency. Source: www.fueleconomy.gov

4.2

Measures of Central Tendency

OBJECTIVES

- Find the mean
- Find the median
- Find the mode

Who is the best running back in professional football? How much does a typical house cost? What is the most popular television program? The answers to questions like these have one thing in common: They are based on averages. To compare the capabilities of athletes, we compute their average performances. This computation usually involves the ratio of two totals, such as (total yards gained)/(total number of carries) = average gain per carry. In real estate, the average-price house is found by listing the prices of all houses for sale (from lowest to highest) and selecting the price in the middle. Television programs are rated by the average number of households tuned in to each particular program.

Rather than listing every data point in a large distribution of numbers, people tend to summarize the data by selecting a representative number, calling it the average. Three figures—the *mean*, the *median*, and the *mode*—describe the “average” or “center” of a distribution of numbers. These averages are known collectively as the **measures of central tendency**.

The Mean



© Reuters/Corbis

In 2001, Barry Bonds of the San Francisco Giants hit 73 home runs and set an all-time record in professional baseball. On average, how many home runs per season did Bonds hit? See Exercise 13.

The **mean** is the average people are most familiar with; it can also be the most misleading. Given a collection of n data points, x_1, x_2, \dots, x_n , the mean is found by adding up the data and dividing by the number of data points:

$$\text{mean of } n \text{ data points} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

If the data are collected from a sample, then the mean is denoted by \bar{x} (read “ x bar”); if the data are collected from an entire population, then the mean is denoted by μ (lowercase Greek letter “mu”). Unless stated otherwise, we will assume that the data represent a sample, so the mean will be symbolized by \bar{x} .

Mathematicians have developed a type of shorthand, called **summation notation**, to represent the sum of a collection of numbers. The Greek letter Σ (“sigma”) corresponds to the letter S and represents the word *sum*. Given a group of data points x_1, x_2, \dots, x_n , we use the symbol Σx to represent their sum; that is, $\Sigma x = x_1 + x_2 + \cdots + x_n$.

DEFINITION OF THE MEAN

Given a sample of n data points, x_1, x_2, \dots, x_n , the **mean**, denoted by \bar{x} , is

$$\bar{x} = \frac{\Sigma x}{n} \quad \text{or} \quad \bar{x} = \frac{\text{the sum of the data points}}{\text{the number of data points}}$$

Many scientific calculators have statistical functions built into them. These functions allow you to enter the data points and press the “ \bar{x} -bar” button to obtain the mean. Consult your manual for specific instructions.

EXAMPLE 1

FINDING THE MEAN: SINGLE VALUES AND MIXED UNITS In 2005, Lance Armstrong won his seventh consecutive Tour de France bicycle race. No one in the 100-year history of the race has won so many times. (Miguel Indurain of Spain won five times from 1991 to 1995.) Lance’s winning times are given in Figure 4.50. Find the mean winning time of Lance Armstrong’s Tour de France victory rides.

Year	1999	2000	2001	2002	2003	2004	2005
Time (h:m:s)	91:32:16	92:33:08	86:17:28	82:05:12	83:41:12	83:36:02	86:15:02

FIGURE 4.50 Lance Armstrong’s Tour de France winning times. *Source: San Francisco Chronicle.*

SOLUTION

To find the mean, we must add up the data and divide by the number of data points. However, before we can sum the data, they must be converted to a common unit, say, minutes. Therefore, we multiply the number of hours by 60 (minutes per hour), divide the number of seconds by 60 (seconds per minute), and add each result to the number of minutes. Lance Armstrong’s 1999 winning time of 91 hours, 32 minutes, 16 seconds is converted to minutes as follows:

$$\begin{aligned} & \left(91 \text{ hours} \times \frac{60 \text{ minutes}}{1 \text{ hour}} \right) + 32 \text{ minutes} + \left(16 \text{ seconds} \times \frac{1 \text{ minute}}{60 \text{ seconds}} \right) \\ &= 5,460 \text{ minutes} + 32 \text{ minutes} + 0.266666 \dots \text{ minutes} \\ &= 5,492.267 \text{ minutes} \quad \text{rounding off to three decimal places} \end{aligned}$$

In a similar fashion, all of Lance Armstrong’s winning times are converted to minutes, and the results are given in Figure 4.51.

Year	1999	2000	2001	2002	2003	2004	2005
Time (min)	5492.267	5553.133	5177.467	4925.200	5021.200	5016.033	5175.033

FIGURE 4.51 Winning time in minutes.

$$\begin{aligned} \bar{x} &= \frac{\Sigma x}{n} \\ &= \frac{5492.267 + 5553.133 + 5177.467 + 4925.200 + 5021.200 + 5016.033 + 5175.033}{7} \\ &= \frac{36360.333}{7} \\ &= 5194.333 \text{ minutes} \end{aligned}$$

Converting back to hours, minutes, and seconds, we find that Lance Armstrong’s mean winning time is 86 hours, 34 minutes, 20 seconds.

THE FAR SIDE® BY GARY LARSON



The Far Side® by Gary Larson © 1988 FarWorks, Inc. All Rights Reserved. The Far Side® and the Larson® signature are registered trademarks of FarWorks, Inc. Used with permission.

“Bob and Ruth! Come on in. ... Have you met Russell and Bill, our 1.5 children?”

If a sample of 460 families have 690 children altogether, then the mean number of children per family is $\bar{x} = 1.5$.

EXAMPLE 2

FINDING THE MEAN: GROUPED DATA In 2008, the U.S. Bureau of Labor Statistics tabulated a survey of workers’ ages and wages. The frequency distribution in Figure 4.52 summarizes the age distribution of workers who received minimum wage (\$6.55 per hour). Find the mean age of a worker receiving minimum wage.

$y = \text{Age (years)}$	Number of Workers
$16 \leq y < 20$	108,000
$20 \leq y < 25$	53,000
$25 \leq y < 35$	41,000
$35 \leq y < 45$	23,000
$45 \leq y < 55$	29,000
$55 \leq y < 65$	23,000
	$n = 277,000$

FIGURE 4.52 Age distribution of workers receiving minimum wage.
Source: Bureau of Labor Statistics, U.S. Department of Labor.

SOLUTION

To find the mean age of the workers, we must add up the ages of all the workers and divide by 277,000. However, because we are given grouped data, the ages of the individual workers are unknown to us. In this situation, we use the midpoint of each interval as the representative of the interval; consequently, our answer is an approximation.

To find the midpoint of an interval, add the endpoints and divide by 2. For instance, the midpoint of the interval $16 \leq y < 20$ is

$$\frac{16 + 20}{2} = 18$$

We can then say that each of the 108,000 people in the first interval is approximately eighteen years old. Adding up these workers' ages, we obtain

$$\begin{aligned} 18 + 18 + 18 + \cdots + 18 & \text{ (one hundred eight thousand times)} \\ &= (108,000)(18) \\ &= 1,944,000 \text{ years} \end{aligned}$$

If we let f = frequency and x = the midpoint of an interval, the product $f \cdot x$ gives us the total age of the workers in an interval. The results of this procedure for all the workers are shown in Figure 4.53.

$y = \text{Age (years)}$	$f = \text{Frequency}$	$x = \text{Midpoint}$	$f \cdot x$
$16 \leq y < 20$	108,000	18	1,944,000
$20 \leq y < 25$	53,000	22.5	1,192,500
$25 \leq y < 35$	41,000	30	1,230,000
$35 \leq y < 45$	23,000	40	920,000
$45 \leq y < 55$	29,000	50	1,450,000
$55 \leq y < 65$	23,000	60	1,380,000
	$n = 277,000$		$\Sigma(f \cdot x) = 8,116,500$

FIGURE 4.53 Using midpoints of grouped data.

Because $f \cdot x$ gives the sum of the ages of the workers in an interval, the symbol $\Sigma(f \cdot x)$ represents the sum of all the $(f \cdot x)$; that is, $\Sigma(f \cdot x)$ represents the sum of the ages of *all* workers.

The mean age is found by dividing the sum of the ages of all the workers by the number of workers:

$$\begin{aligned} \bar{x} &= \frac{\Sigma(f \cdot x)}{n} \\ &= \frac{8,116,500}{277,000} \\ &= 29.30144 \text{ years} \end{aligned}$$

The mean age of the workers earning minimum wage is approximately 29.3 years.



One common mistake that is made in working with grouped data is to forget to multiply the midpoint of an interval by the frequency of the interval. Another common mistake is to divide by the number of intervals instead of by the total number of data points.

The procedure for calculating the mean when working with grouped data (as illustrated in Example 2) is summarized in the following box.

CALCULATING THE MEAN: GROUPED DATA

Given a frequency distribution containing several groups of data, the mean \bar{x} can be found by using the following formula:

$$\bar{x} = \frac{\sum(f \cdot x)}{n} \quad \text{where } x = \text{the midpoint of a group, } f = \text{the frequency of the group, and } n = \sum f$$

EXAMPLE 3

FINDING THE MEAN: REPEATED DATA Ten college students were comparing their wages earned at part-time jobs. Nine earned \$10.00 per hour working at jobs ranging from waiting on tables to working in a bookstore. The tenth student earned \$200.00 per hour modeling for a major fashion magazine. Find the mean wage of the ten students.

SOLUTION

The data point \$10.00 is repeated nine times, so we multiply it by its frequency. Thus, the mean is as follows:

$$\begin{aligned} \bar{x} &= \frac{\sum(f \cdot x)}{n} \\ &= \frac{(9 \cdot 10) + (1 \cdot 200)}{10} \\ &= \frac{290}{10} \\ &= 29 \end{aligned}$$

The mean wage of the students is \$29.00 per hour.

Example 3 seems to indicate that the average wage of the ten students is \$29.00 per hour. Is that a reasonable figure? If nine out of ten students earn \$10.00 per hour, can we justify saying that their average wage is \$29.00? Of course not! Even though the mean wage is \$29.00, it is not a convincing “average” for this specific group of data. The mean is inflated because one student made \$200.00 per hour. This wage is called an **outlier** (or **extreme value**) because it is significantly different from the rest of the data. Whenever a collection of data has extreme values, the mean can be greatly affected and might not be an accurate measure of the average.

The Median

The **median** is the “middle value” of a distribution of numbers. To find it, we first put the data in numerical order. (If a number appears more than once, we include it as many times as it occurs.) If there is an odd number of data points, the median is the middle data point; if there is an even number of data points, the median is defined to be the mean of the two middle values. In either case, the median separates the distribution into two equal parts. Thus, the median can be viewed as an “average.” (The word *median* is also used to describe the strip that runs down the middle of a freeway; half the freeway is on one side, and half is on the other. This common usage is in keeping with the statistical meaning.)

EXAMPLE 4

FINDING THE MEDIAN: ODD VERSUS EVEN NUMBERS OF DATA

Find the median of the following sets of raw data.

- a. 2 8 3 12 6 2 11
 b. 2 8 3 12 6 2 11 8

SOLUTION

- a. First, we put the data in order from smallest to largest. Because there is an odd number of data points, ($n = 7$), we pick the middle one:

$$\begin{array}{ccccccc} 2 & 2 & 3 & 6 & 8 & 11 & 12 \\ & & & \uparrow & & & \\ & & & \text{middle value} & & & \end{array}$$

The median is 6. (Notice that 6 is the fourth number in the list.)

- b. We arrange the data first. Because there is an even number of data points ($n = 8$), we pick the two middle values and find their mean:

$$\begin{array}{cccccccc} 2 & 2 & 3 & 6 & 8 & 8 & 11 & 12 \\ & & & \uparrow & \uparrow & & & \\ & & & \frac{(6 + 8)}{2} = 7 & & & & \end{array}$$

Therefore, the median is 7. (Notice that 7 is halfway between the fourth and fifth numbers in the list.)

In Example 4, we saw that when $n = 7$, the median was the fourth number in the list, and that when $n = 8$, the median was halfway between the fourth and fifth numbers in the list. Consequently, the *location* of the median depends on n , the number of numbers in the set of data. The formula

$$L = \frac{n + 1}{2}$$

can be used to find the location, L , of the median; when $n = 7$,

$$L = \frac{7 + 1}{2} = 4$$

(the median is the fourth number in the list), and when $n = 8$,

$$L = \frac{8 + 1}{2} = 4.5$$

(the median is halfway between the fourth and fifth numbers in the list).

LOCATION OF THE MEDIAN

Given a sample of n data points, the location of the median can be found by using the following formula:

$$L = \frac{n + 1}{2}$$

Once the data have been arranged from smallest to largest, the median is the L th number in the list.

EXAMPLE 5

FINDING THE LOCATION AND VALUE OF THE MEDIAN Find the median wage for the ten students in Example 3.

SOLUTION

First, we put the ten wages in order:

10 10 10 10 10 10 10 10 10 200

Because there are $n = 10$ data points, the location of the median is

$$L = \frac{10 + 1}{2} = 5.5$$

That is, the median is halfway between the fifth and sixth numbers. To find the median, we add the fifth and sixth numbers and divide by 2. Now, the fifth number is 10 and the sixth number is 10, so the median is

$$\frac{10 + 10}{2} = 10$$

Therefore, the median wage is \$10.00. This is a much more meaningful “average” than the mean of \$29.00.

If a collection of data contains extreme values, the median, rather than the mean, is a better indicator of the “average” value. For instance, in discussions of real estate, the median is usually used to express the “average” price of a house. (Why?) In a similar manner, when the incomes of professionals are compared, the median is a more meaningful representation. The discrepancy between mean (average) income and median income is illustrated in the following news article. Although the article is dated (it’s from 1992), it is very informative as to the differences between the mean and the median.

FEATURED IN THE NEWS**DOCTORS’ AVERAGE INCOME IN U.S. IS NOW \$177,000**

Washington—The average income of the nation’s physicians rose to \$177,400 in 1992, up 4 percent from the year before, the American Medical Association said yesterday. . . .

The average income figures, compiled annually by the AMA and based on a telephone survey of more than 4,100 physicians, ranged from a low of \$111,800 for general practitioners and family practice doctors to a high of \$253,300 for radiologists.

Physicians’ median income was \$148,000 in 1992, or 6.5 percent

more than a year earlier. Half the physicians earned more than that and half earned less.

The average is pulled higher than the median by the earnings of the highest paid surgeons, anesthesiologists and other specialists at the top end of the scale. . . .

Here are the AMA’s average and median net income figures by specialty for 1992:

General/family practice:
\$111,800, \$100,000.
Internal medicine:
\$159,300, \$130,000.

Surgery: \$244,600, \$207,000.
Pediatrics: \$121,700, \$112,000.
Obstetrics/gynecology:
\$215,100, \$190,000.
Radiology: \$253,300, \$240,000.
Psychiatry: \$130,700, \$120,000.
Anesthesiology:
\$228,500, \$220,000.
Pathology: \$189,800, \$170,000.
Other: \$165,400, \$150,000.

Associated Press. Reprinted With Permission.

EXAMPLE 6

FINDING THE LOCATION AND VALUE OF THE MEDIAN The students in Ms. Kahlo’s art class were asked how many siblings they had. The frequency distribution in Figure 4.54 summarizes the responses. Find the median number of siblings.

Number of Siblings	Number of Responses
0	2
1	8
2	5
3	6

FIGURE 4.54 Frequency distribution of data.

SOLUTION

The frequency distribution indicates that two students had no (0) siblings, eight students had one (1) sibling, five students had two (2) siblings, and six students had three (3) siblings. Therefore, there were $n = 2 + 8 + 5 + 6 = 21$ students in the class; consequently, there are twenty-one data points. Listing the data in order, we have

$$0, 0, 1, \dots, 1, 2, \dots, 2, 3, \dots, 3$$

Because there are $n = 21$ data points, the location of the median is

$$L = \frac{21 + 1}{2} = 11$$

That is, the median is the eleventh number. Because the first ten numbers are 0s and 1s ($2 + 8 = 10$), the eleventh number is a 2. Consequently, the median number of siblings is 2.

The Mode

The third measure of central tendency is the **mode**. The mode is the most frequent number in a collection of data; that is, it is the data point with the highest frequency. Because it represents the most common number, the mode can be viewed as an average. A distribution of data can have more than one mode or none at all.

EXAMPLE 7

FINDING THE MODE Find the mode(s) of the following sets of raw data:

- 4 10 1 8 5 10 5 10
- 4 9 1 10 1 10 4 9
- 9 6 1 8 3 10 3 9

SOLUTION

- The mode is 10, because it has the highest frequency (3).
- There is no mode, because each number has the same frequency (2).
- The distribution has two modes—namely, 3 and 9—each with a frequency of 2. A distribution that has two modes is called *bimodal*.

In summarizing a distribution of numbers, it is most informative to list all three measures of central tendency. It helps to avoid any confusion or misunderstanding in situations in which the word *average* is used. In the hands of someone with questionable intentions, numbers can be manipulated to mislead people. In his book *How to Lie with Statistics*, Darrell Huff states, “The secret language of statistics, so appealing in a fact-minded culture, is employed to sensationalize, inflate, confuse, and oversimplify. Statistical methods and statistical terms are necessary in

reporting the mass data of social and economic trends, business conditions, ‘opinion’ polls, the census. But without writers who use the words with honesty and understanding, and readers who know what they mean, the result can only be semantic nonsense.” An educated public should always be on the alert for oversimplification of data via statistics. The next time someone mentions an “average,” ask “Which one?” Although people might not intentionally try to mislead you, their findings can be misinterpreted if you do not know the meaning of their statistics and the method by which the statistics were calculated.

4.2 EXERCISES

In Exercises 1–4, find the mean, median, and mode of the given set of raw data.

1. 9 12 8 10 9 11 12
15 20 9 14 15 21 10
- ▶ 2. 20 25 18 30 21 25 32 27
32 35 19 26 38 31 20 23
- ▶ 3. 1.2 1.8 0.7 1.5 1.0 0.7 1.9 1.7 1.2
0.8 1.7 1.3 2.3 0.9 2.0 1.7 1.5 2.2
- ▶ 4. 0.07 0.02 0.09 0.04 0.10 0.08 0.07 0.13
0.05 0.04 0.10 0.07 0.04 0.01 0.11 0.08
- ▶ 5. Find the mean, median, and mode of each set of data.
 - a. 9 9 10 11 12 15
 - b. 9 9 10 11 12 102
 - c. How do your answers for parts (a) and (b) differ (or agree)? Why?
- ▶ 6. Find the mean, median, and mode for each set of data.
 - a. 80 90 100 110 110 140
 - b. 10 90 100 110 110 210
 - c. How do your answers for parts (a) and (b) differ (or agree)? Why?
7. Find the mean, median, and mode of each set of data.
 - a. 2 4 6 8 10 12
 - b. 102 104 106 108 110 112
 - c. How are the data in part (b) related to the data in part (a)?
 - d. How do your answers for parts (a) and (b) compare?
- ▶ 8. Find the mean, median, and mode of each set of data.
 - a. 12 16 20 24 28 32
 - b. 600 800 1,000 1,200 1,400 1,600
 - c. How are the data in part (b) related to the data in part (a)?
 - d. How do your answers for parts (a) and (b) compare?
9. Kaitlin Mowry is a member of the local 4-H club and has six mini Rex rabbits that she enters in regional rabbit competition shows. The weights of the rabbits are given in Figure 4.55. Find the mean, median, and mode of the rabbits’ weights.

Weight (lb:oz)	3:12	4:03	3:06	3:15	3:12	4:02
----------------	------	------	------	------	------	------

FIGURE 4.55 Weights of rabbits.

- ▶ 10. As was stated in Example 1, Lance Armstrong won the Tour de France bicycle race every year from 1999 to 2005. His margins of victory (time difference of the second place finisher) are given in Figure 4.56. Find the mean, median, and mode of Lance Armstrong’s victory margins.

Year	1999	2000	2001	2002	2003	2004	2005
Margin (m:s)	7:37	6:02	6:44	7:17	1:01	6:19	4:40

FIGURE 4.56 Lance Armstrong’s Tour de France victory margins. Source: *San Francisco Chronicle*.

- ▶ 11. Jerry Rice holds the all-time record in professional football for scoring touchdowns. The number of touchdown receptions (TDs) for each of his seasons is given in Figure 4.57. Find the mean, median, and mode of the number of touchdown receptions per year by Rice.

Year	TDs	Year	TDs
1985	3	1995	15
1986	15	1996	8
1987	22	1997	1
1988	9	1998	9
1989	17	1999	5
1990	13	2000	7
1991	14	2001	9
1992	10	2002	7
1993	15	2003	2
1994	13	2004	3

FIGURE 4.57 Touchdown receptions for Jerry Rice. Source: <http://sportsillustrated.cnn.com/football/nfl/players/>.

▶ Selected exercises available online at www.webassign.net/brookscole

- ▶ 12. Wayne Gretzky, known as “The Great One,” holds the all-time record in professional hockey for scoring goals. The number of goals for each of his seasons is given in Figure 4.58. Find the mean, median, and mode of the number of goals per season by Gretzky.

Season	Goals	Season	Goals
1979–80	51	1989–90	40
1980–81	55	1990–91	41
1981–82	92	1991–92	31
1982–83	71	1992–93	16
1983–84	87	1993–94	38
1984–85	73	1994–95	11
1985–86	52	1995–96	23
1986–87	62	1996–97	25
1987–88	40	1997–98	23
1988–89	54	1998–99	9

FIGURE 4.58 Goals made by Wayne Gretzky.
Source: *The World Almanac*.

- ▶ 14. Michael Jordan has been recognized as an extraordinary player in professional basketball, especially in terms of the number of points per game he has scored. Jordan’s average number of points per game (PPG) for each of his seasons is given in Figure 4.60. Find the mean, median, and mode of the average points per game made per season by Jordan.

Season	PPG	Season	PPG
1984–85	28.2	1992–93	32.6
1985–86	22.7	1994–95	26.9
1986–87	37.1	1995–96	30.4
1987–88	35.0	1996–97	29.6
1988–89	32.5	1997–98	28.7
1989–90	33.6	2001–02	22.9
1990–91	31.5	2002–03	20.0
1991–92	30.1		

FIGURE 4.60 Points per game made by Michael Jordan.
Source: <http://www.nba.com/playerfile/>.

13. Barry Bonds of the San Francisco Giants set an all-time record in professional baseball by hitting 73 home runs in one season (2001). The number of home runs (HR) for each of his seasons in professional baseball is given in Figure 4.59. Find the mean, median, and mode of the number of home runs hit per year by Bonds.

Year	HR	Year	HR
1986	16	1997	40
1987	25	1998	37
1988	24	1999	34
1989	19	2000	49
1990	33	2001	73
1991	25	2002	46
1992	34	2003	45
1993	46	2004	45
1994	37	2005	5
1995	33	2006	26
1996	42	2007	28

FIGURE 4.59 Home runs hit by Barry Bonds.
Source: <http://sportsillustrated.cnn.com/baseball/mlb/players/>.

- ▶ 15. The frequency distribution in Figure 4.61 lists the results of a quiz given in Professor Gilbert’s statistics class. Find the mean, median, and mode of the scores.

Score	Number of Students
10	3
9	10
8	9
7	8
6	10
5	2

FIGURE 4.61 Quiz scores in Professor Gilbert’s statistics class.

16. Todd Booth, an avid jogger, kept detailed records of the number of miles he ran per week during the past year. The frequency distribution in Figure 4.62 summarizes his records. Find the mean, median, and mode of the number of miles per week that Todd ran.

Miles Run per Week	Number of Weeks
0	5
1	4
2	10
3	9
4	10
5	7
6	3
7	4

FIGURE 4.62 Miles run by Todd Booth.

- ▶ **17.** To study the output of a machine that fills boxes with cereal, a quality control engineer weighed 150 boxes of Brand X cereal. The frequency distribution in Figure 4.63 summarizes his findings. Find the mean weight of the boxes of cereal.

$x = \text{Weight (ounces)}$	Number of Boxes
$15.3 \leq x < 15.6$	13
$15.6 \leq x < 15.9$	24
$15.9 \leq x < 16.2$	84
$16.2 \leq x < 16.5$	19
$16.5 \leq x \leq 16.8$	10

FIGURE 4.63 Amount of Brand X cereal per box.

- ▶ **18.** To study the efficiency of its new price-scanning equipment, a local supermarket monitored the amount of time its customers had to wait in line. The frequency distribution in Figure 4.64 summarizes the findings. Find the mean amount of time spent in line.

$x = \text{Time (minutes)}$	Number of Customers
$0 \leq x < 1$	79
$1 \leq x < 2$	58
$2 \leq x < 3$	64
$3 \leq x < 4$	40
$4 \leq x \leq 5$	35

FIGURE 4.64 Time spent waiting in a supermarket checkout line.

- ▶ **19.** Katrina must take five exams in a math class. If her scores on the first four exams are 71, 69, 85, and 83, what score does she need on the fifth exam for her overall mean to be
 - a. at least 70?
 - b. at least 80?
 - c. at least 90?
- ▶ **20.** Eugene must take four exams in a geography class. If his scores on the first three exams are 91, 67, and 83, what score does he need on the fourth exam for his overall mean to be
 - a. at least 70?
 - b. at least 80?
 - c. at least 90?
- 21.** The mean salary of twelve men is \$58,000, and the mean salary of eight women is \$42,000. Find the mean salary of all twenty people.
- ▶ **22.** The mean salary of twelve men is \$52,000, and the mean salary of four women is \$84,000. Find the mean salary of all sixteen people.
- ▶ **23.** Maria drove from Chicago, Illinois, to Milwaukee, Wisconsin, a distance of ninety miles, at a mean speed of 60 miles per hour. On her return trip, the traffic was much heavier, and her mean speed was 45 miles per hour. Find Maria's mean speed for the round trip.

HINT: Divide the total distance by the total time.
- ▶ **24.** Sully drove from Atlanta, Georgia, to Birmingham, Alabama, a distance of 150 miles, at a mean speed of 50 miles per hour. On his return trip, the traffic was much lighter, and his mean speed was 60 miles per hour. Find Sully's mean speed for the round trip.

HINT: Divide the total distance by the total time.
- ▶ **25.** The mean age of a class of twenty-five students is 23.4 years. How old would a twenty-sixth student have to be for the mean age of the class to be 24.0 years?
- 26.** The mean age of a class of fifteen students is 18.2 years. How old would a sixteenth student have to be for the mean age of the class to be 21.0 years?
- 27.** The mean salary of eight employees is \$40,000, and the median is \$42,000. The highest-paid employee gets a \$6,000 raise.
 - a. What is the new mean salary of the eight employees?
 - b. What is the new median salary of the eight employees?
- ▶ **28.** The mean salary of ten employees is \$32,000, and the median is \$30,000. The highest-paid employee gets a \$5,000 raise.
 - a. What is the new mean salary of the ten employees?
 - b. What is the new median salary of the ten employees?
- 29.** The number of civilians holding government jobs in various federal departments and their monthly payrolls for September 2007 are given in Figure 4.65.

Department	Number of Civilian Workers	Monthly Payroll (thousands of dollars)
Department of Education	4,201	44,497
Department of Health and Human Services	62,502	576,747
Environmental Protection Agency	18,119	195,320
Department of Homeland Security	159,447	1,251,754
Department of the Navy	175,722	712,172
Department of the Air Force	157,182	635,901
Department of the Army	245,599	678,044

FIGURE 4.65 Monthly earnings for civilian jobs (September 2007). *Source:* U.S. Office of Personnel Management.

- **31.** The age composition of the population of the United States in the year 2000 is given in Figure 4.67.
- Find the mean age of all people in the United States under the age of 85.
 - Replace the interval “85 and over” with the interval $85 \leq x \leq 100$ and find the mean age of all people in the United States.
- **30.** The ages of the nearly 4 million women who gave birth in the United States in 2001 are given in Figure 4.66. Find the mean age of these women.

Age (years)	Number of Women
$15 \leq x < 20$	549,000
$20 \leq x < 25$	872,000
$25 \leq x < 30$	897,000
$30 \leq x < 35$	859,000
$35 \leq x < 40$	452,000
$40 \leq x < 45$	137,000

FIGURE 4.66 Ages of women giving birth in 2001. *Source:* U.S. Bureau of the Census.

Age (years)	Number of People (thousands)
$0 < x < 5$	19,176
$5 \leq x < 10$	20,550
$10 \leq x < 15$	20,528
$15 \leq x < 25$	39,184
$25 \leq x < 35$	39,892
$35 \leq x < 45$	44,149
$45 \leq x < 55$	37,678
$55 \leq x < 65$	24,275
$65 \leq x < 85$	30,752
85 and over	4,240

FIGURE 4.67 Age composition of the population of the United States in the year 2000. *Source:* U.S. Bureau of the Census.

In Exercises 32 and 33, use the age composition of the 14,980,000 students enrolled in institutions of higher education in the United States during 2000, as given in Figure 4.68.

Age of Males (years)	Number of Students
$14 \leq x < 18$	94,000
$18 \leq x < 20$	1,551,000
$20 \leq x < 22$	1,420,000
$22 \leq x < 25$	1,091,000
$25 \leq x < 30$	865,000
$30 \leq x < 35$	521,000
35 and over	997,000
total	6,539,000

FIGURE 4.68 Age composition of students in higher education. *Source:* U.S. National Center for Education Statistics.

Age of Females (years)	Number of Students
$14 \leq x < 18$	78,000
$18 \leq x < 20$	1,907,000
$20 \leq x < 22$	1,597,000
$22 \leq x < 25$	1,305,000
$25 \leq x < 30$	1,002,000
$30 \leq x < 35$	664,000
35 and over	1,888,000
total	8,441,000

FIGURE 4.68 Continued.

32. a. Find the mean age of all male students in higher education under 35.
 b. Replace the interval “35 and over” with the interval $35 \leq x \leq 60$ and find the mean age of all male students in higher education.
33. a. Find the mean age of all female students in higher education under 35.
 b. Replace the interval “35 and over” with the interval $35 \leq x \leq 60$ and find the mean age of all female students in higher education.



Answer the following questions using complete sentences and your own words.

• CONCEPT QUESTIONS

34. What are the three measures of central tendency? Briefly explain the meaning of each.

35. Suppose the mean of Group I is A and the mean of Group II is B . We combine Groups I and II to form Group III. Is the mean of Group III equal to $\frac{A+B}{2}$? Explain.
36. Why do we use the midpoint of an interval when calculating the mean of grouped data?



37. What were last month’s “average” high and low temperatures in your favorite city? Pick a city that interests you and obtain the high and low temperatures for each day last month. Print the data, and submit them as evidence in answering the following.
- a. Find the mean, median, and mode of the daily high temperatures.
 b. How does the mean daily high temperature last month compare to the mean seasonal high temperature for last month? That is, were the high temperatures last month above or below the normal high temperatures for the month?
 c. Find the mean, median, and mode of the daily low temperatures.
 d. How does the mean daily low temperature last month compare to the mean seasonal low temperature for last month? That is, were the low temperatures last month above or below the normal low temperatures for the month?

Some useful links for this web project are listed on the text web site:

www.cengage.com/math/johnson

4.3 Measures of Dispersion

OBJECTIVES

- Find the standard deviation of a data set
- Examine the dispersion of data relative to the mean

To settle an argument over who was the better bowler, George and Danny agreed to bowl six games, and whoever had the highest “average” would be considered best. Their scores were as shown in Figure 4.69.

George	185	135	200	185	250	155
Danny	182	185	188	185	180	190

FIGURE 4.69 Bowling scores.

Each bowler then arranged his scores from lowest to highest and computed the mean, median, and mode:

$$\begin{array}{l}
 \text{George} \left\{ \begin{array}{l}
 135 \quad 155 \quad 185 \quad 185 \quad 200 \quad 250 \\
 \text{mean} = \frac{\text{sum of scores}}{6} = \frac{1,110}{6} = 185 \\
 \text{median} = \text{middle score} = \frac{185 + 185}{2} = 185 \\
 \text{mode} = \text{most common score} = 185
 \end{array} \right. \\
 \\
 \text{Danny} \left\{ \begin{array}{l}
 180 \quad 182 \quad 185 \quad 185 \quad 188 \quad 190 \\
 \text{mean} = \frac{1,110}{6} = 185 \\
 \text{median} = \frac{185 + 185}{2} = 185 \\
 \text{mode} = 185
 \end{array} \right.
 \end{array}$$

Much to their surprise, George's mean, median, and mode were exactly the same as Danny's! Using the measures of central tendency alone to summarize their performances, the bowlers appear identical. Even though their averages were identical, however, their performances were not; George was very erratic, while Danny was very consistent. Who is the better bowler? On the basis of high score, George is better. On the basis of consistency, Danny is better.

George and Danny's situation points out a fundamental weakness in using only the measures of central tendency to summarize data. In addition to finding the averages of a set of data, the consistency, or spread, of the data should also be taken into account. This is accomplished by using **measures of dispersion**, which determine how the data points differ from the average.

Deviations

It is clear from George and Danny's bowling scores that it is sometimes desirable to measure the relative consistency of a set of data. Are the numbers consistently bunched up? Are they erratically spread out? To measure the dispersion of a set of data, we need to identify an average or typical distance between the data points and the mean. The difference between a single data point x and the mean \bar{x} is called the **deviation from the mean** (or simply the **deviation**) and is given by $(x - \bar{x})$. A data point that is close to the mean will have a small deviation, whereas data points far from the mean will have large deviations, as shown in Figure 4.70.

To find the typical deviation of the data points, you might be tempted to add up all the deviations and divide by the total number of data points, thus finding the "average" deviation. Unfortunately, this process leads nowhere. To see why, we will find the mean of the deviations of George's bowling scores.

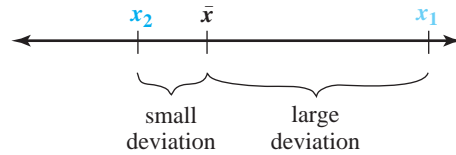


FIGURE 4.70 Large versus small deviations.

EXAMPLE 1

SOLUTION

Score (x)	Deviation (x - 185)
135	-50
155	-30
185	0
185	0
200	15
250	65
	Sum = 0

FIGURE 4.71 Deviations of George's score.

FINDING THE DEVIATIONS OF A DATA SET George bowled six games, and his scores were 185, 135, 200, 185, 250, and 155. Find the mean of the scores, the deviation of each score, and the mean of the deviations.

$$\bar{x} = \frac{\text{sum of scores}}{6} = \frac{1,110}{6} = 185$$

The mean score is 185.

To find the deviations, subtract the mean from each score, as shown in Figure 4.71.

$$\text{mean of the deviations} = \frac{\text{sum of deviations}}{6} = \frac{0}{6} = 0$$

The mean of the deviations is zero.

In Example 1, the sum of the deviations of the data is zero. This *always* happens; that is, $\sum (x - \bar{x}) = 0$ for any set of data. The negative deviations are the “culprits”—they will always cancel out the positive deviations. Therefore, to use deviations to study the spread of the data, we must modify our approach and convert the negatives into positives. We do this by squaring each deviation.

Variance and Standard Deviation

Before proceeding, we must be reminded of the difference between a population and a sample. A population is the universal set of all possible items under study; a sample is any group or subset of items selected from the population. (Samples are used to study populations.) In this context, George's six bowling scores represent a sample, not a population; because we do not know the scores of *all* the games George has ever bowled, we are limited to a sample. Unless otherwise specified, we will consider any given set of data to represent a sample, not an entire population.

To measure the typical deviation contained within a set of data points, we must first find the **variance** of the data. Given a sample of n data points, the variance of the data is found by squaring each deviation, adding the squares, and then dividing the sum by the number $(n - 1)$.*

Because we are working with n data points, you might wonder why we divide by $n - 1$ rather than by n . The answer lies in the study of inferential statistics. Recall that inferential statistics deal with the drawing of conclusions concerning the nature of a population based on observations made within a sample. Hence, the variance of a sample can be viewed as an estimate of the variance of the population. However, because the population will vary more than the sample (a population has more data points), dividing the sum of the squares of the sample deviations by n would underestimate the true variance of the entire population.

*If the n data points represent the entire population, the population variance, denoted by σ^2 , is found by squaring each deviation, adding the squares, and then dividing the sum by n .

SAMPLE VARIANCE DEVIATION

Given a sample of n data points, x_1, x_2, \dots, x_n , the **variance** of the data, denoted by s^2 , is

$$s^2 = \frac{\sum(x - \bar{x})^2}{n - 1}$$

The variance of a sample is found by dividing the sum of the squares of the deviations by $n - 1$. The symbol s^2 is a reminder that the deviations have been squared.

To compensate for this underestimation, statisticians have determined that dividing the sum of the squares of the deviations by $n - 1$ rather than by n produces the best estimate of the true population variance.

Variance is the tool with which we can obtain a measure of the typical deviation contained within a set of data. However, because the deviations have been squared, we must perform one more operation to obtain the desired result: We must take the square root. The square root of variance is called the **standard deviation** of the data.

STANDARD DEVIATION DEFINITION

Given a sample of n data points, x_1, x_2, \dots, x_n , the **standard deviation** of the data, denoted by s , is

$$s = \sqrt{\text{variance}}$$

To find the standard deviation of a set of data, first find the variance and then take the square root of the variance.

EXAMPLE 2

FINDING THE VARIANCE AND STANDARD DEVIATION: SINGLE VALUES George bowled six games, and his scores were 185, 135, 200, 185, 250, and 155. Find the standard deviation of his scores.

SOLUTION

To find the standard deviation, we must first find the variance. The mean of the six data points is 185. The necessary calculations for finding variance are shown in Figure 4.72.

Data (x)	Deviation ($x - 185$)	Deviation Squared ($x - 185$) ²
135	-50	$(-50)^2 = 2,500$
155	-30	$(-30)^2 = 900$
185	0	$(0)^2 = 0$
185	0	$(0)^2 = 0$
200	15	$(15)^2 = 225$
250	65	$(65)^2 = 4,225$
		Sum = 7,850

FIGURE 4.72 Finding variance.

$$\begin{aligned}\text{variance} &= \frac{\text{sum of the squares of the deviations}}{n - 1} \\ s^2 &= \frac{7,850}{6 - 1} \\ &= \frac{7,850}{5} \\ &= 1,570\end{aligned}$$

The variance is $s^2 = 1,570$. Taking the square root, we have

$$\begin{aligned}s &= \sqrt{1,570} \\ &= 39.62322551 \dots\end{aligned}$$

It is customary to round off s to one place more than the original data. Hence, the standard deviation of George's bowling scores is $s = 39.6$ points.

Because they give us information concerning the spread of data, variance and standard deviation are called **measures of dispersion**. Standard deviation (and variance) is a relative measure of the dispersion of a set of data; the larger the standard deviation, the more spread out the data. Consider George's standard deviation of 39.6. This appears to be high, but what exactly constitutes a "high" standard deviation? Unfortunately, because it is a relative measure, there is no hard-and-fast distinction between a "high" and a "low" standard deviation.

By itself, the standard deviation of a set of data might not be very informative, but standard deviations are very useful in comparing the relative consistencies of two sets of data. Given two groups of numbers of the same type (for example, two sets of bowling scores, two sets of heights, or two sets of prices), the set with the lower standard deviation contains data that are more consistent, whereas the data with the higher standard deviation are more spread out. Calculating the standard deviation of Danny's six bowling scores, we find $s = 3.7$. Since Danny's standard deviation is less than George's, we infer that Danny is more consistent. If George's standard deviation is less than 39.6 the next time he bowls six games, we would infer that his game has become more consistent (the scores would not be spread out as far).

Alternative Methods for Finding Variance

The procedure for calculating variance is very direct: First find the mean of the data, then find the deviation of each data point, and finally divide the sum of the squares of the deviations by $(n - 1)$. However, using the definition of sample variance to find the variance can be rather tedious. Fortunately, many scientific calculators are programmed to find the variance (and standard deviation) if you just push a few buttons. Consult your manual to utilize the statistical capabilities of your calculator.

If your calculator does not have built-in statistical functions, you might still be able to take a shortcut in calculating variance. Instead of using the definition of variance (as in Example 2), we can use an alternative formula that contains the two sums Σx and Σx^2 , where Σx represents the sum of the data and Σx^2 represents the sum of the squares of the data, as shown in the following box.

ALTERNATIVE FORMULA FOR SAMPLE VARIANCE

Given a sample of n data points, x_1, x_2, \dots, x_n , the **variance** of the data, denoted by s^2 , can be found by

$$s^2 = \frac{1}{(n-1)} \left[\sum x^2 - \frac{(\sum x)^2}{n} \right]$$

Note: $\sum x^2$ means “square each data point, then add”; $(\sum x)^2$ means “add the data points, then square.”

Although we will not prove it, this Alternative Formula for Sample Variance is algebraically equivalent to the sample variance definition; given any set of data, either method will produce the same answer. At first glance, the Alternative Formula might appear to be more difficult to use than the definition. Do not be fooled by its appearance! As we will see, the Alternative Formula is relatively quick and easy to apply.

EXAMPLE 3

FINDING THE VARIANCE AND STANDARD DEVIATION: ALTERNATIVE FORMULA Using the Alternative Formula for Sample Variance, find the standard deviation of George’s bowling scores as given in Example 2.

SOLUTION

x	x^2
135	18,225
155	24,025
185	34,225
185	34,225
200	40,000
250	62,500
$\Sigma x = 1,110$	$\Sigma x^2 = 213,200$

FIGURE 4.73

Data and data squared.

Recall that George’s scores were 185, 135, 200, 185, 250, and 155. To find the standard deviation, we must first find the variance.

The Alternative Formula for Sample Variance requires that we find the sum of the data and the sum of the squares of the data. These calculations are shown in Figure 4.73. Applying the Alternative Formula for Sample Variance, we have

$$\begin{aligned} s^2 &= \frac{1}{(n-1)} \left[\sum x^2 - \frac{(\sum x)^2}{n} \right] \\ &= \frac{1}{6-1} \left[213,200 - \frac{(1,110)^2}{6} \right] \\ &= \frac{1}{5} [213,200 - 205,350] \\ &= \frac{7,850}{5} = 1,570 \end{aligned}$$

The variance is $s^2 = 1,570$. (Note that this is the same as the variance calculated in Example 2 using the definition of sample variance.)

Taking the square root, we have

$$\begin{aligned} s &= \sqrt{1,570} \\ &= 39.62322551 \dots \end{aligned}$$

Rounded off, the standard deviation of George’s bowling scores is $s = 39.6$ points.

When we are working with grouped data, the individual data points are unknown. In such cases, the midpoint of each interval should be used as the representative value of the interval.

EXAMPLE 4

FIND THE VARIANCE AND STANDARD DEVIATION: GROUPED DATA

In 2008, the U.S. Bureau of Labor Statistics tabulated a survey of workers' ages and wages. The frequency distribution in Figure 4.74 summarizes the age distribution of workers who received minimum wage (\$6.55 per hour). Find the standard deviation of the ages these workers.

$y = \text{Age (years)}$	Number of Workers
$16 \leq y < 20$	108,000
$20 \leq y < 25$	53,000
$25 \leq y < 35$	41,000
$35 \leq y < 45$	23,000
$45 \leq y < 55$	29,000
$55 \leq y < 65$	23,000
	$n = 277,000$

FIGURE 4.74 Age distribution of workers receiving minimum wage.
Source: Bureau of Labor Statistics, U.S. Department of Labor.

SOLUTION

Because we are given grouped data, the first step is to determine the midpoint of each interval. We do this by adding the endpoints and dividing by 2.

To utilize the Alternative Formula for Sample Variance, we must find the sum of the data and the sum of the squares of the data. The sum of the data is found by multiplying each midpoint by the frequency of the interval and adding the results; that is, $\Sigma(f \cdot x)$. The sum of the squares of the data is found by squaring each midpoint, multiplying by the corresponding frequency, and adding; that is, $\Sigma(f \cdot x^2)$. The calculations are shown in Figure 4.75.

$y = \text{Age (years)}$	$f = \text{Frequency}$	$x = \text{Midpoint}$	$f \cdot x$	$f \cdot x^2$
$16 \leq y < 20$	108,000	18	1,944,000	34,992,000
$20 \leq y < 25$	53,000	22.5	1,192,500	26,831,250
$25 \leq y < 35$	41,000	30	1,230,000	36,900,000
$35 \leq y < 45$	23,000	40	920,000	36,800,000
$45 \leq y < 55$	29,000	50	1,450,000	72,500,000
$55 \leq y < 65$	23,000	60	1,380,000	82,800,000
	$n = 277,000$		$\Sigma(f \cdot x) = 8,116,500$	$\Sigma(f \cdot x^2) = 290,823,250$

FIGURE 4.75 Finding variance of grouped data.

Applying the Alternative Formula for Sample Variance, we have

$$\begin{aligned} s^2 &= \frac{1}{n-1} \left[\Sigma(f \cdot x^2) - \frac{(\Sigma f \cdot x)^2}{n} \right] \\ &= \frac{1}{277,000 - 1} \left[290,823,250 - \frac{(8,116,500)^2}{277,000} \right] \\ &= \frac{1}{276,999} [290,823,250 - 237,825,170.6] \\ &= \frac{52,998,079.4}{276,999} \end{aligned}$$

The variance is $s^2 = 191.3294972$. Taking the square root, we have

$$\begin{aligned} s &= \sqrt{191.3294972} \\ &= 13.83219062 \dots \end{aligned}$$

Rounded off, the standard deviation of the ages of the workers receiving minimum wage is $s = 13.8$ years.

The procedure for calculating variance when working with grouped data (as illustrated in Example 4) is summarized in the following box.

ALTERNATIVE FORMULA FOR SAMPLE VARIANCE: GROUPED DATA

Given a frequency distribution containing several groups of data, the variance s^2 can be found by

$$s^2 = \frac{1}{(n-1)} \left[\Sigma(f \cdot x^2) - \frac{(\Sigma f \cdot x)^2}{n} \right]$$

where x = the midpoint of a group, f = the frequency of the group, and $n = \Sigma f$

To obtain the best analysis of a collection of data, we should use the measures of central tendency and the measures of dispersion in conjunction with each other. The most common way to combine these measures is to determine what percent of the data lies within a specified number of standard deviations of the mean. The phrase “one standard deviation of the mean” refers to all numbers within the interval $[\bar{x} - s, \bar{x} + s]$, that is, all numbers that differ from \bar{x} by at most s . Likewise, “two standard deviations of the mean” refers to all numbers within the interval $[\bar{x} - 2s, \bar{x} + 2s]$. One, two, and three standard deviations of the mean are shown in Figure 4.76.

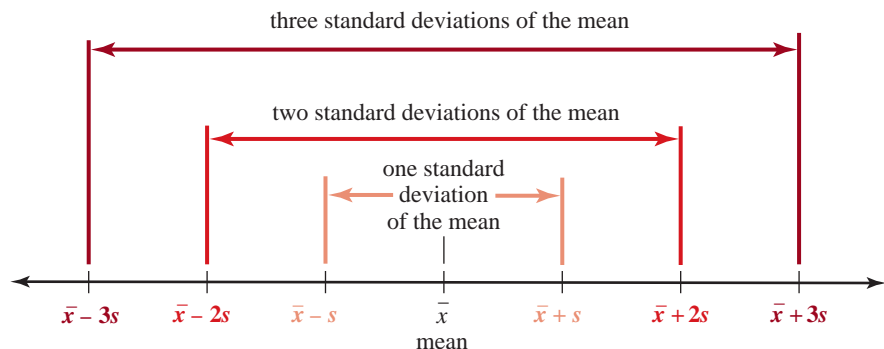


FIGURE 4.76 One, two, and three standard deviations of the mean.

EXAMPLE 5

FINDING THE PERCENTAGE OF DATA WITHIN A SPECIFIED INTERVAL Paki Mowry is a rabbit enthusiast and has eleven mini Rex rabbits that she enters in regional rabbit competition shows. The weights of the rabbits are given in Figure 4.77. What percent of the rabbits' weights lie within one standard deviation of the mean?

Weight (lb:oz)	3:10	4:02	3:06	3:15	3:12	4:01
	3:11	4:00	4:03	3:13	3:15	

FIGURE 4.77 Weights of rabbits in pounds and ounces.

SOLUTION

First, the weights must be converted to a common unit, say, ounces. Therefore, we multiply the number of pounds by 16 (ounces per pound) and add the given number of ounces. For instance, 3 pounds, 10 ounces is converted to ounces as follows:

$$\begin{aligned} 3 \text{ pounds, } 10 \text{ ounces} &= \left(3 \text{ pounds} \times \frac{16 \text{ ounces}}{\text{pound}} \right) + 10 \text{ ounces} \\ &= 48 \text{ ounces} + 10 \text{ ounces} \\ &= 58 \text{ ounces} \end{aligned}$$

The converted weights are given in Figure 4.78.

Weight (ounces)	58	66	54	63	60	65
	59	64	67	61	63	

FIGURE 4.78 Weights of rabbits in ounces.

Now we find the mean and standard deviation.

Summing the eleven data points, we have $\Sigma x = 680$. Summing the squares of the data points, we have $\Sigma x^2 = 42,186$. The mean is

$$\bar{x} = \frac{680}{11} = 61.81818181 \dots = 61.8 \text{ ounces (rounded to one decimal place)}$$

Using the Alternative Formula for Sample Variance, we have

$$\begin{aligned} s^2 &= \frac{1}{(n-1)} \left[\Sigma x^2 - \frac{(\Sigma x)^2}{n} \right] \\ &= \frac{1}{(11-1)} \left[42,186 - \frac{(680)^2}{11} \right] \\ &= \frac{1}{10} \left[\frac{1646}{11} \right] && \text{subtracting fractions with LCD = 11} \\ &= \frac{823}{55} && \text{reducing} \end{aligned}$$

The variance is $\frac{823}{55} = 14.96363636 \dots$. Taking the square root, we have

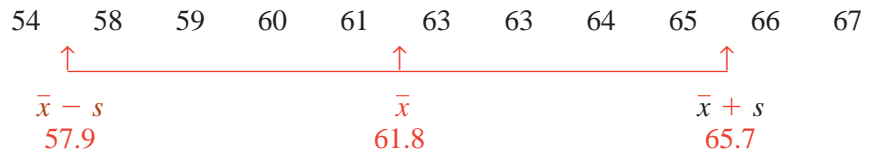
$$s = \sqrt{823/55} = 3.868285972 \dots$$

The standard deviation is 3.9 ounces (rounded to one decimal place).

To find one standard deviation of the mean, we add and subtract the standard deviation to and from the mean:

$$\begin{aligned} [\bar{x} - s, \bar{x} + s] &= [61.8 - 3.9, 61.8 + 3.9] \\ &= [57.9, 65.7] \end{aligned}$$

Arranging the data from smallest to largest, we see that eight of the eleven data points are between 57.9 and 65.7:



Therefore, $\frac{8}{11} = 0.7272727272 \dots$, or 72.7%, of the data lie within one standard deviation of the mean.

4.3 EXERCISES

- ▶ 1. Perform each task, given the following sample data:
3 8 5 3 10 13
- Use the Sample Variance Definition to find the variance and standard deviation of the data.
 - Use the Alternative Formula for Sample Variance to find the variance and standard deviation of the data.
- ▶ 2. Perform each task, given the following sample data:
6 10 12 12 11 17 9
- Use the Sample Variance Definition to find the variance and standard deviation of the data.
 - Use the Alternative Formula for Sample Variance to find the variance and standard deviation of the data.
- ▶ 3. Perform each task, given the following sample data:
10 10 10 10 10 10
- Find the variance of the data.
 - Find the standard deviation of the data.
- ▶ 4. Find the mean and standard deviation of each set of data.
- 2 4 6 8 10 12
 - 102 104 106 108 110 112
 - How are the data in (b) related to the data in (a)?
 - How do your answers for (a) and (b) compare?
- ▶ 5. Find the mean and standard deviation of each set of data.
- 12 16 20 24 28 32
 - 600 800 1,000 1,200 1,400 1,600
 - How are the data in (b) related to the data in (a)?
 - How do your answers for (a) and (b) compare?
- ▶ 6. Find the mean and standard deviation of each set of data.
- 50 50 50 50 50
 - 46 50 50 50 54
 - 5 50 50 50 95
 - How do your answers for (a), (b), and (c) compare?
- ▶ 7. Joey and Dee Dee bowled five games at the Rock 'n' Bowl Lanes. Their scores are given in Figure 4.79.
- Find the mean score of each bowler. Who has the highest mean?
 - Find the standard deviation of each bowler's scores.
 - Who is the more consistent bowler? Why?

Joey	144	171	220	158	147
Dee Dee	182	165	187	142	159

FIGURE 4.79 Bowling scores.

- ▶ 8. Paki surveyed the price of unleaded gasoline (self-serve) at gas stations in Novato and Lafayette. The raw data, in dollars per gallon, are given in Figure 4.80.
- Find the mean price in each city. Which city has the lowest mean?
 - Find the standard deviation of prices in each city.
 - Which city has more consistently priced gasoline? Why?

Novato	2.899	3.089	3.429	2.959	2.999	3.099
Lafayette	3.595	3.389	3.199	3.199	3.549	3.349

FIGURE 4.80 Price (in dollars) of one gallon of unleaded gasoline.

▶ Selected exercises available online at www.webassign.net/brookscole

- 9. Kaitlin Mowry is a member of the local 4-H club and has six mini Rex rabbits that she enters in regional rabbit competition shows. The weights of the rabbits are given in Figure 4.81. Find the standard deviation of the rabbits' weights.

Weight (lb:oz)	4:12	5:03	4:06	3:15	4:12	4:08
----------------	------	------	------	------	------	------

FIGURE 4.81 Weights of rabbits.

- 10. As was stated in Example 1 of Section 4.2, Lance Armstrong won the Tour de France bicycle race every year from 1999 to 2005. His margins of victory (time difference of the second place finisher) are given in Figure 4.82. Find the standard deviation of Lance Armstrong's victory margins.

Year	1999	2000	2001	2002	2003	2004	2005
Margin	7:37	6:02	6:44	7:17	1:01	6:19	4:40

FIGURE 4.82 Lance Armstrong's Tour de France victory margins. *Source: San Francisco Chronicle.*

11. Barry Bonds of the San Francisco Giants set an all-time record in professional baseball by hitting 73 home runs in one season (2001). The number of home runs (HR) for each of his seasons in professional baseball is given in Figure 4.83. Find the standard deviation of the number of home runs hit per year by Bonds.

Year	HR	Year	HR
1986	16	1997	40
1987	25	1998	37
1988	24	1999	34
1989	19	2000	49
1990	33	2001	73
1991	25	2002	46
1992	34	2003	45
1993	46	2004	45
1994	37	2005	5
1995	33	2006	26
1996	42	2007	28

FIGURE 4.83 Barry Bonds's home runs.

- 12. Michael Jordan has been recognized as an extraordinary player in professional basketball, especially in terms of the number of points per game he has scored. Jordan's average number of points per game (PPG) for each of his seasons is given in Figure 4.84. Find the standard deviation of the average points per game made per season by Jordan.

Season	PPG	Season	PPG
1984–85	28.2	1992–93	32.6
1985–86	22.7	1994–95	26.9
1986–87	37.1	1995–96	30.4
1987–88	35.0	1996–97	29.6
1988–89	32.5	1997–98	28.7
1989–90	33.6	2001–02	22.9
1990–91	31.5	2002–03	20.0
1991–92	30.1		

FIGURE 4.84 Michael Jordan's points per game.

- 13. The Truly Amazing Dudes are a group of comic acrobats. The heights (in inches) of the ten acrobats are as follows:

68 50 70 67 72 78 69 68 66 67



© Michael P. Gadowski/Photo Researchers, Inc.

Is your height or weight "average"? These characteristics can vary considerably within any specific group of people. The mean is used to represent the average, and the standard deviation is used to measure the "spread" of a collection of data.

- a. Find the mean and standard deviation of the heights.
- b. What percent of the data lies within one standard deviation of the mean?
- c. What percent of the data lies within two standard deviations of the mean?
- 14. The weights (in pounds) of the ten Truly Amazing Dudes are as follows:
- 152 196 144 139 166 83 186 157 140 138
- a. Find the mean and standard deviation of the weights.
- b. What percent of the data lies within one standard deviation of the mean?
- c. What percent of the data lies within two standard deviations of the mean?
15. The normal monthly rainfall in Seattle, Washington, is given in Figure 4.85.

Month	Jan.	Feb.	Mar.	Apr.	May	June
Inches	5.4	4.0	3.8	2.5	1.8	1.6

Month	July	Aug.	Sept.	Oct.	Nov.	Dec.
Inches	0.9	1.2	1.9	3.3	5.7	6.0

FIGURE 4.85 Monthly rainfall in Seattle, WA. *Source:* U.S. Department of Commerce.

- a. Find the mean and standard deviation of the monthly rainfall in Seattle.
- b. What percent of the year will the monthly rainfall be within one standard deviation of the mean?
- c. What percent of the year will the monthly rainfall be within two standard deviations of the mean?
- 16. The normal monthly rainfall in Phoenix, Arizona, is given in Figure 4.86.

Month	Jan.	Feb.	Mar.	Apr.	May	June
Inches	0.7	0.7	0.9	0.2	0.1	0.1

Month	July	Aug.	Sept.	Oct.	Nov.	Dec.
Inches	0.8	1.0	0.9	0.7	0.7	1.0

FIGURE 4.86 Monthly rainfall in Phoenix, AZ. *Source:* U.S. Department of Commerce.

- a. Find the mean and standard deviation of the monthly rainfall in Phoenix.
- b. What percent of the year will the monthly rainfall be within one standard deviation of the mean?
- c. What percent of the year will the monthly rainfall be within two standard deviations of the mean?

17. The frequency distribution in Figure 4.87 lists the results of a quiz given in Professor Gilbert's statistics class.

Score	Number of Students	Score	Number of Students
10	5	7	8
9	10	6	3
8	6	5	2

FIGURE 4.87 Quiz scores in Professor Gilbert's statistics class.

- a. Find the mean and standard deviation of the scores.
- b. What percent of the data lies within one standard deviation of the mean?
- c. What percent of the data lies within two standard deviations of the mean?
- d. What percent of the data lies within three standard deviations of the mean?
- 18. Amy surveyed the prices for a quart of a certain brand of motor oil. The sample data, in dollars per quart, is summarized in Figure 4.88.

Price per Quart (dollars)	Number of Stores
1.99	2
2.09	5
2.19	10
2.29	13
2.39	9
2.49	3

FIGURE 4.88 Price for a quart of motor oil.

- a. Find the mean and the standard deviation of the prices.
- b. What percent of the data lies within one standard deviation of the mean?
- c. What percent of the data lies within two standard deviations of the mean?
- d. What percent of the data lies within three standard deviations of the mean?
- 19. To study the output of a machine that fills boxes with cereal, a quality control engineer weighed 150 boxes of Brand X cereal. The frequency distribution in Figure 4.89 summarizes his findings. Find the standard deviation of the weight of the boxes of cereal.

$x = \text{Weight (ounces)}$	Number of Boxes
$15.3 \leq x < 15.6$	13
$15.6 \leq x < 15.9$	24
$15.9 \leq x < 16.2$	84
$16.2 \leq x < 16.5$	19
$16.5 \leq x < 16.8$	10

FIGURE 4.89 Amount of Brand X cereal per box.

- 20. To study the efficiency of its new price-scanning equipment, a local supermarket monitored the amount of time its customers had to wait in line. The frequency distribution in Figure 4.90 summarizes the findings. Find the standard deviation of the amount of time spent in line.

$x = \text{Time (minutes)}$	Number of Customers
$0 \leq x < 1$	79
$1 \leq x < 2$	58
$2 \leq x < 3$	64
$3 \leq x < 4$	40
$4 \leq x \leq 5$	35

FIGURE 4.90 Time spent waiting in a supermarket checkout line.

21. The ages of the nearly 4 million women who gave birth in the United States in 2001 are given in Figure 4.91. Find the standard deviation of the ages of these women.

Age (years)	Number of Women
$15 \leq x < 20$	549,000
$20 \leq x < 25$	872,000
$25 \leq x < 30$	897,000
$30 \leq x < 35$	859,000
$35 \leq x < 40$	452,000
$40 \leq x < 45$	137,000

FIGURE 4.91 Ages of women giving birth in 2001. *Source:* U.S. Bureau of the Census.

- 22. The age composition of the population of the United States in the year 2000 is given in Figure 4.92. Replace the interval “85 and over” with the interval

$85 \leq x \leq 100$ and find the standard deviation of the ages of all people in the United States.

Age	Number of People (thousands)
$0 < x < 5$	19,176
$5 \leq x < 10$	20,550
$10 \leq x < 15$	20,528
$15 \leq x < 25$	39,184
$25 \leq x < 35$	39,892
$35 \leq x < 45$	44,149
$45 \leq x < 55$	37,678
$55 \leq x < 65$	24,275
$65 \leq x < 85$	30,752
85 and over	4,240

FIGURE 4.92 Age composition of the population of the United States in the year 2000. *Source:* U.S. Bureau of the Census.



Answer the following questions using complete sentences and your own words.

• **CONCEPT QUESTIONS**

23. a. When studying the dispersion of a set of data, why are the deviations from the mean squared?
 b. What effect does squaring have on a deviation that is less than 1?
 c. What effect does squaring have on a deviation that is greater than 1?
 d. What effect does squaring have on the data’s units?
 e. Why is it necessary to take a square root when calculating standard deviation?
24. Why do we use the midpoint of an interval when calculating the standard deviation of grouped data?

WEB PROJECT

25. This project is a continuation of Exercise 37 in Section 4.2. How did last month’s daily high and low temperatures vary in your favorite city? Pick a city that interests you, and obtain the high and low temperatures for each day last month. Print the data, and submit them as evidence in answering the following.
- a. Find the standard deviation of the daily high temperatures.

- b. Find the standard deviation of the daily low temperatures.
- c. Comparing your answers to parts (a) and (b), what can you conclude?

Some useful links for this web project are listed on the text web site:

www.cengage.com/math/johnson

• PROJECTS

26. The purpose of this project is to explore the variation in the pricing of a common commodity. Go to several different stores that sell food (the more stores the better) and record the price of one gallon of whole milk.
- Compute the mean, median, and mode of the data.
 - Compute the standard deviation of the data.
27. The purpose of this project is to explore the variation in the pricing of a common commodity. Go to several different gas stations (the more the better) and record the price of one gallon of premium gasoline (91 octane).
- Compute the mean, median, and mode of the data.
 - Compute the standard deviation of the data.
 - What percent of the data lie within one standard deviation of the mean?
 - What percent of the data lie within two standard deviations of the mean?
 - What percent of the data lie within three standard deviations of the mean?



MEASURES OF CENTRAL TENDENCY AND DISPERSION ON A GRAPHING CALCULATOR

In Examples 1, 2, and 3 of this section, we found the mean, median, mode, variance, and standard deviation of George's bowling scores. This work can be done quickly and easily on either a graphing calculator or Excel.



TECHNOLOGY AND MEASURES OF CENTRAL TENDENCY AND DISPERSION

Calculating the Mean, the Variance, and the Standard Deviation

ON A TI-83/84:

- Enter the data from Example 1 of this section in list L_1 as discussed in Section 4.1.
- Press **STAT**.
- Select "1-Var Stats" from the "CALC" menu.
- When "1-Var" appears on the screen, press **2nd** **L1** **ENTER**.

ON A CASIO:

- Enter the data from Example 1 of this section as described in Section 4.1.
- Press **CALC** (i.e., **F2**).
- Press **1VAR** (i.e., **F1**).

```

1-Var Stats
 $\bar{x}$ =185
 $\Sigma x$ =1110
 $\Sigma x^2$ =213200
 $S_x$ =39.62322551
 $\sigma_x$ =36.17089069
 $\downarrow$ n=6

```

```

1-Var Stats
 $\uparrow$ n=6
minX=135
Q1=155
Med=185
Q3=200
maxX=250

```

FIGURE 4.93

After computing the mean, standard deviation, and more on a TI graphing calculator.

The above steps will result in the first screen in Figure 4.93. This screen gives the mean, the sample standard deviation (S_x on a TI, $\chi\sigma n-1$ on a Casio), the population standard deviation (σ_x on a TI, $\chi\sigma n$ on a Casio), and the number of data points (n). The second screen can be obtained by pressing the down arrow. It gives the minimum and maximum data points (minX and maxX, respectively) as well as the median (Med).

Calculating the Sample Variance

The above work does not yield the sample variance. To find it, follow these steps:

- Quit the statistics mode.
- Get S_x on the screen.
- Press **VAR**, select “Statistics,” and then select “ S_x ” from the “X/Y” menu.
- Once S_x is on the screen, square it by pressing **x²** **ENTER**. The variance is 1,570.

Calculating the Mean, the Variance, and the Standard Deviation with Grouped Data

To calculate the mean and standard deviation from the frequency distribution that utilizes grouped data, follow the same steps except:

ON A TI-83/84:

- Enter the midpoints of the classes in list L_1 .
- Enter the frequencies of those classes in list L_2 . Each frequency must be less than 100.
- After “1-Var Stats” appears on the screen, press **2nd** **L₁** **,** **2nd** **L₂** **ENTER**.



MEASURES OF CENTRAL TENDENCY AND DISPERSION ON EXCEL

1. Enter the data from Example 1 of this section as discussed in Section 4.1. See column A of Figure 4.94.
2. Have Excel compute the mean, standard deviation, etc.
 - Use your mouse to select “Tools” at the very top of the screen. Then pull your mouse down until “Data Analysis” is highlighted, and let go. (If “Data Analysis” is not listed under “Tools,” then follow the instructions given in Section 4.1.)
 - In the “Data Analysis” box that appears, use your mouse to highlight “Descriptive Statistics” and press the “OK” button.
 - In the “Descriptive Statistics” box that appears, use your mouse to click on the white rectangle that follows “Input Range,” and then use your mouse to draw a box around all of the bowling scores. (To draw the box, move your mouse to cell A2, press the mouse button, and move the mouse down until all of the entered scores are enclosed in a box.) This should cause “\$A\$2:\$A\$7” to appear in the Input Range rectangle.

◇	A	B	C
1	George's scores	Column 1	
2	185		
3	135	Mean	185
4	200	Standard Errc	16.1761141
5	185	Median	185
6	250	Mode	185
7	155	Standard Dev	39.6232255
8		Sample Varia	1570
9		Kurtosis	0.838675
10		Skewness	0.60763436
11		Range	115
12		Minimum	135
13		Maximum	250
14		Sum	1110
15		Count	6

FIGURE 4.94 After computing the mean, standard deviation, and more on Excel.

- Be certain that there is a dot in the button to the left of “Output Range.” Then click on the white rectangle that follows “Output Range,” and use your mouse to click on cell B1. This should cause “\$B\$1” to appear in the Output Range rectangle.
- Select “Summary Statistics.”
- Use your mouse to press the “OK” button.

After you complete this step, your spreadsheet should look like that in Figure 4.94. To compute the population standard deviation, first type “pop std dev” in cell B16, and then click on cell C16. Press the *fx* button on the Excel ribbon at the top of the screen. In the “Paste Function” box that appears, click on “Statistical” under “Paste Function category” and on “STDEV.P” under “Function name.” Then press the “OK” button. In the “STDEV.P” box that appears, click on the white rectangle that follows “Number 1.” Then use your mouse to draw a box around all of the bowling scores. Press the “OK” button and the population standard deviation will appear in cell C16.

EXERCISES

28. Do Double Stuf Oreos really contain twice as much “stuf”? In 1999, Marie Revak and Jihan William’s “interest was piqued by the unusual spelling of the word stuf and the bold statement ‘twice the filling’ on the package.” So they conducted an experiment in which they weighed the amount of filling in a sample of traditional Oreos and Double Stuf Oreos. The data from that experiment are in Figure 4.95 (weight in grams). An unformatted spreadsheet containing these data can be downloaded from the text web site: www.cengage.com/math/johnson
- Find the mean of these data.
 - Find the standard deviation.



Fostling Research

- Did you choose the population or sample standard deviation? Why?
- What do your results say about Double Stuf Oreos? Interpret the results of both parts (a) and (b).

Double Stuf Oreos:																		
4.7	6.5	5.5	5.6	5.1	5.3	5.4	5.4	3.5	5.5	6.5	5.9	5.4	4.9	5.6	5.7	5.3	6.9	6.5
6.3	4.8	3.3	6.4	5.0	5.3	5.5	5.0	6.0	5.7	6.3	6.0	6.3	6.1	6.0	5.8	5.8	5.9	6.2
5.9	6.5	6.5	6.1	5.8	6.0	6.2	6.2	6.0	6.8	6.2	5.4	6.6	6.2					
Traditional Oreos:																		
2.9	2.8	2.6	3.5	3.0	2.4	2.7	2.4	2.5	2.2	2.6	2.6	2.9	2.6	2.6	3.1	2.9	2.4	2.8
3.8	3.1	2.9	3.0	2.1	3.8	3.0	3.0	2.8	2.9	2.7	3.2	2.8	3.1	2.7	2.8	2.6	2.6	3.0
2.8	3.5	3.3	3.3	2.8	3.1	2.6	3.5	3.5	3.1	3.1								

FIGURE 4.95 Amount of Oreo filling. *Source:* Revak, Marie A. and Jihan G. Williams, “Sharing Teaching Ideas: The Double Stuf Dilemma,” *The Mathematics Teacher*, November 1999, Volume 92, Issue 8, page 674.

Rank	City	Air Quality Index (% days AQI ranked good)	Property Crime (incidents per 1000)	Median Commute Time (in minutes)
1	Louisville, CO	74.0%	15	19.5
2	Chanhausen, MN	N.A.	16	23.1
3	Papillion, NE	92.0%	16	19.3
4	Middleton, WI	76.0%	27	16.3
5	Milton, MA	81.0%	9	27.7
6	Warren, NJ	N.A.	21	28.7
7	Keller, TX	65.0%	14	29.4
8	Peachtree City, GA	84.0%	12	29.3
9	Lake St. Louis, MO	79.0%	16	26.8
10	Mukilteo, WA	89.0%	27	22.3

FIGURE 4.96 The best places to live. *Source:* CNNMoney.com.

29. Every year, CNNMoney.com lists “the best places to live.” They compare the “best small towns” on several bases, one of which is the “quality of life.” See Figure 4.96. An unformatted spreadsheet containing these data can be downloaded from the text web site: www.cengage.com/math/johnson
- Find the mean of the air quality indices.
 - Would the population standard deviation or the sample standard deviation be more appropriate here? Why?
 - Find the standard deviation from part (b).
 - What do your results say about the air quality index of all of the best places to live in the United States?
30. Every year, CNNMoney.com lists “the best places to live.” They compare the “best small towns” on several bases, one of which is the “quality of life.” See Figure 4.96. An unformatted spreadsheet containing these data can be downloaded from the text web site: www.cengage.com/math/johnson
- Find the mean of the commute times.
 - Would the population standard deviation or the sample standard deviation be more appropriate here? Why or why not?
 - Find the standard deviation from part (b).
 - What do your results say about commuting in all of the best places to live in the United States?
31. Every year, CNNMoney.com lists “the best places to live.” They compare the “best small towns” on several bases, one of which is the “quality of life.” See Figure 4.96. An unformatted spreadsheet containing these data can be downloaded from the text web site: www.cengage.com/math/johnson
- Find the mean of the property crime levels.
 - Would the population standard deviation or the sample standard deviation be more appropriate here? Why or why not?
 - Find the standard deviation from part (b).
 - What do your results say about property crime in all of the best places to live in the United States?
32. Figure 4.97 gives EPA fuel efficiency ratings for 2010 family sedans with an automatic transmission and the smallest available engine. An unformatted spreadsheet containing these data can be downloaded from the text web site: www.cengage.com/math/johnson
- For each of the two manufacturing regions, find the mean and an appropriate standard deviation for city driving, and defend your choice of standard deviations.
 - What do the results of part (a) tell you about the two regions?
33. Figure 4.97 gives EPA fuel efficiency ratings for 2010 family sedans with an automatic transmission and the smallest available engine. An unformatted spreadsheet containing these data can be downloaded from the text web site: www.cengage.com/math/johnson
- For each of the two manufacturing regions, find the mean and an appropriate standard deviation for highway driving, and defend your choice of standard deviations.
 - What do the results of part (a) tell you about the two regions?

American Cars	City mpg	Highway mpg	Asian Cars	City mpg	Highway mpg
Ford Fusion hybrid	41	36	Toyota Prius	51	48
Chevrolet Malibu hybrid	26	34	Nissan Altima hybrid	35	33
Ford Fusion fwd	24	34	Toyota Camry Hybrid	33	34
Mercury Milan	23	34	Kia Forte	27	36
Chevrolet Malibu	22	33	Hyundai Elantra	26	35
Saturn Aura	22	33	Nissan Altima	23	32
Dodge Avenger	21	30	Subaru Legacy awd	23	31
Buick Lacrosse	17	27	Toyota Camry	22	33
			Hyundai Sonata	22	32
			Kia Optima	22	32
			Honda Accord	22	31
			Mazda 6	21	30
			Mitsubishi Galant	21	30
			Hyunda Azera	18	26

FIGURE 4.97 Fuel efficiency. *Source:* www.fueleconomy.gov.

4.4 The Normal Distribution

OBJECTIVES

- Find probabilities of the standard normal distribution
- Find probabilities of a nonstandard normal distribution
- Find the value of a normally distributed variable that will produce a specific probability

Sets of data may exhibit various trends or patterns. Figure 4.98 shows a histogram of the weights of bags of corn chips. Notice that most of the data are near the “center” and that the data taper off at either end. Furthermore, the histogram is nearly symmetric; it is almost the same on both sides. This type of distribution (nearly symmetric, with most of the data in the middle) occurs quite often in many different situations. To study the composition of such distributions, statisticians have created an ideal **bell-shaped curve** describing a **normal distribution**, as shown in Figure 4.99.

Before we can study the characteristics and applications of a normal distribution, we must make a distinction between different types of variables.

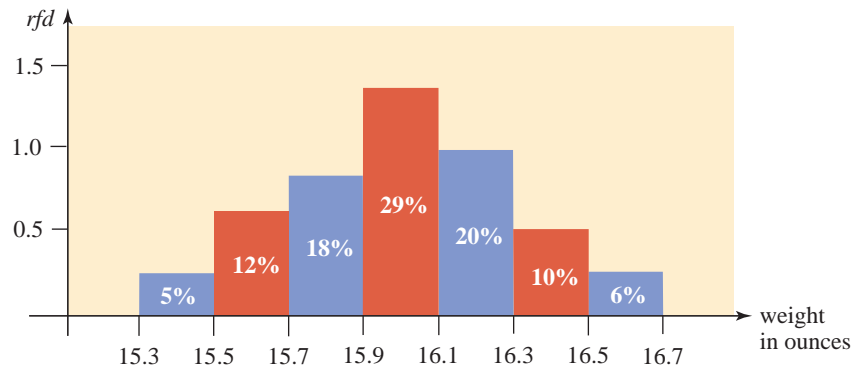


FIGURE 4.98 Weights of bags of corn chips.

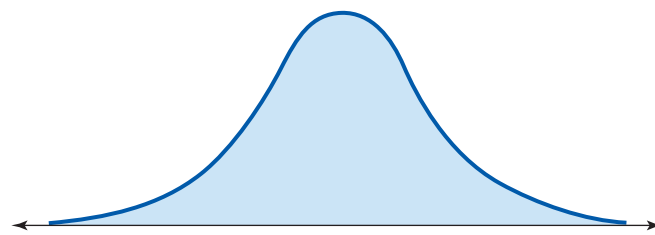


FIGURE 4.99 Normal distribution.

Discrete versus Continuous Variables

The number of children in a family is variable, because it varies from family to family. In listing the number of children, only whole numbers (0, 1, 2, and so on) can be used. In this respect, we are limited to a collection of discrete, or separate, values. A variable is **discrete** if there are “gaps” between the possible variable values. Consequently, any variable that involves counting is discrete.

On the other hand, a person’s height or weight does not have such a restriction. When someone grows, he or she does not instantly go from 67 inches to 68 inches; a person grows continuously from 67 inches to 68 inches, attaining all possible values in between. For this reason, height is called a continuous variable. A variable is **continuous** if it can assume *any value* in an interval of real numbers (see Figure 4.100). Consequently, any variable that involves measurement is continuous; someone might claim to be 67 inches tall and to weigh 152 pounds, but the true values might be 67.13157 inches and 151.87352 pounds. Heights and weights are expressed (discretely) as whole numbers solely for convenience; most people do not have rulers or bathroom scales that allow them to obtain measurements that are accurate to ten or more decimal places!

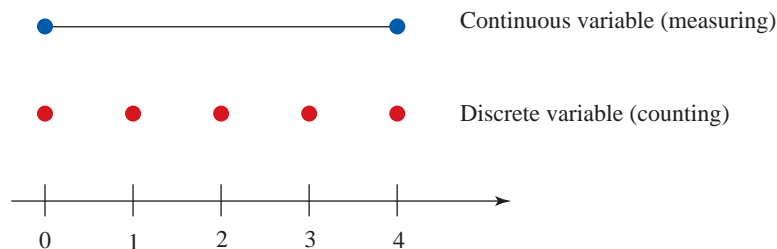


FIGURE 4.100 Discrete versus continuous variables.

Normal Distributions

The collection of all possible values that a discrete variable can assume forms a countable set. For instance, we can list all the possible numbers of children in a family. In contrast, a continuous variable will have an uncountable number of possibilities because it can assume any value in an interval. For instance, the weights (a continuous variable) of bags of corn chips could be *any* value x such that $15.3 \leq x \leq 16.7$.

When we sample a continuous variable, some values may occur more often than others. As we can see in Figure 4.98, the weights are “clustered” near the center of the histogram, with relatively few located at either end. If a continuous variable has a symmetric distribution such that the highest concentration of values is at the center and the lowest is at both extremes, the variable is said to have a **normal distribution** and is represented by a smooth, continuous, bell-shaped curve like that in Figure 4.99*.

The normal distribution, which is found in a wide variety of situations, has two main qualities: (1) the frequencies of the data points nearer the center or “average” are increasingly higher than the frequencies of data points far from the center, and (2) the distribution is symmetric (one side is a mirror image of the other). *Because of these two qualities, the mean, median, and mode of a normal distribution all coincide at the center of the distribution.*

Just like any other collection of numbers, the spread of normal distribution is measured by its standard deviation. It can be shown that for any normal distribution, slightly more than two-thirds of the data (68.26%) will lie within one standard deviation of the mean, 95.44% will lie within two standard deviations, and virtually all the data (99.74%) will lie within three standard deviations of the mean. Recall that μ (the Greek letter “mu”) represents the mean of a population and σ (the Greek letter “sigma”) represents the standard deviation of the population. The spread of a normal distribution, with μ and σ used to represent the mean and standard deviation, is shown in Figure 4.101.

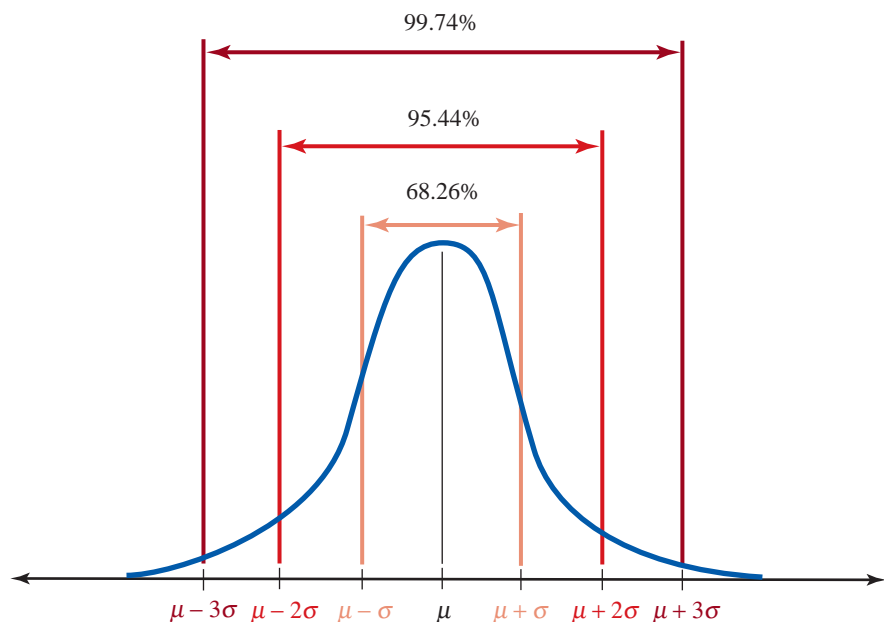


FIGURE 4.101 The spread of a normal distribution.

* This is an informal definition only. The formal definition of a normal distribution involves the number pi, the natural exponential e^x , and the mean and variance of the distribution.

EXAMPLE 1

ANALYZING THE DISPERSION OF A NORMAL DISTRIBUTION The heights of a large group of people are assumed to be normally distributed. Their mean height is 66.5 inches, and the standard deviation is 2.4 inches. Find and interpret the intervals representing one, two, and three standard deviations of the mean.

SOLUTION

The mean is $\mu = 66.5$, and the standard deviation is $\sigma = 2.4$.

1. *One standard deviation of the mean:*

$$\begin{aligned}\mu \pm 1\sigma &= 66.5 \pm 1(2.4) \\ &= 66.5 \pm 2.4 \\ &= [64.1, 68.9]\end{aligned}$$

Therefore, approximately 68% of the people are between 64.1 and 68.9 inches tall.

2. *Two standard deviations of the mean:*

$$\begin{aligned}\mu \pm 2\sigma &= 66.5 \pm 2(2.4) \\ &= 66.5 \pm 4.8 \\ &= [61.7, 71.3]\end{aligned}$$

Therefore, approximately 95% of the people are between 61.7 and 71.3 inches tall.

3. *Three standard deviations of the mean:*

$$\begin{aligned}\mu \pm 3\sigma &= 66.5 \pm 3(2.4) \\ &= 66.5 \pm 7.2 \\ &= [59.3, 73.7]\end{aligned}$$

Nearly all of the people (99.74%) are between 59.3 and 73.7 inches tall.

In Example 1, we found that virtually all the people under study were between 59.3 and 73.7 inches tall. A clothing manufacturer might want to know what percent of these people are shorter than 66 inches or what percent are taller than 73 inches. Questions like these can be answered by using probability and a normal distribution. (We will do this in Example 6.)

Probability, Area, and Normal Distributions

In Chapter 3, we mentioned that relative frequency is really a type of probability. If 3 out of every 100 people have red hair, you could say that the relative frequency of red hair is $\frac{3}{100}$ (or 3%), or you could say that the probability of red hair $p(x = \text{red hair})$ is 0.03. Therefore, to find out what percent of the people in a population are taller than 73 inches, we need to find $p(x > 73)$, the probability that x is greater than 73, where x represents the height of a randomly selected person.

Recall that a sample space is the set S of all possible outcomes of a random experiment. Consequently, the probability of a sample space must always equal 1; that is, $p(S) = 1$ (or 100%). If the sample space S has a normal distribution, its outcomes and their respective probabilities can be represented by a bell curve.

Recall that when constructing a histogram, relative frequency density (*rfd*) was used to measure the heights of the rectangles. Consequently, the *area* of a rectangle gave the relative frequency (percent) of data contained in an interval. In a similar manner, we can imagine a bell curve being a histogram composed of infinitely many “skinny” rectangles, as in Figure 4.102.

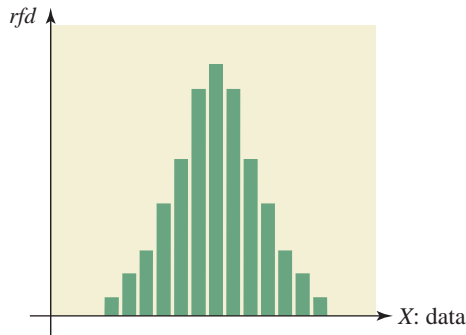


FIGURE 4.102 Symmetric, bell-shaped histogram.

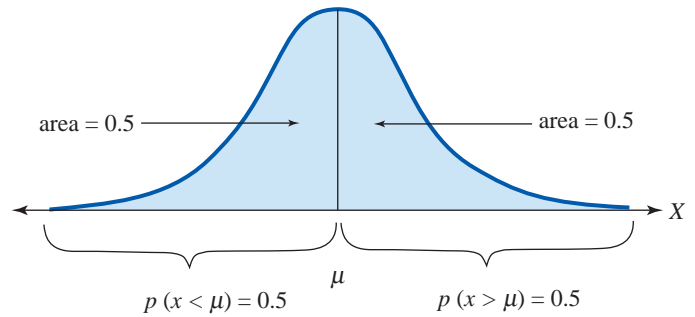


FIGURE 4.103 Total probability equals 1 (or 100%).

For a normal distribution, the outcomes nearer the center of the distribution occur more frequently than those at either end; the distribution is denser in the middle and sparser at the extremes. This difference in density is taken into account by consideration of the area under the bell curve; the center of the distribution is denser, contains more area, and has a higher probability of occurrence than the extremes. Consequently, we use the area under the bell curve to represent the probability of an outcome. Because $p(S) = 1$, we define the entire area under the bell curve to equal 1.

Because a normal distribution is symmetric, 50% of the data will be greater than the mean, and 50% will be less. (The mean and the median coincide in a symmetric distribution.) Therefore, the probability of randomly selecting a number x greater than the mean is $p(x > \mu) = 0.5$, and that of selecting a number x less than the mean is $p(x < \mu) = 0.5$, as shown in Figure 4.103.

To find the probability that a randomly selected number x is between two values (say a and b), we must determine the area under the curve from a to b ; that is, $p(a < x < b) = \text{area under the bell curve from } x = a \text{ to } x = b$, as shown in Figure 4.104(a). Likewise, the probability that x is greater than or less than any specific number is given by the area of the tail, as shown in Figure 4.104(b). To find probabilities involving data that are normally distributed, we must find the area of the appropriate region under the bell curve.

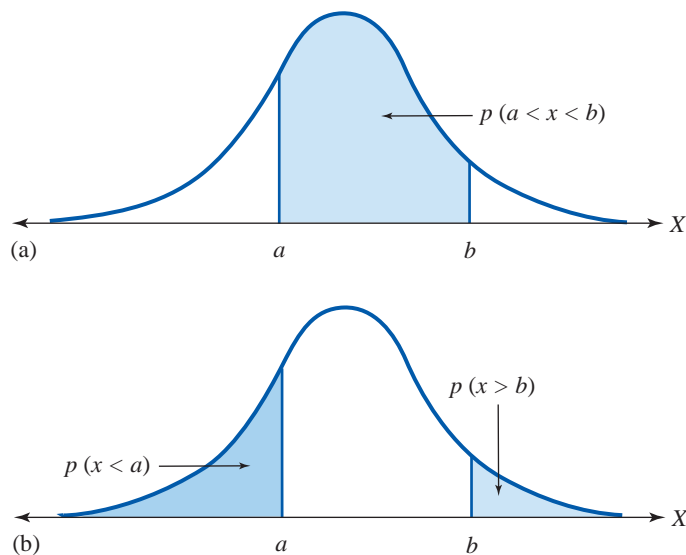


FIGURE 4.104 Regions under a bell curve.

HISTORICAL NOTE

CARL FRIEDRICH GAUSS, 1777–1855

Dubbed “the Prince of Mathematics,” Carl Gauss is considered by many to be one of the greatest mathematicians of all time. At the age of three, Gauss is said to have discovered an arithmetic error in his father’s book-keeping. The child prodigy was encouraged by his teachers and excelled throughout his early schooling. When he was fourteen, Gauss was introduced to Ferdinand, the Duke of Brunswick. Impressed with the youth, the duke gave Gauss a yearly stipend and sponsored his education for many years.

In 1795, Gauss enrolled at Göttingen University, where he remained for three years. While at Göttingen, Gauss had complete academic freedom; he was not required to attend lectures, he had no required conferences with professors or tutors, and he did not take exams. Much of his time was spent studying independently in the library. For reasons unknown to us, Gauss left the university in 1798 without a diploma. Instead, he sent his dissertation to the University of Helmstedt and in 1799 was awarded his degree without the usual oral examination.

In 1796, Gauss began his famous mathematical diary. Discovered forty years after his death, the 146 sometimes cryptic entries exhibit the diverse range of topics that Gauss pondered and pioneered. The first entry was



© alky-images

Gauss’s discovery (at the age of nineteen) of a method for constructing a seventeen-sided polygon with a compass and a straightedge. Other entries include important results in number theory, algebra, calculus, analysis, astronomy, electricity, magnetism,

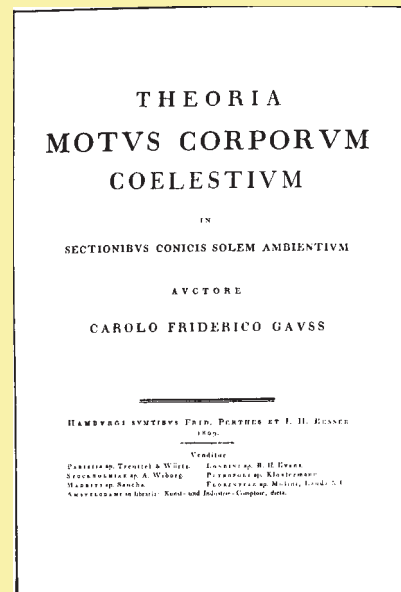
the foundations of geometry, and probability.

At the dawn of the nineteenth century, Gauss began his lifelong study of astronomy. On January 1, 1801, the Italian astronomer Giuseppe Piazzi discovered Ceres, the first of the known planetoids (minor planets or asteroids). Piazzi and others observed Ceres for forty-one days, until it was lost behind the sun. Because of his interest in the mathematics of astronomy, Gauss turned his attention to Ceres. Working with a minimal amount of data, he successfully calculated the orbit of Ceres. At the end of the year, the planetoid was rediscovered in exactly the spot that Gauss had predicted!

To obtain the orbit of Ceres, Gauss utilized his method of least squares, a technique for dealing with experimental error. Letting x represent the error between an experimentally obtained value and the true value it represents, Gauss’s theory involved minimizing x^2 —that is, obtaining the least square of the error. Theorizing that the probability of a small error was higher than that of a large

error, Gauss subsequently developed the normal distribution, or bell-shaped curve, to explain the probabilities of the random errors. Because of his pioneering efforts, some mathematicians refer to the normal distribution as the Gaussian distribution.

In 1807, Gauss became director of the newly constructed observatory at Göttingen. He held the position until his death some fifty years later.



© Stock Montage

Published in 1809, Gauss’s *Theoria Motus Corporum Coelestium* (Theory of the Motion of Heavenly Bodies) contained rigorous methods of determining the orbits of planets and comets from observational data via the method of least squares. It is a landmark in the development of modern mathematical astronomy and statistics.

The Standard Normal Distribution

All normal distributions share the following features: they are symmetric, bell-shaped curves, and virtually all the data (99.74%) lie within three standard deviations of the mean. Depending on whether the standard deviation is large or small, the bell curve will be either flat and spread out or peaked and narrow, as shown in Figure 4.105.

To find the area under any portion of any bell curve, mathematicians have devised a means of comparing the proportions of any curve with the proportions of a

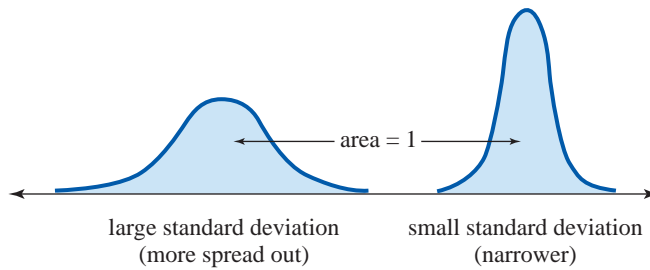


FIGURE 4.105 Large versus small standard deviation.

special curve defined as “standard.” To find probabilities involving normally distributed data, we utilize the bell curve associated with the standard normal distribution.

The **standard normal distribution** is the normal distribution whose mean is 0 and standard deviation is 1, as shown in Figure 4.106. The standard normal distribution is also called the **z-distribution**; we will always use the letter z to refer to the standard normal. By convention, we will use the letter x to refer to any other normal distribution.

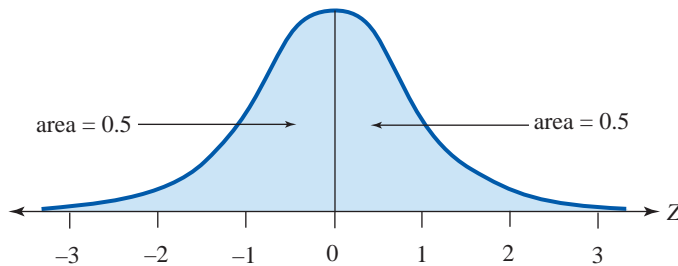
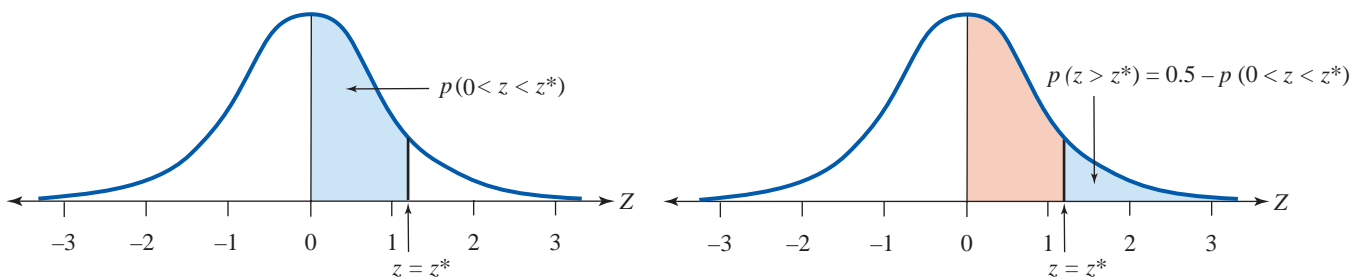


FIGURE 4.106 The standard normal distribution (mean = 0, standard deviation = 1).

Tables have been developed for finding areas under the standard normal curve using the techniques of calculus. Graphing calculators will also give these areas. We will use the table in Appendix F to find $p(0 < z < z^*)$, the probability that z is between 0 and a positive number z^* , as shown in Figure 4.107(a). The table in Appendix F is known as the **body table** because it gives the probability of an interval located in the middle, or body, of the bell curve.

The tapered end of a bell curve is known as a **tail**. To find the probability of a tail—that is, to find $p(z > z^*)$ or $p(z < z^*)$ where z^* is a positive real number—subtract the probability of the corresponding body from 0.5, as shown in Figure 4.107(b).



(a) Area found by using the body table (Appendix F)

(b) Area of a tail, found by subtracting the corresponding body area from 0.5

FIGURE 4.107

EXAMPLE 2

FINDING PROBABILITIES OF THE STANDARD NORMAL DISTRIBUTION Find the following probabilities (that is, the areas), where z represents the standard normal distribution.

- a. $p(0 < z < 1.25)$ b. $p(z > 1.87)$

SOLUTION

- a. As a first step, it is always advisable to draw a picture of the z -curve and shade in the desired area. We will use the body table directly, because we are working with a central area (see Figure 4.108).

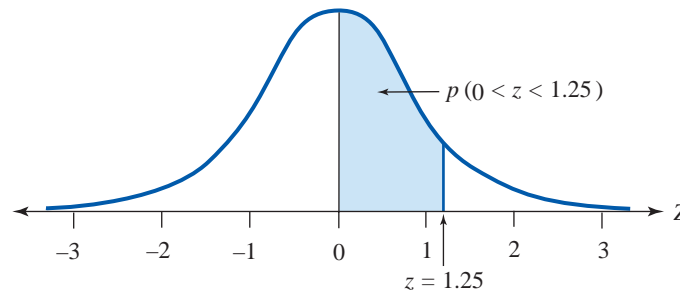


FIGURE 4.108 A central region, or body.

The z -numbers are located along the left edge and the top of the table. Locate the whole number and the first-decimal-place part of the number (1.2) along the left edge; then locate the second-decimal-place part of the number (0.05) along the top. The desired probability (area) is found at the intersection of the row and column of the two parts of the z -number. Thus, $p(0 < z < 1.25) = 0.3944$, as shown in Figure 4.109.

z^*	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0										
0.1										
0.2										
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4										
1.5										

FIGURE 4.109 A portion of the body table.

Hence, we could say that about 39% of the z -distribution lies between $z = 0$ and $z = 1.25$.

- b. To find the area of a tail, we subtract the corresponding body area from 0.5, as shown in Figure 4.110. Therefore,

$$\begin{aligned} p(z > 1.87) &= 0.5 - p(0 < z < 1.87) \\ &= 0.5 - 0.4692 \\ &= 0.0308 \end{aligned}$$

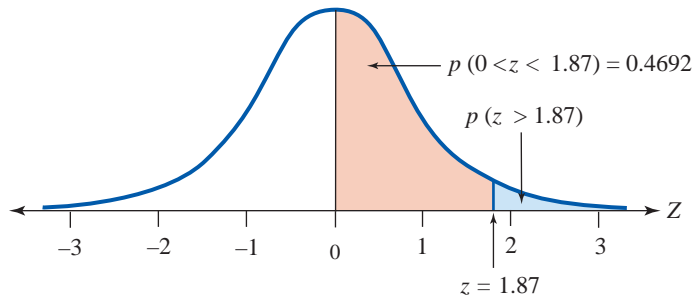


FIGURE 4.110 Finding the area of a tail.

The body table can also be used to find areas other than those given explicitly as $p(0 < z < z^*)$ and $p(z > z^*)$ where z^* is a positive number. By adding or subtracting two areas, we can find probabilities of the type $p(a < z < b)$, where a and b are positive or negative numbers, and probabilities of the type $p(z < c)$, where c is a positive or negative number.

EXAMPLE 3

FINDING PROBABILITIES OF THE STANDARD NORMAL DISTRIBUTION Find the following probabilities (the areas), where z represents the standard normal distribution.

- a. $p(0.75 < z < 1.25)$ b. $p(-0.75 < z < 1.25)$

SOLUTION

- a. Because the required region, shown in Figure 4.111, doesn't begin exactly at $z = 0$, we cannot look up the desired area directly in the body table. Whenever z is between two nonzero numbers, we will take an indirect approach to finding the required area.

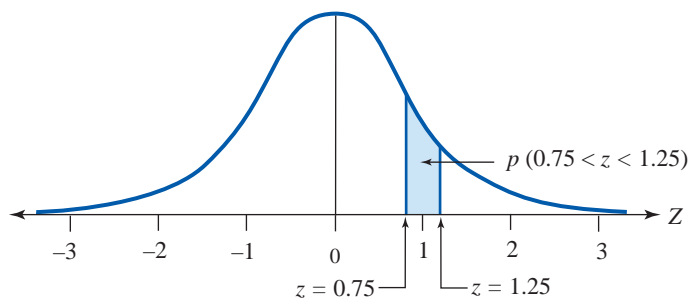


FIGURE 4.111 A strip.

The total area under the curve from $z = 0$ to $z = 1.25$ can be divided into two portions: the area under the curve from 0 to 0.75 and the area under the curve from 0.75 to 1.25.

To find the area of the “strip” between $z = 0.75$ and $z = 1.25$, we *subtract* the area of the smaller body (from $z = 0$ to $z = 0.75$) from that of the larger body (from $z = 0$ to $z = 1.25$), as shown in Figure 4.112.

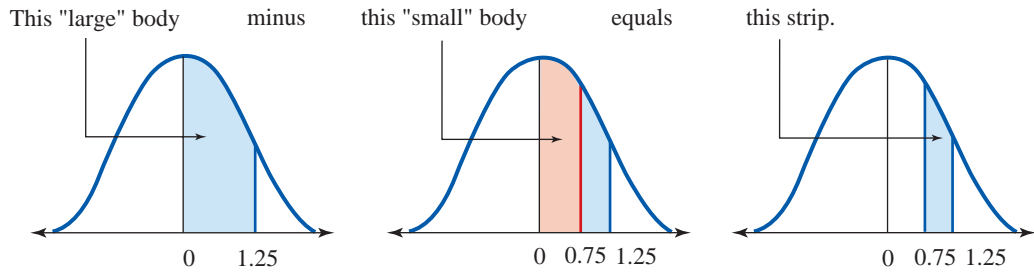


FIGURE 4.112 Area of a strip.

$$\begin{aligned} \text{area of strip} &= \text{area of large body} - \text{area of small body} \\ p(0.75 < z < 1.25) &= p(0 < z < 1.25) - p(0 < z < 0.75) \\ &= 0.3944 - 0.2734 \\ &= 0.1210 \end{aligned}$$

Therefore, $p(0.75 < z < 1.25) = 0.1210$. Hence, we could say that about 12.1% of the z -distribution lies between $z = 0.75$ and $z = 1.25$.

- b.** The required region, shown in Figure 4.113, can be divided into two regions: the area from $z = -0.75$ to $z = 0$ and the area from $z = 0$ to $z = 1.25$.

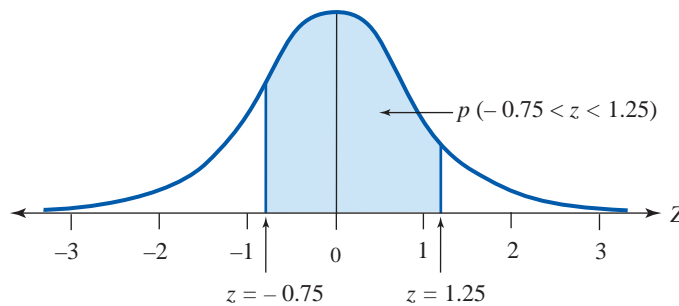


FIGURE 4.113 A central region.

To find the total area of the region between $z = -0.75$ and $z = 1.25$, we *add* the area of the "left" body (from $z = -0.75$ to $z = 0$) to the area of the "right" body (from $z = 0$ to $z = 1.25$), as shown in Figure 4.114.

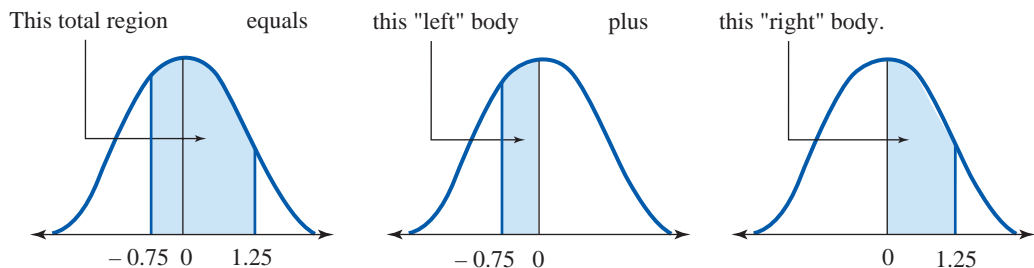


FIGURE 4.114 Left body plus right body.

This example is different from our previous examples in that it contains a negative z -number. A glance at the tables reveals that negative numbers are not included! However, recall that normal distributions are symmetric. Therefore, the area of the body from $z = -0.75$ to $z = 0$ is the same as that from $z = 0$ to $z = 0.75$; that is, $p(-0.75 < z < 0) = p(0 < z < 0.75)$. Therefore,

$$\begin{aligned}
 \text{total area of region} &= \text{area of left body} + \text{area of right body} \\
 p(-0.75 < z < 1.25) &= p(-0.75 < z < 0) + p(0 < z < 1.25) \\
 &= p(0 < z < 0.75) + p(0 < z < 1.25) \\
 &= 0.2734 + 0.3944 \\
 &= 0.6678
 \end{aligned}$$

Therefore, $p(-0.75 < z < 1.25) = 0.6678$. Hence, we could say that about 66.8% of the z -distribution lies between $z = -0.75$ and $z = 1.25$.

EXAMPLE 4

FINDING PROBABILITIES OF THE STANDARD NORMAL DISTRIBUTION Find the following probabilities (the areas), where z represents the standard normal distribution.

- $p(z < 1.25)$
- $p(z < -1.25)$

SOLUTION

- The required region is shown in Figure 4.115. Because 50% of the distribution lies to the left of 0, we can add 0.5 to the area of the body from $z = 0$ to $z = 1.25$:

$$\begin{aligned}
 p(z < 1.25) &= p(z < 0) + p(0 < z < 1.25) \\
 &= 0.5 + 0.3944 \\
 &= 0.8944
 \end{aligned}$$

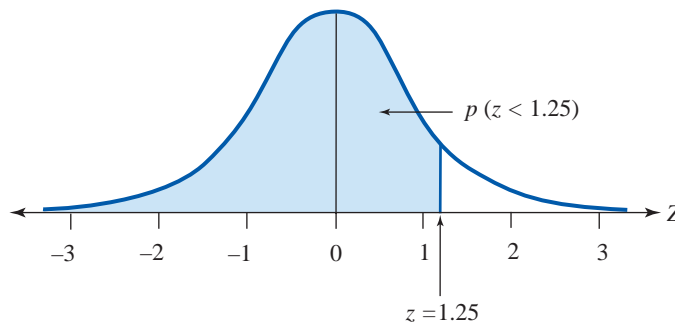


FIGURE 4.115 A body plus 50%.

Therefore, $p(z < 1.25) = 0.8944$. Hence, we could say that about 89.4% of the z -distribution lies to the left of $z = 1.25$.

- The required region is shown in Figure 4.116. Because a normal distribution is symmetric, the area of the left tail ($z < -1.25$) is the same as the area of the corresponding right tail ($z > 1.25$).

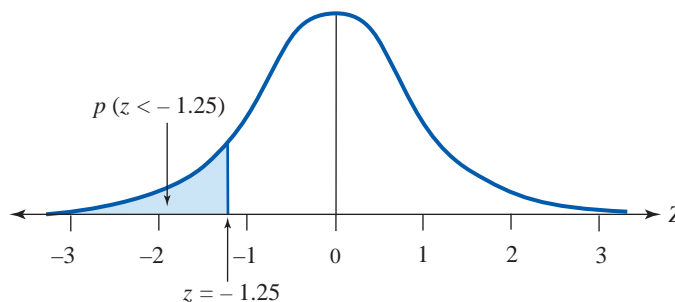


FIGURE 4.116 A left tail.

Therefore,

$$\begin{aligned} p(z < -1.25) &= p(z > 1.25) \\ &= 0.5 - p(0 < z < 1.25) \\ &= 0.5 - 0.3944 \\ &= 0.1056 \end{aligned}$$

Hence, we could say that about 10.6% of the z -distribution lies to the left of $z = -1.25$.

Converting to the Standard Normal

Weather forecasters in the United States usually report temperatures in degrees Fahrenheit. Consequently, if a temperature is given in degrees Celsius, most people convert it to Fahrenheit in order to judge how hot or cold it is. A similar situation arises when we are working with a normal distribution. Suppose we know that a large set of data is normally distributed with a mean value of 68 and a standard deviation of 4. What percent of the data will lie between 65 and 73? We are asked to find $p(65 < x < 73)$. To find this probability, we must first convert the given normal distribution to the standard normal distribution and then look up the approximate z -numbers.

The body table (Appendix F) applies to the standard normal z -distribution. When we are working with any other normal distribution (denoted by X), we must first convert the x -distribution into the standard normal z -distribution. This conversion is done with the help of the following rule.

Given a number x , its corresponding z -number counts the number of standard deviations the number lies from the mean. For example, suppose the mean and standard deviation of a normal distribution are $\mu = 68$ and $\sigma = 4$. The z -number corresponding to $x = 78$ is

$$z = \frac{x - \mu}{\sigma} = \frac{78 - 68}{4} = 2.5$$

This implies that $x = 78$ lies two and one-half standard deviations above the mean, 68. Similarly, for $x = 65$,

$$z = \frac{65 - 68}{4} = -0.75$$

Therefore, $x = 65$ lies three-quarters of a standard deviation below the mean, 68.

CONVERTING A NORMAL DISTRIBUTION INTO THE STANDARD NORMAL Z

Every number x in a given normal distribution has a corresponding number z in the standard normal distribution. The **z -number** that corresponds to the number x is

$$z = \frac{x - \mu}{\sigma}$$

where μ is the mean and σ the standard deviation of the given normal distribution.

EXAMPLE 5

FINDING A PROBABILITY OF A NONSTANDARD NORMAL DISTRIBUTION Suppose a population is normally distributed with a mean of 24.6 and a standard deviation of 1.3. What percent of the data will lie between 25.3 and 26.8?

SOLUTION

We are asked to find $p(25.3 < x < 26.8)$, the area of the region shown in Figure 4.117. Because we need to find the area of the strip between 25.3 and 26.8, we must find the body of each and subtract, as in part (a) of Example 3.

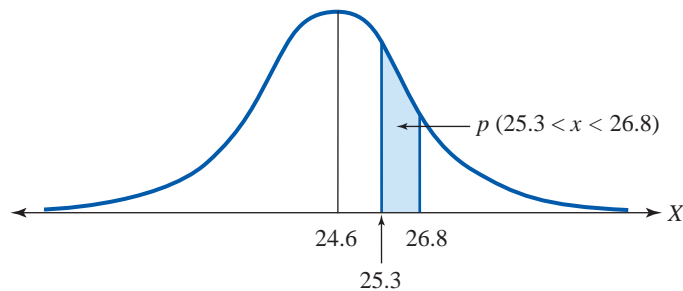


FIGURE 4.117 A strip.

Using the Conversion Formula $z = (x - \mu)/\sigma$ with $\mu = 24.6$ and $\sigma = 1.3$, we first convert $x = 25.3$ and $x = 26.8$ into their corresponding z -numbers.

Converting $x = 25.3$

$$\begin{aligned} z &= \frac{x - \mu}{\sigma} \\ &= \frac{25.3 - 24.6}{1.3} \\ &= 0.5384615 \\ &= 0.54 \end{aligned}$$

Converting $x = 26.8$

$$\begin{aligned} z &= \frac{x - \mu}{\sigma} \\ &= \frac{26.8 - 24.6}{1.3} \\ &= 1.6923077 \\ &= 1.69 \end{aligned}$$

rounding off z -numbers to two decimal places

Therefore,

$$\begin{aligned} p(25.3 < x < 26.8) &= p(0.54 < z < 1.69) \\ &= p(0 < z < 1.69) - p(0 < z < 0.54) \\ &= 0.4545 - 0.2054 && \text{using the body table} \\ &= 0.2491 \end{aligned}$$

Assuming a normal distribution, approximately 24.9% of the data will lie between 25.3 and 26.8.

EXAMPLE 6

FINDING PROBABILITIES OF A NONSTANDARD NORMAL DISTRIBUTION The heights of a large group of people are assumed to be normally distributed. Their mean height is 68 inches, and the standard deviation is 4 inches. What percentage of these people are the following heights?

- a. taller than 73 inches b. between 60 and 75 inches

SOLUTION

a. Let x represent the height of a randomly selected person. We need to find $p(x > 73)$, the area of a tail, as shown in Figure 4.118.

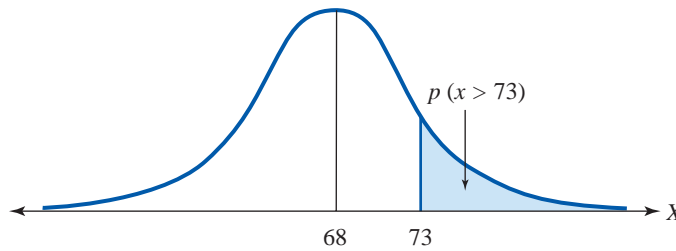


FIGURE 4.118 A right tail.

First, we must convert $x = 73$ to its corresponding z -number. Using the Conversion Formula with $x = 73$, $\mu = 68$, and $\sigma = 4$, we have

$$\begin{aligned} z &= \frac{x - \mu}{\sigma} \\ &= \frac{73 - 68}{4} \\ &= 1.25 \end{aligned}$$

Therefore,

$$\begin{aligned} p(x > 73) &= p(z > 1.25) \\ &= 0.5 - p(0 < z < 1.25) \\ &= 0.5 - 0.3944 \\ &= 0.1056 \end{aligned}$$

Approximately 10.6% of the people will be taller than 73 inches.

- b.** We need to find $p(60 < x < 75)$, the area of the central region shown in Figure 4.119. Notice that we will be adding the areas of the two bodies.

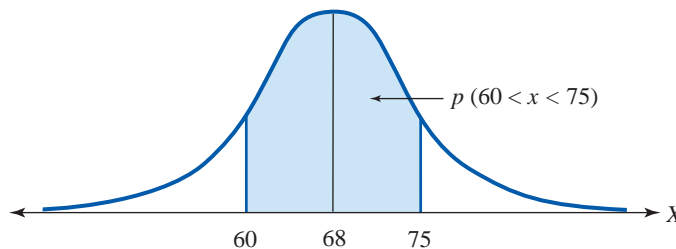


FIGURE 4.119 A central region.

First, we convert $x = 60$ and $x = 75$ to their corresponding z -numbers:

$$\begin{aligned} p(60 < x < 75) &= p\left(\frac{60 - 68}{4} < z < \frac{75 - 68}{4}\right) \\ &= p(-2.00 < z < 1.75) \\ &= p(-2.00 < z < 0) + p(0 < z < 1.75) \\ &= p(0 < z < 2.00) + p(0 < z < 1.75) \\ &= 0.4772 + 0.4599 \\ &= 0.9371 \end{aligned}$$

using the
Conversion
Formula

expressing the
area as two
bodies
using
symmetry
using the
body table

Approximately 93.7% of the people will be between 60 and 75 inches tall.

All the preceding examples involved finding probabilities that contained only the strict $<$ or $>$ inequalities, never \leq or \geq inequalities; the endpoints were

never included. What if the endpoints are included? How does $p(a < x < b)$ compare with $p(a \leq x \leq b)$? Because probabilities for continuous data are found by determining *area* under a curve, including the endpoints does not affect the probability! The probability of a single point $p(x = a)$ is 0, because there is no “area” over a single point. (We obtain an area only when we are working with an interval of numbers.) Consequently, if x represents continuous data, then $p(a \leq x \leq b) = p(a < x < b)$; it makes no difference whether the endpoints are included.

EXAMPLE 7

FINDING THE VALUE OF A VARIABLE THAT WILL PRODUCE A SPECIFIC PROBABILITY Tall Dudes is a clothing store that specializes in fashions for tall men. Its informal motto is “Our customers are taller than 80% of the rest.” Assuming the heights of men to be normally distributed with a mean of 67 inches and a standard deviation of 5.5 inches, find the heights of Tall Dudes’ clientele.

SOLUTION

Let c = the height of the shortest customer at Tall Dudes, and let x represent the height of a randomly selected man. We are given that the heights of all men are normally distributed with $\mu = 67$ and $\sigma = 5.5$.

Assuming Tall Dudes’ clientele to be taller than 80% of all men implies that $x < c$ 80% of the time and $x > c$ 20% of the time. Hence, we can say that the probability of selecting someone shorter than the shortest tall dude is $p(x < c) = 0.80$ and that the probability of selecting a tall dude is $p(x > c) = 0.20$, as shown in Figure 4.120.

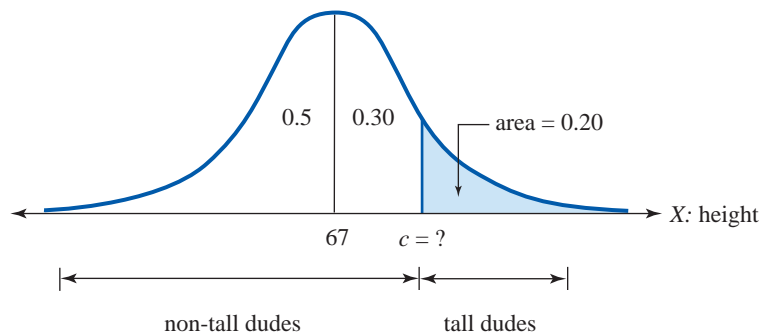


FIGURE 4.120 Finding a 20% right tail.

We are given that the area of the right tail is 0.20 (and that of the right body is 0.30), and we need to find the appropriate cutoff number c . This is exactly the reverse of all the previous examples, in which we were given the cutoff numbers and asked to find the area. Thus, our goal is to find the z -number that corresponds to a body of area 0.30 and convert it into its corresponding x -number.

When we scan through the *interior* of the body table, the number closest to the desired area of 0.30 is 0.2995, which is the area of the body when $z = 0.84$. This means that $p(0 < z < 0.84) = 0.2995$, as shown in Figure 4.121.

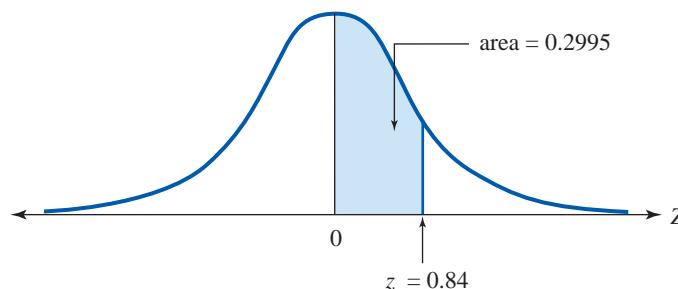


FIGURE 4.121 A body close to 30%.

Therefore, the number c that we are seeking lies 0.84 standard deviations above the mean. All that remains is to convert $z = 0.84$ into its corresponding x -number by substituting $x = c$, $z = 0.84$, $\mu = 67$, and $\sigma = 5.5$ into the Conversion Formula:

$$z = \frac{x - \mu}{\sigma}$$

$$0.84 = \frac{c - 67}{5.5}$$

$$(5.5)(0.84) = c - 67$$

$$4.62 = c - 67$$

$$c = 71.62 \quad (\approx 72 \text{ inches, or 6 feet})$$

Therefore, Tall Dudes caters to men who are at least 71.62 inches tall, or about 6 feet tall.

4.4 EXERCISES

- The weights (in ounces) of several bags of corn chips are given in Figure 4.122. Construct a histogram for the data using the groups $15.40 \leq x < 15.60$, $15.60 \leq x < 15.80$, \dots , $16.60 \leq x < 16.80$. Do the data appear to be approximately normally distributed? Explain.

16.08	16.49	15.61	16.66	15.80	15.87
16.02	15.82	16.48	16.08	15.63	16.02
16.00	16.25	15.41	16.22	16.04	15.68
16.45	16.41	16.01	15.82	16.08	15.82
16.29	16.26	16.05	16.25	15.86	

FIGURE 4.122 Weights (in ounces) of bags of corn chips.

- The time (in minutes) spent waiting in line for several students at the campus bookstore are given in Figure 4.124. Construct a histogram for the data using the groups $0 \leq x < 1$, $1 \leq x < 2$, \dots , $4 \leq x < 5$. Do the data appear to be normally distributed? Explain.

1.75	2	0.5	1.5	0
1	2.5	0.75	2.25	3
1.5	3.5	2.5	1	0.5
1.75	0.25	4.5	0	1.5

FIGURE 4.124 Time (in minutes) spent waiting in line.

- The weights (in grams) of several bags of chocolate chip cookies are given in Figure 4.123. Construct a histogram for the data using the groups $420 \leq x < 430$, $430 \leq x < 440$, \dots , $480 \leq x < 490$. Do the data appear to be approximately normally distributed? Explain.

451	435	482	449	454	451
479	448	432	423	461	475
467	453	448	459	454	444
475	461	450	466	446	458

FIGURE 4.123 Weights (in grams) of bags of chocolate chip cookies.

- A die was rolled several times, and the results are given in Figure 4.125. Construct a histogram for the data using the single-values 1, 2, 3, 4, 5, and 6. Do the data appear to be normally distributed? Explain.

5	3	1	1	4	2
6	4	1	2	5	5
4	3	1	6	4	2
3	1	2	5	2	1
4	6	5	1	4	5

FIGURE 4.125 Rolling a die.

Selected exercises available online at www.webassign.net/brookscole

- ▶ 5. What percent of the standard normal z -distribution lies between the following values?
- $z = 0$ and $z = 1$
 - $z = -1$ and $z = 0$
 - $z = -1$ and $z = 1$
- (Note: This interval represents one standard deviation of the mean.)
- ▶ 6. What percent of the standard normal z -distribution lies between the following values?
- $z = 0$ and $z = 2$
 - $z = -2$ and $z = 0$
 - $z = -2$ and $z = 2$
- (Note: This interval represents two standard deviations of the mean.)
7. What percent of the standard normal z -distribution lies between the following values?
- $z = 0$ and $z = 3$
 - $z = -3$ and $z = 0$
 - $z = -3$ and $z = 3$
- (Note: This interval represents three standard deviations of the mean.)
- ▶ 8. What percent of the standard normal z -distribution lies between the following values?
- $z = 0$ and $z = 1.5$
 - $z = -1.5$ and $z = 0$
 - $z = -1.5$ and $z = 1.5$
- (Note: This interval represents one and one-half standard deviations of the mean.)
9. A population is normally distributed with mean 24.7 and standard deviation 2.3.
- Find the intervals representing one, two, and three standard deviations of the mean.
 - What percentage of the data lies in each of the intervals in part (a)?
 - Draw a sketch of the bell curve.
- ▶ 10. A population is normally distributed with mean 18.9 and standard deviation 1.8.
- Find the intervals representing one, two, and three standard deviations of the mean.
 - What percent of the data lies in each of the intervals in part (a)?
 - Draw a sketch of the bell curve.
- ▶ 11. Find the following probabilities.
- $p(0 < z < 1.62)$
 - $p(1.30 < z < 1.84)$
 - $p(-0.37 < z < 1.59)$
 - $p(z < -1.91)$
 - $p(-1.32 < z < -0.88)$
 - $p(z < 1.25)$
- ▶ 12. Find the following probabilities.
- $p(0 < z < 1.42)$
 - $p(1.03 < z < 1.66)$
 - $p(-0.87 < z < 1.71)$
 - $p(z < -2.06)$
 - $p(-2.31 < z < -1.18)$
 - $p(z < 1.52)$
- ▶ 13. Find c such that each of the following is true.
- $p(0 < z < c) = 0.1331$
 - $p(c < z < 0) = 0.4812$
 - $p(-c < z < c) = 0.4648$
 - $p(z > c) = 0.6064$
 - $p(z > c) = 0.0505$
 - $p(z < c) = 0.1003$
14. Find c such that each of the following is true.
- $p(0 < z < c) = 0.3686$
 - $p(c < z < 0) = 0.4706$
 - $p(-c < z < c) = 0.2510$
 - $p(z > c) = 0.7054$
 - $p(z > c) = 0.0351$
 - $p(z < c) = 0.2776$
15. A population X is normally distributed with mean 250 and standard deviation 24. For each of the following values of x , find the corresponding z -number. Round off your answers to two decimal places.
- | | |
|--------------|--------------|
| a. $x = 260$ | b. $x = 240$ |
| c. $x = 300$ | d. $x = 215$ |
| e. $x = 321$ | f. $x = 197$ |
- ▶ 16. A population X is normally distributed with mean 72.1 and standard deviation 9.3. For each of the following values of x , find the corresponding z -number. Round off your answers to two decimal places.
- | | |
|-------------|-------------|
| a. $x = 90$ | b. $x = 80$ |
| c. $x = 75$ | d. $x = 70$ |
| e. $x = 60$ | f. $x = 50$ |
- ▶ 17. A population is normally distributed with mean 36.8 and standard deviation 2.5. Find the following probabilities.
- | | |
|-------------------------|-------------------------|
| a. $p(36.8 < x < 39.3)$ | b. $p(34.2 < x < 38.7)$ |
| c. $p(x < 40.0)$ | d. $p(32.3 < x < 41.3)$ |
| e. $p(x = 37.9)$ | f. $p(x > 37.9)$ |
- ▶ 18. A population is normally distributed with mean 42.7 and standard deviation 4.7. Find the following probabilities.
- | | |
|-------------------------|-------------------------|
| a. $p(42.7 < x < 47.4)$ | b. $p(40.9 < x < 44.1)$ |
| c. $p(x < 50.0)$ | d. $p(33.3 < x < 52.1)$ |
| e. $p(x = 45.3)$ | f. $p(x > 45.3)$ |

- ▶ 19. The mean weight of a box of cereal filled by a machine is 16.0 ounces, with a standard deviation of 0.3 ounce. If the weights of all the boxes filled by the machine are normally distributed, what percent of the boxes will weigh the following amounts?
- less than 15.5 ounces
 - between 15.8 and 16.2 ounces
- ▶ 20. The amount of time required to assemble a component on a factory assembly line is normally distributed with a mean of 3.1 minutes and a standard deviation of 0.6 minute. Find the probability that a randomly selected employee will take the given amount of time to assemble the component.
- more than 4.0 minutes
 - between 2.0 and 2.5 minutes
21. The time it takes an acrylic paint to dry is normally distributed. If the mean is 2 hours 36 minutes with a standard deviation of 24 minutes, find the probability that the drying time will be as follows.
- less than 2 hours 15 minutes
 - between 2 and 3 hours
- HINT:* Convert everything to minutes (or to hours).
- ▶ 22. The shrinkage in length of a certain brand of blue jeans is normally distributed with a mean of 1.1 inches and a standard deviation of 0.2 inch. What percent of this brand of jeans will shrink the following amounts?
- more than 1.5 inches
 - between 1.0 and 1.25 inches
23. The mean volume of a carton of milk filled by a machine is 1.0 quart, with a standard deviation of 0.06 quart. If the volumes of all the cartons are normally distributed, what percent of the cartons will contain the following amounts?
- at least 0.9 quart
 - at most 1.05 quarts
- ▶ 24. The amount of time between taking a pain reliever and getting relief is normally distributed with a mean of 23 minutes and a standard deviation of 4 minutes. Find the probability that the time between taking the medication and getting relief is as follows.
- at least 30 minutes
 - at most 20 minutes
- ▶ 25. The results of a statewide exam for assessing the mathematics skills of realtors were normally distributed with a mean score of 72 and a standard deviation of 12. The realtors who scored in the top 10% are to receive a special certificate, while those in the bottom 20% will be required to attend a remedial workshop.
- What score does a realtor need in order to receive a certificate?
 - What score will dictate that the realtor attend the workshop?
- ▶ 26. Professor Harde assumes that exam scores are normally distributed and wants to grade “on the curve.” The mean score was 58, with a standard deviation of 16.
- If she wants 14% of the students to receive an A, find the minimum score to receive an A.
 - If she wants 19% of the students to receive a B, find the minimum score to receive a B.
27. The time it takes an employee to package the components of a certain product is normally distributed with $\mu = 8.5$ minutes and $\sigma = 1.5$ minutes. To boost productivity, management has decided to give special training to the 34% of employees who took the greatest amount of time to package the components. Find the amount of time taken to package the components that will indicate that an employee should get special training.
- ▶ 28. The time it takes an employee to package the components of a certain product is normally distributed with $\mu = 8.5$ and $\sigma = 1.5$ minutes. As an incentive, management has decided to give a bonus to the 20% of employees who took the shortest amount of time to package the components. Find the amount of time taken to package the components that will indicate that an employee should get a bonus.



Answer the following questions using complete sentences and your own words.

• CONCEPT QUESTIONS

- What are the characteristics of a normal distribution?
- Are all distributions of data normally distributed? Support your answer with an example.
- Why is the total area under a bell curve equal to 1?
- Why are there no negative z -numbers in the body table?
- When converting an x -number to a z -number, what does a negative z -number tell you about the location of the x -number?
- Is it logical to assume that the heights of all high school students in the United States are normally distributed? Explain.
- Is it reasonable to assume that the ages of all high school students in the United States are normally distributed? Explain.

• HISTORY QUESTIONS

- Who is known as “the Prince of Mathematics”? Why?
- What mathematician was instrumental in the creation of the normal distribution? What application prompted this person to create the normal distribution?

4.5

Polls and Margin of Error

OBJECTIVES

- Find the margin of error in a poll
- Determine the effect of sample size on the margin of error
- Find the level of confidence for a specific sample and margin of error
- Find the minimum sample size to obtain a specified margin of error

One of the most common applications of statistics is the evaluation of the results of surveys and public opinion polls. Most editions of the daily newspaper contain the results of at least one poll. Headlines announce the attitude of the nation toward a myriad of topics ranging from the actions of politicians to controversial current issues, such as abortion, as shown in the newspaper article on the next page. How are these conclusions reached? What do they mean? How valid are they? In this section, we investigate these questions and obtain results concerning the “margin of error” associated with the reporting of “public opinion.”

Sampling and Inferential Statistics

The purpose of conducting a survey or poll is to obtain information about a population—for example, adult Americans. Because there are approximately 230 million Americans over the age of eighteen, it would be very difficult, time-consuming, and expensive to contact every one of them. The only realistic alternative is to poll a sample and use the science of inferential statistics to draw conclusions about the population as a whole. Different samples have different characteristics depending on, among other things, the age, sex, education, and locale of the people in the sample. Therefore, it is of the utmost importance that a sample be representative of the population. Obtaining a representative sample is the most difficult aspect of inferential statistics.

Another problem facing pollsters is determining *how many* people should be selected for the sample. Obviously, the larger the sample, the more likely that it will reflect the population. However, larger samples cost more money, so a limited budget will limit the sample size. Conducting surveys can be very costly, even for a small to moderate sample. For example, a survey conducted in 1989 by the Gallup Organization that contacted 1,005 adults and 500 teenagers would have cost \$100,000 (the pollsters donated their services for this survey). The results of this poll indicated that Americans thought the “drug crisis” was the nation’s top problem (stated by 27% of the adults and 32% of the teenagers).

After a sample has been selected and its data have been analyzed, information about the sample is generalized to the entire population. Because 27% of the 1,005 adults in a poll stated that the drug crisis was the nation’s top problem, we would like to conclude that 27% of *all* adults have the same belief. Is this a valid generalization? That is, how confident is the pollster that the feelings of the people in the sample reflect those of the population?

FEATURED IN THE NEWS

MORE AMERICANS “PRO-LIFE” THAN “PRO-CHOICE” FOR THE FIRST TIME

Princeton, NJ—A new Gallup Poll, conducted May 7–10, 2009, finds 51% of Americans calling themselves “pro-life” on the issue of abortion and 42% “pro-choice.” This is the first time a majority of U.S. adults have identified themselves as pro-life since Gallup began asking this question in 1995. The new results, obtained from Gallup’s annual Values and Beliefs survey, represent a significant shift from a year ago, when 50% were pro-choice and 44% pro-life. Prior to now, the highest percentage identifying as pro-life was 46%, in both August 2001 and May 2002.

The source of the shift in abortion views is clear in the Gallup Values and

Belief survey. The percentage of Republicans (including independents who lean Republican) calling themselves “pro-life” rose by 10 points over the past year, from 60% to 70%, while there has been essentially no change in the views of Democrats and Democratic leaners. Similarly, by ideology, all of the increase in pro-life sentiment is seen among self-identified conservatives and moderates; the abortion views of political liberals have not changed.

A year ago, Gallup found more women calling themselves pro-choice than pro-life, by 50% to 43%, while men were more closely divided: 49% pro-choice, 46% pro-life. Now, because of heightened pro-life sentiment among both groups, women as well as

men are more likely to be pro-life. Men and women have been evenly divided on the issue in previous years; however, this is the first time in nine years of Gallup Values surveys that significantly more men and women are pro-life than pro-choice.

Results are based on telephone interviews with 1,015 national adults, aged 18 and older, conducted May 7–10, 2009. For results based on the total sample of national adults, one can say with 95% confidence that the maximum margin of sampling error is ± 3 percentage points.

By Lydia Saad
Gallup News Service

Sample Proportion versus Population Proportion

If x members (for example, people, automobiles, households) in a sample of size n have a certain characteristic, then the proportion of the sample, or **sample proportion**, having this characteristic is given by $\frac{x}{n}$. For instance, in a sample of $n = 70$ automobiles, if $x = 14$ cars have a defective fan switch, then the proportion of the sample having a defective switch is $\frac{14}{70} = 0.2$, or 20%. The true proportion of the entire population, or **population proportion**, having the characteristic is represented by the letter P . A sample proportion $\frac{x}{n}$ is an estimate of the population proportion P .

Sample proportions $\frac{x}{n}$ vary from sample to sample; some will be larger than P , and some will be smaller. Of the 1,005 adults in the Gallup Poll sample mentioned on page 295, 27% viewed the drug crisis as the nation’s top problem. If a different sample of 1,005 had been chosen, 29% might have had this view. If still another 1,005 had been selected, this view might have been shared by only 25%. We will assume that the sample proportions $\frac{x}{n}$ are normally distributed around the population proportion P . The set of all sample proportions, along with their probabilities of occurring, can be represented by a bell curve like the one in Figure 4.126.

In general, a sample estimate is not 100% accurate; although a sample proportion might be close to the true population proportion, it will have an error term associated with it. The difference between a sample estimate and the true (population) value is called the **error of the estimate**. We can use a bell curve (like the one in Figure 4.126) to predict the probable error of a sample estimate.

Before developing this method of predicting the error term, we need to introduce some special notation. The symbol z_α (read “z alpha”) will be used to represent the positive z -number that has a right body of area α . That is, z_α is the number such that $p(0 < z < z_\alpha) = \alpha$, as shown in Figure 4.127.

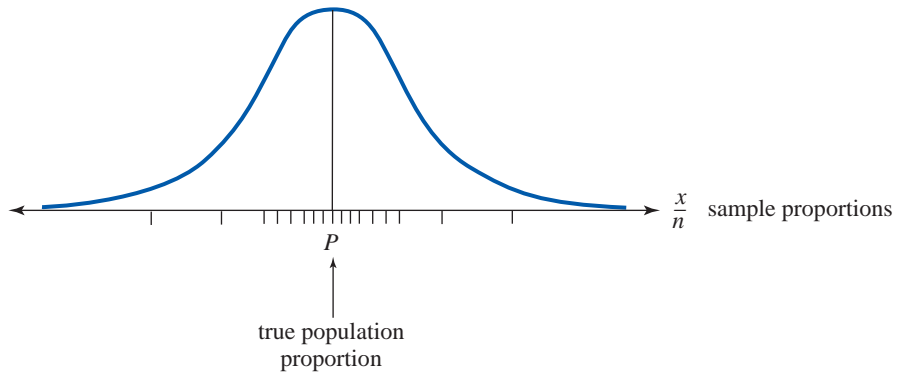


FIGURE 4.126 Sample proportions normally distributed around the true (population) proportion.

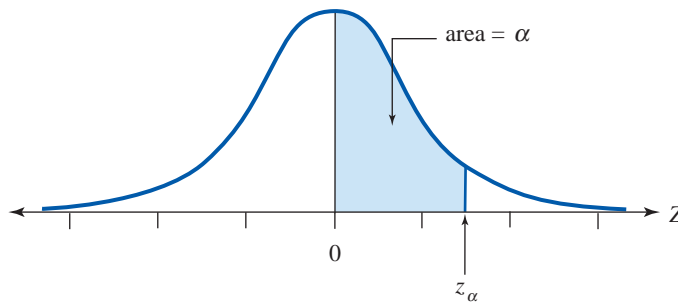


FIGURE 4.127 $p(0 < z < z_\alpha) = \alpha$.

EXAMPLE 1

FINDING A VALUE OF THE STANDARD NORMAL DISTRIBUTION THAT WILL PRODUCE A SPECIFIC PROBABILITY Use the body table in Appendix F to find the following values:

- a. $z_{0.3925}$ b. $z_{0.475}$ c. $z_{0.45}$ d. $z_{0.49}$

SOLUTION

- a. $z_{0.3925}$ represents the z -number that has a body of area 0.3925. Looking at the interior of the body table, we find 0.3925 and see that it corresponds to the z -number 1.24—that is, $p(0 < z < 1.24) = 0.3925$. Therefore, $z_{0.3925} = 1.24$.
- b. In a similar manner, we find that a body of area 0.4750 corresponds to $z = 1.96$. Therefore, $z_{0.475} = 1.96$.
- c. Looking through the interior of the table, we cannot find a body of area 0.45. However, we do find a body of 0.4495 (corresponding to $z = 1.64$) and a body of 0.4505 (corresponding to $z = 1.65$). Because the desired body is *exactly halfway* between the two listed bodies, we use a z -number that is exactly halfway between the two listed z -numbers, 1.64 and 1.65. Therefore, $z_{0.45} = 1.645$.
- d. We cannot find a body of the desired area, 0.49, in the interior of the table. The closest areas are 0.4898 (corresponding to $z = 2.32$) and 0.4901 (corresponding to $z = 2.33$). Because the desired area (0.49) is *closer* to 0.4901, we use $z = 2.33$. Therefore, $z_{0.49} = 2.33$.

Margin of Error

Sample proportions $\frac{x}{n}$ vary from sample to sample; some will have a small error, and some will have a large error. Knowing that sample estimates have inherent errors, statisticians make predictions concerning the largest possible error associated with

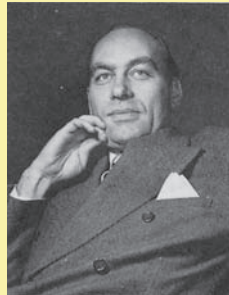
HISTORICAL NOTE

GEORGE H. GALLUP, 1901–1984

To many people, the name *Gallup* is synonymous with opinion polls. George Horace Gallup, the founder of the American Institute of Public Opinion, began his news career while attending the University of Iowa. During his junior year as a student of journalism, Gallup became the editor of his college newspaper, the *Daily Iowan*. After receiving his bachelor's degree in 1923, Gallup remained at the university nine years as an instructor of journalism.

In addition to teaching, Gallup continued his own studies of human nature and public opinion. Interest in how the public reacts to advertisements and perceives various issues of the day led Gallup to combine his study of journalism with the study of psychology. In 1925, he received his master's degree in psychology. Gallup's studies culminated in 1928 with his doctoral thesis, *A New Technique for Objective Methods for Measuring Reader Interest in Newspapers*. Gallup's new technique of polling the public was to utilize a stratified sample, that is, a sample that closely mirrors the composition of the entire population. Gallup contended that a stratified sample of 1,500 people was sufficient to obtain reliable estimates. For his pioneering work in this new field, Gallup was awarded his Ph.D. in journalism in 1928.

Gallup founded the American Institute of Public Opinion in 1935 with the stated purpose "to impartially measure and report public opinion on political and social issues of the day without regard to the rightness or wisdom of the



© Corbis/Bettman

views expressed." His first triumph was his prediction of the winner of the 1936 presidential election between Franklin D. Roosevelt and Alfred Landon. While many, including the prestigious *Literary Digest*, predicted that Landon would win, Gallup correctly predicted Roosevelt as the winner.

Gallup Polls have correctly predicted all presidential elections since, with the exception of the 1948 race between Thomas Dewey and Harry S. Truman. Much to his embarrassment, Gallup predicted Dewey as the winner. Truman won the election with 49.9% of the vote, while Gallup had predicted that he would receive only 44.5%. Gallup's explanation was that he had ended his poll too far in advance of election day and had disregarded the votes of those who were undecided. Of his error, Gallup said, "We are continually experimenting and continually learning."

Although some people criticize the use of polls, citing their potential influence and misuse, Gallup considered the public opinion poll to be "one of the most useful instruments of democracy ever devised." Answering the charge that he and his polls influenced elections, Gallup retorted, "One

might as well insist that a thermometer makes the weather!" In addition, Gallup confessed that he had not voted in a presidential election since 1928. Above all, Gallup wanted to ensure the impartiality of his polls.

Besides polling people regarding their choices in presidential campaigns, Gallup was the first pollster to ask the public to rate a president's performance and popularity. Today, these "presidential report cards" are so common we may take them for granted. In addition to presidential politics, Gallup also dealt with sociological issues, asking questions such as "What is the most important problem facing the country?"

Polling has become a multimillion-dollar business. In 2009, the Gallup Organization had revenues totaling \$271.6 million, and it had over 2,000 employees. Today, Gallup Polls are syndicated in newspapers across the country and around the world.



AP Photo/Byron Rollins

The Gallup Organization predicted that Thomas Dewey would win the 1948 presidential election. Much to Gallup's embarrassment, Harry S. Truman won the election and triumphantly displayed a newspaper containing Gallup's false prediction. With the exception of this election, Gallup Polls have correctly predicted every presidential election since 1936.

a sample estimate. This error is called the **margin of error** of the estimate and is denoted by **MOE**. Because the margin of error is a prediction, we cannot guarantee that it is absolutely correct; that is, the probability that a prediction is correct might be 0.95, or it might be 0.75.

In the field of inferential statistics, the probability that a prediction is correct is referred to as the **level of confidence** of the prediction. For example, we

might say that we are 95% confident that the maximum error of an opinion poll is plus or minus 3 percentage points; that is, if 100 samples were analyzed, 95 of them would have proportions that differ from the true population proportion by an amount less than or equal to 0.03, and 5 of the samples would have an error greater than 0.03.

Assuming that sample proportions are normally distributed around the population proportion (as in Figure 4.126), we can use the z -distribution to determine the margin of error associated with a sample proportion. In general, the margin of error depends on the sample size and the level of confidence of the estimate.

MARGIN OF ERROR FORMULA

Given a sample size n , the **margin of error**, denoted by **MOE**, for a poll involving sample proportions is

$$\text{MOE} = \frac{z_{\alpha/2}}{2\sqrt{n}}$$

where α represents the level of confidence of the poll. That is, the probability is α that the sample proportion has an error of at most MOE.

EXAMPLE 2

FINDING THE MARGIN OF ERROR Assuming a 90% level of confidence, find the margin of error associated with each of the following:

- a. sample size $n = 275$ b. sample size $n = 750$

SOLUTION

- a. The margin of error depends on two things: the sample size and the level of confidence. For a 90% level of confidence, $\alpha = 0.90$. Hence, $\frac{\alpha}{2} = 0.45$, and $z_{\alpha/2} = z_{0.45} = 1.645$.

Substituting this value and $n = 275$ into the MOE formula, we have the following:

$$\begin{aligned} \text{MOE} &= \frac{z_{\alpha/2}}{2\sqrt{n}} \\ &= \frac{1.645}{2\sqrt{275}} \\ &= 0.049598616 \dots \\ &\approx 0.050 && \text{rounding off to three decimal places} \\ &= 5.0\% \end{aligned}$$

When we are polling a sample of 275 people, we can say that we are 90% confident that the maximum possible error in the sample proportion will be plus or minus 5.0 percentage points.

- b. For a 90% level of confidence, $\alpha = 0.90$, $\frac{\alpha}{2} = 0.45$, and $z_{\alpha/2} = z_{0.45} = 1.645$. Substituting this value and $n = 750$ into the MOE formula, we have the following:

$$\begin{aligned} \text{MOE} &= \frac{z_{\alpha/2}}{2\sqrt{n}} \\ &= \frac{1.645}{2\sqrt{750}} \\ &= 0.030033453 \dots \\ &\approx 0.030 && \text{rounding off to three decimal places} \\ &= 3.0\% \end{aligned}$$

When we are polling a sample of 750 people, we can say that we are 90% confident that the maximum possible error in the sample proportion will be plus or minus 3.0 percentage points.

If we compare the margins of error in parts (a) and (b) of Example 2, we notice that by increasing the sample size (from 275 to 750), the margin of error was reduced (from 5.0 to 3.0 percentage points). Intuitively, this should make sense; a larger sample gives a better estimate (has a smaller margin of error).

EXAMPLE 3

FINDING A SAMPLE PROPORTION AND MARGIN OF ERROR To obtain an estimate of the proportion of all Americans who think the president is doing a good job, a random sample of 500 Americans is surveyed, and 345 respond, “The president is doing a good job.”

- Determine the sample proportion of Americans who think the president is doing a good job.
- Assuming a 95% level of confidence, find the margin of error associated with the sample proportion.

SOLUTION

- $n = 500$ and $x = 345$. The sample proportion is

$$\frac{x}{n} = \frac{345}{500} = 0.69$$

Sixty-nine percent of the sample think the president is doing a good job.

- We must find MOE when $n = 500$ and $\alpha = 0.95$. Because $\alpha = 0.95$, $\frac{\alpha}{2} = 0.475$ and $z_{\alpha/2} = z_{0.475} = 1.96$.

Therefore,

$$\begin{aligned} \text{MOE} &= \frac{z_{\alpha/2}}{2\sqrt{n}} \\ &= \frac{1.96}{2\sqrt{500}} \\ &= 0.043826932 \dots \\ &\approx 0.044 && \text{rounding off to three decimal places} \\ &= 4.4\% \end{aligned}$$

The margin of error associated with the sample proportion is plus or minus 4.4 percentage points. We are 95% confident that 69% ($\pm 4.4\%$) of all Americans think the president is doing a good job. In other words, on the basis of our sample proportion ($\frac{x}{n}$) of 69%, we predict (with 95% certainty) that the true population proportion (P) is somewhere between 64.6% and 73.4%.

Example 2 indicated that increasing the sample size will decrease the margin of error, that is, larger samples give better estimates. If larger samples give better estimates, how large should a sample be? This question can be answered by manipulating the margin of error formula. That is, in its original form, we plug values of $z_{\alpha/2}$ (based upon the level of confidence) and n (the sample size) into the formula, and we calculate the margin of error. However, if we first solve the margin

of error formula for n , we will have a new version of the formula for determining how large a sample should be. We proceed as follows:

$$\begin{aligned} \text{MOE} &= \frac{z_{\alpha/2}}{2\sqrt{n}} && \text{the margin of error formula} \\ \sqrt{n} \cdot (\text{MOE}) &= \frac{z_{\alpha/2}}{2} && \text{multiplying each side by } \sqrt{n} \\ \sqrt{n} &= \frac{z_{\alpha/2}}{2(\text{MOE})} && \text{dividing each side by MOE} \\ n &= \left(\frac{z_{\alpha/2}}{2(\text{MOE})} \right)^2 && \text{squaring both sides} \end{aligned}$$

SAMPLE SIZE FORMULA

The required sample size n , necessary to have a desired margin of error of at most MOE, is given

$$n = \left(\frac{z_{\alpha/2}}{2(\text{MOE})} \right)^2$$

where $z_{\alpha/2}$ is determined by the given level of confidence.

EXAMPLE 4

FINDING A SAMPLE SIZE With a 98% level of confidence, how large should a sample be so that the margin of error is at most 4%?

SOLUTION

For a 98% level of confidence, $\alpha = 0.98$, $\frac{\alpha}{2} = 0.49$, and $z_{\alpha/2} = z_{0.49} = 2.33$. Substituting this value and $\text{MOE} = 0.04$ into the sample size formula, we have

$$\begin{aligned} n &= \left(\frac{z_{\alpha/2}}{2(\text{MOE})} \right)^2 \\ &= \left(\frac{2.33}{2(0.04)} \right)^2 \\ &= 848.265625 \dots \end{aligned}$$

Therefore, we should have roughly 848 and “one-quarter” people ($0.265625 \approx 0.25 = \frac{1}{4}$) in the sample. However, we cannot have part of a person, so it is customary to round this number *up* to the next highest whole number (include the whole person!); thus, the required sample size is $n = 849$ people.

EXAMPLE 5

FINDING THE MARGIN OF ERROR FOR DIFFERENT LEVELS OF CONFIDENCE The article shown on the next page was released by the Gallup Organization in October 2009.

- The poll states that 44% of the Americans questioned think that the laws covering firearm sales should be made more strict. Assuming a 95% level of confidence (the most commonly used level of confidence), find the margin of error associated with the survey.
- Assuming a 98% level of confidence, find the margin of error associated with the survey.

FEATURED IN THE NEWS

IN U.S., RECORD-LOW SUPPORT FOR STRICTER GUN LAWS

Princeton, NJ—Gallup finds a new low of 44% of Americans saying the laws covering firearm sales should be made more strict. That is down 5 points in the last year and 34 points from the high of 78% recorded the first time the question was asked, in 1990.

Today, Americans are as likely to say the laws governing gun sales should be kept as they are now (43%) as to say they should be made more strict. Until this year, Gallup had always found a significantly higher percentage advocating stricter laws. At the same time, 12% of Americans believe the laws should be

less strict, which is low in an absolute sense but ties the highest Gallup has measured for this response.

These results are based on Gallup's annual Crime Poll, conducted October 1–4, 2009.

The Poll also shows a new low in the percentage of Americans favoring a ban on handgun possession except by the police and other authorized persons, a question that dates back to 1959. Only 28% now favor such a ban. The high point in support for a handgun-possession ban was 60% in the initial measurement in 1959. Since then, less than a majority has been in favor, and support has been below 40% since December 1993.

The trends on the questions about gun-sale laws and a handgun-possession ban indicate that Americans' attitudes have moved toward being more pro-gun rights. But this is not due to a growth in personal gun ownership, which has held steady around 30% this decade, or to an increase in household gun ownership, which has been steady in the low 40% range since 2000.

Results are based on telephone interviews with 1,013 national adults, aged 18 and older, conducted October 1–4, 2009.

By Jeffrey M. Jones
Gallup News Service

SOLUTION

- a. We must find MOE when $n = 1,013$ and $\alpha = 0.95$. Because $\alpha = 0.95$, $\alpha/2 = 0.475$ and $z_{\alpha/2} = z_{0.475} = 1.96$.

Therefore,

$$\begin{aligned} \text{MOE} &= \frac{z_{\alpha/2}}{2\sqrt{n}} \\ &= \frac{1.96}{2\sqrt{1,013}} \\ &= 0.030790827 \dots \\ &\approx 0.031 \end{aligned}$$

rounding off to three decimal places

The margin of error associated with the survey is plus or minus 3.1%. We are 95% confident that $44\% \pm 3.1\%$ of all Americans think that the laws covering firearm sales should be made more strict.

- b. We must find MOE when $n = 1,013$ and $\alpha = 0.98$. Because $\alpha = 0.98$, $\alpha/2 = 0.49$ and $z_{\alpha/2} = z_{0.49} = 2.33$.

Therefore,

$$\begin{aligned} \text{MOE} &= \frac{z_{\alpha/2}}{2\sqrt{n}} \\ &= \frac{2.33}{2\sqrt{1,013}} \\ &= 0.036603381 \dots \\ &\approx 0.037 \end{aligned}$$

rounding off to three decimal places

The margin of error associated with the survey is plus or minus 3.7%. We are 98% confident that $44\% \pm 3.7\%$ of all Americans think that the laws covering firearm sales should be made more strict.

FEATURED IN THE NEWS

WHO SUPPORTS MARIJUANA LEGALIZATION? SUPPORT RISING; VARIES MOST BY AGE AND GENDER

Since the late 1960s, Gallup has periodically asked Americans whether the use of marijuana should be made legal in the United States. Although a majority of Americans have consistently opposed the idea of legalizing marijuana, public support has slowly increased over the years. In 1969, just 12% of Americans supported making marijuana legal, but by 1977, roughly one in four endorsed it. Support edged up to 31% in 2000, and now, about a third of Americans say marijuana should be legal.

Support for marijuana legalization varies greatest by gender and age. Overall, younger Americans (aged 18

to 29) are essentially divided, with 47% saying marijuana should be legal and 50% saying it should not be. Support for legalization is much lower among adults aged 30 to 64 (35%) and those aged 65 and older (22%). Men (39%) are somewhat more likely than women (30%) to support the legalization of marijuana in the country.

Americans residing in the western parts of the country are more likely than those living elsewhere to support the legalization of marijuana. These differences perhaps result from the fact that six Western states have, in various ways, already legalized marijuana for medicinal use. Overall, the data show that Westerners are divided about marijuana, with 47% saying it should be legal and 49% saying it should not be.

No more than a third of adults living in other parts of the country feel marijuana should be legal.

Support for legalizing marijuana is much lower among Republicans than it is among Democrats or independents. One in five Republicans (21%) say marijuana should be made legal in this country, while 37% of Democrats and 44% of independents share this view.

*Results are based on telephone interviews with 2,034 national adults, aged 18 and older, conducted Aug. 3–5, 2001, Nov. 10–12, 2003, and Oct. 21–23, 2005. For results based on the total sample of national adults, one can say with 95% confidence that the maximum margin of sampling error is ± 2 percentage points.

By Joseph Carroll,
Gallup Poll Assistant Editor

If we compare the margins of error found in parts (a) and (b) of Example 4, we notice that as the level of confidence went up (from 95% to 98%) the margin of error increased (from 3.1% to 3.7%). Intuitively, if we want to be more confident in our predictions, we should give our prediction more leeway (a larger margin of error).

When the results of the polls are printed in a newspaper, the sample size, level of confidence, margin of error, date of survey, and location of survey may be given as a footnote, as shown in the above article.

EXAMPLE 6

VERIFYING A STATED MARGIN OF ERROR Verify the margin of error stated in the article shown above.

SOLUTION

The footnote to the article states that for a sample size of 2,034 and a 95% level of confidence, the margin of error is $\pm 2\%$, that is, $\text{MOE} = 0.02$ for $n = 2,034$ and $\alpha = 0.95$. Because $\alpha = 0.95$, $\frac{\alpha}{2} = 0.475$ and $z_{\alpha/2} = z_{0.475} = 1.96$. Therefore,

$$\begin{aligned}\text{MOE} &= \frac{z_{\alpha/2}}{2\sqrt{n}} \\ &= \frac{1.96}{2\sqrt{2,034}} \\ &= 0.02195127 \dots \\ &\approx 0.02 \quad \text{rounding off to two decimal places}\end{aligned}$$

Therefore, the stated margin of error of $\pm 2\%$ is correct.

Text not available due to copyright restrictions

News articles do not always mention the level of confidence of a survey. However, if the sample size and margin of error are given, the level of confidence can be determined, as shown in Example 7.

EXAMPLE 7

FINDING THE LEVEL OF CONFIDENCE OF AN OPINION POLL Do you think that UFOs are real? The news article shown above presents the results of a Roper poll pertaining to this question. Find the level of confidence of this poll.

SOLUTION

We are given $n = 1,021$ and $\text{MOE} = 0.03$. To find α , the level of confidence of the poll, we must first find $z_{\alpha/2}$ and the area of the bodies, as shown in Figure 4.128.

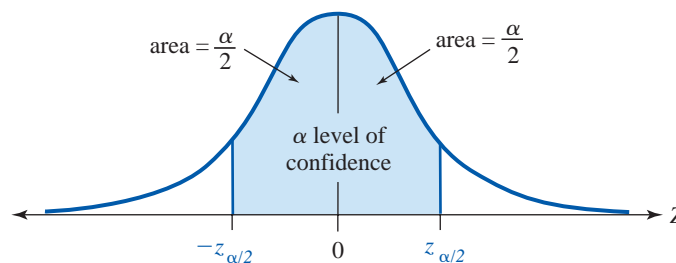


FIGURE 4.128 Level of confidence equals central region.

Substituting the given values into the MOE formula, we have

$$\begin{aligned}\text{MOE} &= \frac{z_{\alpha/2}}{2\sqrt{n}} \\ 0.03 &= \frac{z_{\alpha/2}}{2\sqrt{1,021}}\end{aligned}$$



TOPIC X RANDOM SAMPLING AND OPINION POLLS STATISTICS IN THE REAL WORLD

National opinion polls are pervasive in today's world; hardly a day goes by without the release of yet another glimpse at the American psyche. Topics ranging from presidential performance and political controversy to popular culture and alien abduction vie for our attention in the media: "A recent survey indicates that a majority of Americans. . . ." Where do these results come from? Are they accurate? Can they be taken seriously, or are they mere entertainment and speculation?

Although many people are intrigued by the "findings" of opinion polls, others are skeptical or unsure of the fundamental premise of statistical sampling. How can the opinions of a diverse population of nearly 300 million Americans be measured by a survey of a mere 1,000 to 1,500 individuals? Wouldn't a survey of 10,000 people give better, or more accurate, results than those obtained from the typical sample of about a thousand people? The answer is "no, not necessarily." Although intuition might dictate that a survey's reliability is driven by the size of the sample (bigger is better), in reality, the most important factor in obtaining reliable results is the *method* by which the sample was selected. That is, depending on how it was selected, a sample of 1,000 people can yield far better results than a sample of 10,000, 20,000, or even 50,000 people.

The basic premise in statistical sampling is that the views of a small portion

of a population *can* accurately represent the views of the entire population *if* the sample is selected properly, that is, if the sample is selected randomly. So what does it mean to say that a sample is selected "randomly"? The answer is simple: A sample is random if every member of the population has an *equal chance* of being selected. For example, suppose that a remote island has a population of 1,234 people and we wish to select a random sample of 30 inhabitants to interview. We could write each islander's name on a slip of paper, put all of the slips in a large box, shake up the box, close our eyes, and select 30 slips. Each person on the island would have an equal chance of being selected; consequently, we would be highly confident (say, 95%) that the views of the sample (plus or minus a margin or error) would accurately represent the views of the entire population.

The key to reliable sampling is the selection of a random sample; to select a random sample, each person in the population must have an equal chance of being selected. Realistically, how can this be accomplished with such a vast population of people in the United States or elsewhere? Years ago, the most accurate polls (especially Gallup polls) were based on data gathered from knocking on doors and conducting face-to-face interviews. However, in today's world, almost every American adult has a telephone, and tele-

phone surveys have replaced the door-to-door surveys of the past. Although a poll might state that the target population is "all Americans aged 18 and over," it really means "all Americans aged 18 and over who have an accessible telephone number." Who will be excluded from this population? Typically, active members of the military forces and people in prisons, institutions, or hospitals are excluded from the sampling frame of today's opinion polls.

How can a polling organization obtain a list of the telephone numbers of all Americans aged 18 and over? Typically, no such list exists. Telephone directories are not that useful because of the large number of unlisted telephone numbers. However, using high-speed computers and a procedure known as random digit dialing, polling organizations are able to create a list of all possible phone numbers and thus are able to select a random sample. Finally, each telephone number in the sample is called, and "an American aged 18 and over" (or whatever group is being targeted) is interviewed. If no one answers the phone or the appropriate person is not at home, the polling organization makes every effort to establish contact at a later time. This ensures that the random sampling process is accurately applied and therefore that the results are true to the stated level of confidence with an acceptable margin of error.

$$z_{\alpha/2} = 0.03(2\sqrt{1,021})$$

multiplying each side by $2\sqrt{1,021}$

$$z_{\alpha/2} = 1.917185437 \dots$$

$$z_{\alpha/2} \approx 1.92$$

rounding off to two decimal places

Using the body table in Appendix F, we can find the area under the bell curve between $z = 0$ and $z = 1.92$; that is, $p(0 < z < 1.92) = 0.4726$.

Therefore, $\frac{\alpha}{2} = 0.4726$, and multiplying by 2, we have $\alpha = 0.9452$. Thus, the level of confidence is $\alpha = 0.9452$ (or 95%), as shown in Figure 4.129.

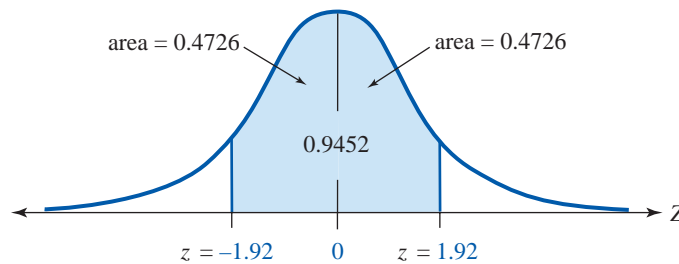


FIGURE 4.129 Finding the level of confidence.

4.5 EXERCISES

In Exercises 1–4, use the body table in Appendix F to find the specified z -number.

1. a. $z_{0.2517}$ b. $z_{0.1217}$ c. $z_{0.4177}$ d. $z_{0.4960}$
- ▶ 2. a. $z_{0.0199}$ b. $z_{0.2422}$ c. $z_{0.4474}$ d. $z_{0.4936}$
- ▶ 3. a. $z_{0.4250}$ b. $z_{0.4000}$ c. $z_{0.3750}$ d. $z_{0.4950}$
- ▶ 4. a. $z_{0.4350}$ b. $z_{0.4100}$ c. $z_{0.2750}$ d. $z_{0.4958}$
- ▶ 5. Find the z -number associated with a 92% level of confidence.
- ▶ 6. Find the z -number associated with a 97% level of confidence.
7. Find the z -number associated with a 75% level of confidence.
8. Find the z -number associated with an 85% level of confidence.

In Exercises 9–22, round off your answers (sample proportions and margins of error) to three decimal places (a tenth of a percent).

9. The Gallup Poll in Example 6 states that one-third (33%) of respondents support general legalization of marijuana. For each of the following levels of confidence, find the margin of error associated with the sample.
 - a. a 90% level of confidence
 - b. a 98% level of confidence
- ▶ 10. The Roper poll in Example 7 states that 56% of the Americans questioned think that UFOs are real. For each of the following levels of confidence, find the margin of error associated with the sample.
 - a. an 80% level of confidence
 - b. a 98% level of confidence
- ▶ 11. A survey asked, “How important is it to you to buy products that are made in America?”

Of the 600 Americans surveyed, 450 responded, “It is important.” For each of the following levels of confidence, find the sample proportion and the margin of error associated with the poll.

 - a. a 90% level of confidence
 - b. a 95% level of confidence
- ▶ 12. In the survey in Exercise 11, 150 of the 600 Americans surveyed responded, “It is not important.” For each of the following levels of confidence, find the sample proportion and the margin of error associated with the poll.
 - a. an 85% level of confidence
 - b. a 98% level of confidence
13. A survey asked, “Have you ever bought a lottery ticket?” Of the 2,710 Americans surveyed, 2,141 said yes, and 569 said no.*
 - a. Determine the sample proportion of Americans who have purchased a lottery ticket.
 - b. Determine the sample proportion of Americans who have not purchased a lottery ticket.
 - c. With a 90% level of confidence, find the margin of error associated with the sample proportions.
- ▶ 14. A survey asked, “Which leg do you put into your trousers first?” Of the 2,710 Americans surveyed, 1,138 said left, and 1,572 said right.*
 - a. Determine the sample proportion of Americans who put their left leg into their trousers first.
 - b. Determine the sample proportion of Americans who put their right leg into their trousers first.
 - c. With a 90% level of confidence, find the margin of error associated with the sample proportions.

* Data from Poretz and Sinrod, *The First Really Important Survey of American Habits* (Los Angeles: Price Stern Sloan Publishing, 1989).

15. A survey asked, “Do you prefer showering or bathing?” Of the 1,220 American men surveyed, 1,049 preferred showering, and 171 preferred bathing. In contrast, 1,043 of the 1,490 American women surveyed preferred showering, and 447 preferred bathing.*
- Determine the sample proportion of American men who prefer showering.
 - Determine the sample proportion of American women who prefer showering.
 - With a 95% level of confidence, find the margin of error associated with the sample proportions.
16. A survey asked, “Do you like the way you look in the nude?” Of the 1,220 American men surveyed, 830 said yes, and 390 said no. In contrast, 328 of the 1,490 American women surveyed said yes, and 1,162 said no.*
- Determine the sample proportion of American men who like the way they look in the nude.
 - Determine the sample proportion of American women who like the way they look in the nude.
 - With a 95% level of confidence, find the margin of error associated with the sample proportions.
- Exercises 17–20 are based on a survey (published in April 2000) of 129,593 students in grades 6–12 conducted by USA WEEKEND magazine.*
17. When asked, “Do you, personally, feel safe from violence in school?” 92,011 said yes, and 37,582 said no.
- Determine the sample proportion of students who said yes.
 - Determine the sample proportion of students who said no.
 - With a 95% level of confidence, find the margin of error associated with the sample proportions.
- ▶ 18. When asked, “Do kids regularly carry weapons in your school?” 14,255 said yes, and 115,338 said no.
- Determine the sample proportion of students who said yes.
 - Determine the sample proportion of students who said no.
 - With a 95% level of confidence, find the margin of error associated with the sample proportions.
19. When asked, “Is there a gun in your home?” 58,317 said yes, 60,908 said no, and 10,368 said they did not know.
- Determine the sample proportion of students who said yes.
 - Determine the sample proportion of students who said no.
 - Determine the sample proportion of students who said they did not know.
 - With a 90% level of confidence, find the margin of error associated with the sample proportions.
20. When asked, “How likely do you think it is that a major violent incident could occur at your school?” 18,143 said very likely, 64,797 said somewhat likely, and 46,653 said not likely at all.
- Determine the sample proportion of students who said “very likely.”
 - Determine the sample proportion of students who said “somewhat likely.”
 - Determine the sample proportion of students who said “not likely at all.”
 - With a 90% level of confidence, find the margin of error associated with the sample proportions.
21. A survey asked, “Can you imagine a situation in which you might become homeless?” Of the 2,503 Americans surveyed, 902 said yes.*
- Determine the sample proportion of Americans who can imagine a situation in which they might become homeless.
 - With a 90% level of confidence, find the margin of error associated with the sample proportion.
 - With a 98% level of confidence, find the margin of error associated with the sample proportion.
 - How does your answer to part (c) compare to your answer to part (b)? Why?
- ▶ 22. A survey asked, “Do you think that homeless people are responsible for the situation they are in?” Of the 2,503 Americans surveyed, 1,402 said no.*
- Determine the sample proportion of Americans who think that homeless people are not responsible for the situation they are in.
 - With an 80% level of confidence, find the margin of error associated with the sample proportion.
 - With a 95% level of confidence, find the margin of error associated with the sample proportion.
 - How does your answer to part (c) compare to your answer to part (b)? Why?
- ▶ 23. A sample consisting of 430 men and 765 women was asked various questions pertaining to international affairs. With a 95% level of confidence, find the margin of error associated with the following samples.
- the male sample
 - the female sample
 - the combined sample
- ▶ 24. A sample consisting of 942 men and 503 women was asked various questions pertaining to the nation’s economy. For a 95% level of confidence, find the margin of error associated with the following samples.

* Data from Poretz and Sinrod, *The First Really Important Survey of American Habits* (Los Angeles: Price Stern Sloan Publishing, 1989).

* Data from Mark Clements, “What Americans Say about the Homeless,” *Parade Magazine*, Jan. 9, 1994: 4–6.

- a. the male sample
 - b. the female sample
 - c. the combined sample
- ▶ 25. A poll pertaining to environmental concerns had the following footnote: “Based on a sample of 1,763 adults, the margin of error is plus or minus 2.5 percentage points.” Find the level of confidence of the poll.

HINT: See Example 7.

26. A poll pertaining to educational goals had the following footnote: “Based on a sample of 2,014 teenagers, the margin of error is plus or minus 2 percentage points.” Find the level of confidence of the poll.

HINT: See Example 7.

27. A recent poll pertaining to educational reforms involved 640 men and 820 women. The margin of error for the combined sample is 2.6%. Find the level of confidence for the entire poll.

HINT: See Example 7.

- ▶ 28. A recent poll pertaining to educational reforms involved 640 men and 820 women. The margin of error is 3.9% for the male sample and 3.4% for the female sample. Find the level of confidence for the male portion of the poll and for the female portion of the poll.

HINT: See Example 7.

In Exercises 29–32, you are planning a survey for which the findings are to have the specified level of confidence. How large should your sample be so that the margin of error is at most the specified amount?

29. 95% level of confidence
- a. margin of error = 3%
 - b. margin of error = 2%
 - c. margin of error = 1%
 - d. Comparing your answers to parts (a)–(c), what conclusion can be made regarding the margin of error and the sample size?
- ▶ 30. 96% level of confidence
- a. margin of error = 3%
 - b. margin of error = 2%
 - c. margin of error = 1%
 - d. Comparing your answers to parts (a)–(c), what conclusion can be made regarding the margin of error and the sample size?
31. 98% level of confidence
- a. margin of error = 3%
 - b. margin of error = 2%
 - c. margin of error = 1%
 - d. Comparing your answers to parts (a)–(c), what conclusion can be made regarding the margin of error and the sample size?

- ▶ 32. 99% level of confidence
- a. margin of error = 3%
 - b. margin of error = 2%
 - c. margin of error = 1%
 - d. Comparing your answers to parts (a)–(c), what conclusion can be made regarding the margin of error and the sample size?



Answer the following questions using complete sentences and your own words.

• CONCEPT QUESTIONS

33. What is a sample proportion? How is it calculated?
34. What is a margin of error? How is it calculated?
35. If the sample size is increased in a survey, would you expect the margin of error to increase or decrease? Why?
36. What is a random sample?
37. What is random digit dialing (RDD)?

• HISTORY QUESTIONS

38. Who founded the American Institute of Public Opinion? When? In what two academic fields did this person receive degrees?



WEB PROJECTS

39. What are the nation’s current opinions concerning major issues in the headlines? Pick a current issue (such as abortion, same-sex marriage, gun control, war, or the president’s approval rating) and find a recent survey regarding the issue. Summarize the results of the survey. Be sure to include the polling organization, date(s) of the survey, sample size, level of confidence, margin of error, and any pertinent information.
40. What is random digit dialing (RDD), and how is it used? Visit the web site of a national polling or news organization, and conduct a search of its FAQs regarding the organization’s methods of sampling and the use of RDD. Write a report in which you summarize the use of RDD by the polling or news organization. Some useful links for this web project are listed on the text web site:
www.cengage.com/math/johnson

• PROJECTS

41. The purpose of this project is to conduct an opinion poll. Select a topic that is relevant to you and/or your community. Create a multiple-choice question to gather people’s opinions concerning this issue.

For example:

If you could vote today, how would you vote on Proposition X?

i. support ii. oppose iii. undecided

- a. Ask fifty people your question, and record their responses. Calculate the sample proportion for each category of response. Use a 95% level of confidence, and calculate the margin of error for your survey.
- b. Ask 100 people your question, and record their responses. Calculate the sample proportion for each category of response. Use a 95% level of confidence, and calculate the margin of error for your survey.
- c. How does the margin of error in part (b) compare to the margin of error in part (a)? Explain.

4.6 Linear Regression

OBJECTIVES

- Find the line of best fit
- Use linear regression to make a prediction
- Calculate the coefficient of linear correlation

When x and y are variables and m and b are constants, the equation $y = mx + b$ has infinitely many solutions of the form (x, y) . A specific ordered pair (x_1, y_1) is a solution of the equation if $y_1 = mx_1 + b$. Because every solution of the given equation lies on a straight line, we say that x and y are *linearly related*.

If we are given two ordered pairs (x_1, y_1) and (x_2, y_2) , we should be able to “work backwards” and find the equation of the line passing through them; assuming that x and y are linearly related, we can easily find the equation of the line passing through the points (x_1, y_1) and (x_2, y_2) . The process of finding the equation of a line passing through given points is known as **linear regression**; the equation thus found is called the **mathematical model** of the linear relationship. Once the model has been constructed, it can be used to make predictions concerning the values of x and y .

EXAMPLE 1

USING A LINEAR EQUATION TO MAKE A PREDICTION Charlie is planning a family reunion and wants to place an order for custom T-shirts from Prints Alive (the local silk-screen printer) to commemorate the occasion. He has ordered shirts from Prints Alive on two previous occasions; on one occasion, he paid \$164 for twenty-four shirts; on another, he paid \$449 for eighty-four. Assuming a linear relationship between the cost of T-shirts and the number ordered, predict the cost of ordering 100 shirts.

SOLUTION

Letting x = the number of shirts ordered and y = the total cost of the shirts, the given data can be expressed as two ordered pairs: $(x_1, y_1) = (24, 164)$ and $(x_2, y_2) = (84, 449)$. We must find $y = mx + b$, the equation of the line passing through the two points.

First, we find m , the slope:

$$\begin{aligned} m &= \frac{y_2 - y_1}{x_2 - x_1} \\ &= \frac{449 - 164}{84 - 24} \\ &= \frac{285}{60} \\ &= 4.75 \end{aligned}$$

Now we use one of the ordered pairs to find b , the y -intercept. Either point will work; we will use $(x_1, y_1) = (24, 164)$.

The slope-intercept form of a line is $y = mx + b$. Solving for b , we obtain

$$\begin{aligned} b &= y - mx \\ &= 164 - 4.75(24) \\ &= 164 - 114 \\ &= 50 \end{aligned}$$

Therefore, the equation of the line is $y = 4.75x + 50$. We use this linear model to predict the cost of ordering $x = 100$ T-shirts.

$$\begin{aligned} y &= 4.75x + 50 \\ &= 4.75(100) + 50 \\ &= 475 + 50 \\ &= 525 \end{aligned}$$

We predict that it will cost \$525 to order 100 T-shirts.

Linear Trends and Line of Best Fit

Example 1 illustrates the fact that two points determine a unique line. To find the equation of the line, we must find the slope and the y -intercept. If we are given more than two points, the points might not be collinear. In collecting real-world data, this is usually the case. However, after the scatter of points is plotted on an x - y coordinate system, it may appear that the points “almost” fit on a line. If a sample of ordered pairs tend to “go in the same general direction,” we say that they exhibit a **linear trend**. See Figure 4.130.

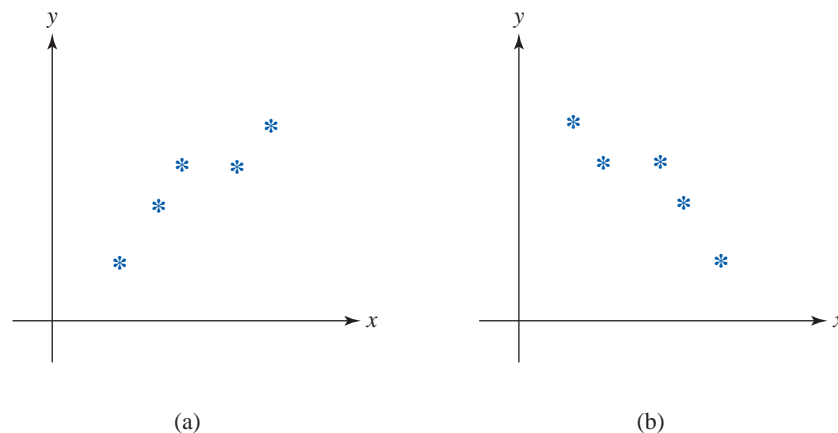


FIGURE 4.130 Ordered pairs that exhibit a linear trend.

When a scatter of points exhibits a linear trend, we construct the line that best approximates the trend. This line is called the **line of best fit** and is denoted by $\hat{y} = mx + b$. The “hat” over the y indicates that the calculated value of y is a prediction based on linear regression. See Figure 4.131.

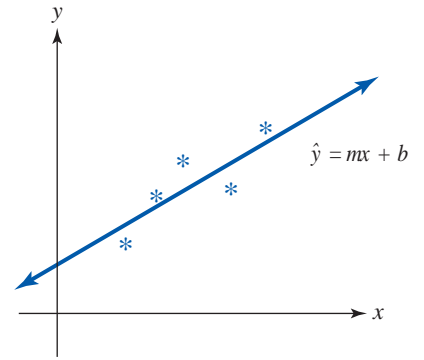


FIGURE 4.131 The line of best fit $\hat{y} = mx + b$.

To calculate the slope and y -intercept of the line of best fit, mathematicians have developed formulas based on the method of least squares. (See the Historical Note on Carl Gauss in Section 4.4.)

Recall that the symbol Σ means “sum.” Therefore, Σx represents the sum of the x -coordinates of the points, and Σy represents the sum of the y -coordinates. To find Σxy , multiply the x - and y -coordinates of each point and sum the results.

LINE OF BEST FIT

Given a sample of n ordered pairs $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, the **line of best fit** (the line that best represents the data) is denoted by $\hat{y} = mx + b$, where the slope m and y -intercept b are given by

$$m = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2} \quad \text{and} \quad b = \bar{y} - m\bar{x}$$

\bar{x} and \bar{y} denote the means of the x - and y -coordinates, respectively.

EXAMPLE 2

FINDING AND GRAPHING THE LINE OF BEST FIT We are given the ordered pairs $(5, 14), (9, 17), (12, 16), (14, 18),$ and $(17, 23)$.

- Find the equation of the line of best fit.
- Plot the given data and sketch the graph of the line of best fit on the same coordinate system.

SOLUTION

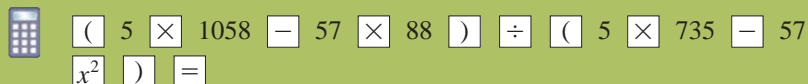
- Organize the data in a table and compute the appropriate sums, as shown in Figure 4.132.


(x, y)	x	x^2	y	xy
$(5, 14)$	5	25	14	$5 \cdot 14 = 70$
$(9, 17)$	9	81	17	$9 \cdot 17 = 153$
$(12, 16)$	12	144	16	$12 \cdot 16 = 192$
$(14, 18)$	14	196	18	$14 \cdot 18 = 252$
$(17, 23)$	17	289	23	$17 \cdot 23 = 391$
$n = 5$ ordered pairs	$\Sigma x = 57$	$\Sigma x^2 = 735$	$\Sigma y = 88$	$\Sigma xy = 1,058$

FIGURE 4.132 Table of sums.

First, we find the slope:

$$\begin{aligned} m &= \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \\ &= \frac{5(1,058) - (57)(88)}{5(735) - (57)^2} \\ &= 0.643192488 \dots \end{aligned}$$



 For graphing calculators, see the instructions on page 319.

Once m has been calculated, we store it in the memory of our calculator. We will need it to calculate b , the y -intercept.

$$\begin{aligned} b &= \bar{y} - m\bar{x} \\ &= \left(\frac{88}{5}\right) - 0.643192488 \left(\frac{57}{5}\right) \\ &= 10.26760564 \dots \end{aligned}$$



Therefore, the line of best fit, $\hat{y} = mx + b$, is

$$\hat{y} = 0.643192488x + 10.26760564$$

Rounding off to one decimal place, we have

$$\hat{y} = 0.6x + 10.3$$

- b.** To graph the line, we need to plot two points. One point is the y -intercept $(0, b) = (0, 10.3)$. To find another point, we pick an appropriate value for x —say, $x = 18$ —and calculate \hat{y} :

$$\begin{aligned} \hat{y} &= 0.6x + 10.3 \\ &= 0.6(18) + 10.3 \\ &= 21.1 \end{aligned}$$

Therefore, the point $(x, \hat{y}) = (18, 21.1)$ is on the line of best fit.

Plotting $(0, 10.3)$ and $(18, 21.1)$, we construct the line of best fit. It is customary to use asterisks (*) to plot the given ordered pairs, as shown in Figure 4.133.

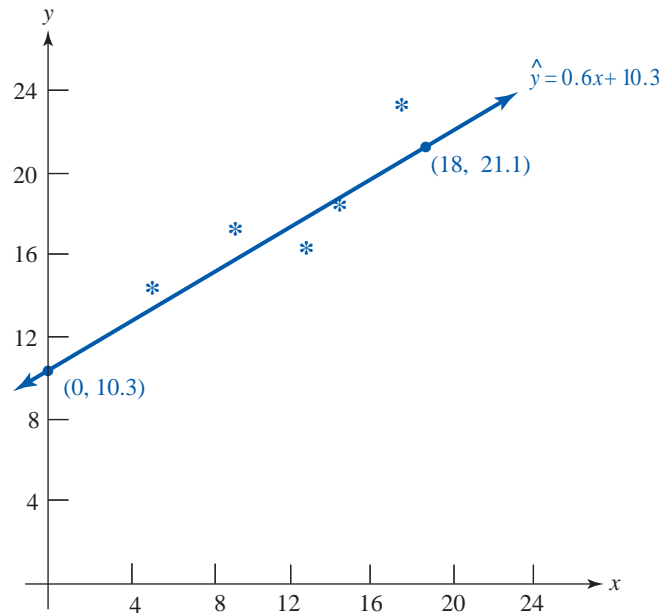


FIGURE 4.133 Scatter diagram and line of best fit.

Coefficient of Linear Correlation

Given a sample of n ordered pairs, we can always find the line of best fit. Does the line accurately portray the data? Will the line give accurate predictions? To answer these questions, we must consider the relative strength of the linear trend exhibited by the given data. If the given points are close to the line of best fit, there is a strong linear relation between x and y ; the line will generate good predictions. If the given points are widely scattered about the line of best fit, there is a weak linear relation, and predictions based on it are probably not reliable.

One way to measure the strength of a linear trend is to calculate the **coefficient of linear correlation**, denoted by r . The formula for calculating r is shown in the following box.

COEFFICIENT OF LINEAR CORRELATION

Given a sample of n ordered pairs, $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$

The **coefficient of linear correlation**, denoted by r , is given by

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \sqrt{n(\sum y^2) - (\sum y)^2}}$$

The calculated value of r is always between -1 and 1 , inclusive; that is, $-1 \leq r \leq 1$. If the given ordered pairs lie perfectly on a line whose slope is *positive*, then the calculated value of r will equal 1 (think 100% perfect with positive slope). In this case, both variables have the same behavior: As one increases (or decreases), so will the other. On the other hand, if the data points fall perfectly on a line whose slope is *negative*, the calculated value of r will equal -1 (think 100% perfect with negative slope). In this case, the variables have opposite behavior: As one increases, the other decreases, and vice versa. See Figure 4.134.

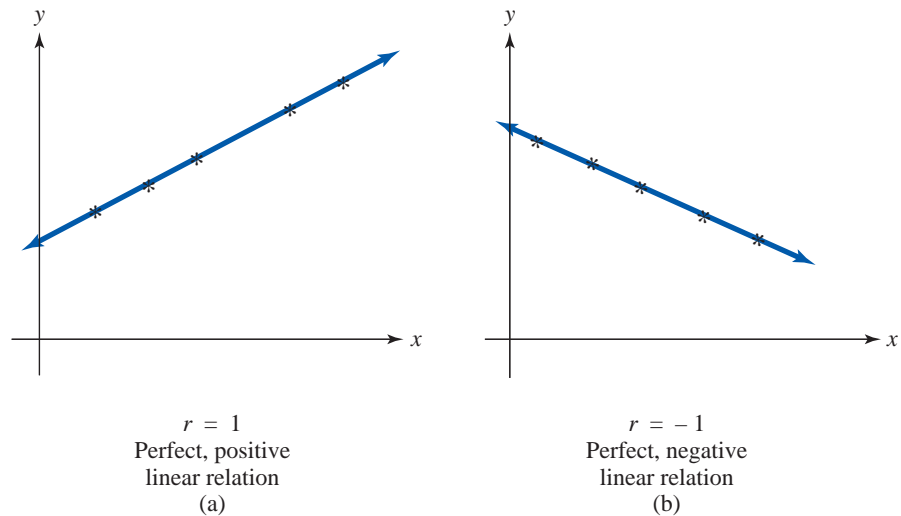


FIGURE 4.134

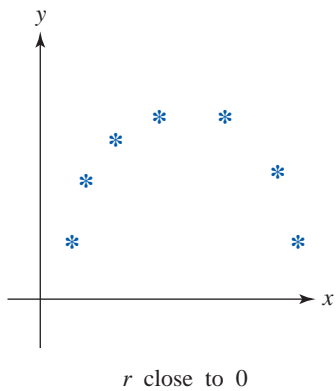


FIGURE 4.135

If the value of r is close to 0, there is little or no *linear* relation between the variables. This does not mean that the variables are not related. It merely means that no *linear* relation exists; the variables might be related in some nonlinear fashion, as shown in Figure 4.135.

In summary, the closer r is to 1 or -1 , the stronger is the linear relation between x and y ; the line of best fit will generate reliable predictions. The closer r is to 0, the weaker is the linear relation; the line of best fit will generate unreliable predictions. If r is positive, the variables have a direct relationship (as one increases, so does the other); if r is negative, the variables have an inverse relationship (as one increases, the other decreases).

EXAMPLE 3

CALCULATING THE COEFFICIENT OF LINEAR CORRELATION

Calculate the coefficient of linear correlation for the ordered pairs given in Example 2.

SOLUTION

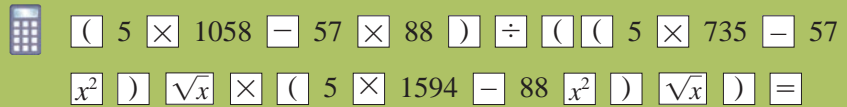
The ordered pairs are (5, 14), (9, 17), (12, 16), (14, 18), and (17, 23). We add a y^2 column to the table in Example 2, as shown in Figure 4.136.

(x, y)	x	x^2	y	y^2	xy
(5, 14)	5	25	14	196	$5 \cdot 14 = 70$
(9, 17)	9	81	17	289	$9 \cdot 17 = 153$
(12, 16)	12	144	16	256	$12 \cdot 16 = 192$
(14, 18)	14	196	18	324	$14 \cdot 18 = 252$
(17, 23)	17	289	23	529	$17 \cdot 23 = 391$
$n = 5$ ordered pairs	$\sum x = 57$	$\sum x^2 = 735$	$\sum y = 88$	$\sum y^2 = 1,594$	$\sum xy = 1,058$

FIGURE 4.136 Table of sums.

Now use the formula to calculate r :

$$\begin{aligned} r &= \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \sqrt{n(\sum y^2) - (\sum y)^2}} \\ &= \frac{5(1,058) - (57)(88)}{\sqrt{5(735) - (57)^2} \sqrt{5(1,594) - (88)^2}} \\ &= 0.883062705 \dots \end{aligned}$$



The calculator interface shows the following sequence of operations: $(5 \times 1058 - 57 \times 88) \div ((5 \times 735 - 57^2) \times (5 \times 1594 - 88^2))^{0.5}$.

The coefficient of linear correlation is reasonably close to 1, so the line of best fit will generate reasonably reliable predictions. (Notice that the data points in Figure 4.133 are fairly close to the line of best fit.)

EXAMPLE 4

USING LINEAR CORRELATION AND REGRESSION TO MAKE PREDICTIONS

Unemployment and personal income are undoubtedly related; we would assume that as the national unemployment rate increases, total personal income would decrease. Figure 4.137 gives the unemployment rate and the total personal income for the United States for various years.

- Use linear regression to predict the total personal income of the United States if the unemployment rate is 5.0%.
- Use linear regression to predict the unemployment rate if the total personal income of the United States is \$10 billion.
- Are the predictions in parts (a) and (b) reliable? Why or why not?

Year	Unemployment Rate (percent)	Total Personal Income (billions)
1975	8.5	\$1.3
1980	7.1	2.3
1985	7.2	3.4
1990	5.6	4.9
1995	5.6	6.1
2000	4.0	8.4
2006	4.6	11.0

FIGURE 4.137 Sources: Bureau of Labor Statistics, U.S. Department of Labor; Bureau of Economic Analysis, U.S. Department of Commerce.

SOLUTION

- Letting x = unemployment rate and y = total personal income, we have $n = 7$ ordered pairs, as shown in Figure 4.138.

Using a calculator, we find the following sums:

$$\sum x^2 = 274.38 \quad \sum y^2 = 271.32 \quad \sum xy = 197.66$$

x (percent)	y (billions)
8.5	\$1.3
7.1	2.3
7.2	3.4
5.6	4.9
5.6	6.1
4.0	8.4
4.6	11.0
$\Sigma x = 42.6$	$\Sigma y = 37.4$

FIGURE 4.138

x = unemployment rate,
 y = total personal income.

First we find the slope:

$$\begin{aligned} m &= \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2} \\ &= \frac{7(197.66) - (42.6)(37.4)}{7(274.38) - (42.6)^2} \\ &= -1.979414542 \dots \end{aligned}$$

Now we calculate b , the y -intercept:

$$\begin{aligned} b &= \bar{y} - m\bar{x} \\ &= \left(\frac{37.4}{7}\right) - (-1.979414542)\left(\frac{42.6}{7}\right) \\ &= 17.3890085 \dots \end{aligned}$$

Therefore, the line of best fit, $\hat{y} = mx + b$, is

$$\hat{y} = -1.979414542x + 17.3890085$$

Rounding off to two decimal places (one more than the data), we have

$$\hat{y} = -1.98x + 17.39$$

Now, substituting $x = 5.0$ (5% unemployment) into the equation of the line of best fit, we have

$$\begin{aligned} \hat{y} &= -1.98x + 17.39 \\ &= -1.98(5.0) + 17.39 \\ &= -9.9 + 17.39 \\ &= 7.49 \end{aligned}$$

If the unemployment rate is 5.0%, we predict that the total personal income of the United States will be approximately \$7.49 billion.

- b. To predict the unemployment rate when the total personal income is \$10 billion, we let $y = 10$, substitute into \hat{y} , and solve for x :

$$\begin{aligned} 10 &= -1.98x + 17.39 \\ 1.98x + 10 &= 17.39 && \text{adding } 1.98x \text{ to both sides} \\ 1.98x &= 17.39 - 10 && \text{subtracting } 10 \text{ from both sides} \\ 1.98x &= 7.39 \\ x &= \frac{7.39}{1.98} && \text{dividing by } 1.98 \\ &= 3.7323232 \dots \end{aligned}$$

We predict that the unemployment rate will be approximately 3.7% when the total personal income is \$10 billion.

- c. To investigate the reliability of our predictions (the strength of the linear trend), we must calculate the coefficient of linear correlation:

$$\begin{aligned} r &= \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2} \sqrt{n(\Sigma y^2) - (\Sigma y)^2}} \\ r &= \frac{7(197.66) - (42.6)(37.4)}{\sqrt{7(274.38) - (42.6)^2} \sqrt{7(271.32) - (37.4)^2}} \\ &= -0.9105239486 \dots \end{aligned}$$

Because r is close to -1 , we conclude that our predictions are very reliable; the linear relationship between x and y is high. Furthermore, since r is negative, we know that y (total personal income) decreases as x (unemployment rate) increases.

4.6 EXERCISES

- ▶ 1. A set of $n = 6$ ordered pairs has the following sums:
- $$\Sigma x = 64 \quad \Sigma x^2 = 814 \quad \Sigma y = 85$$
- $$\Sigma y^2 = 1,351 \quad \Sigma xy = 1,039$$
- Find the line of best fit.
 - Predict the value of y when $x = 11$.
 - Predict the value of x when $y = 19$.
 - Find the coefficient of linear correlation.
 - Are the predictions in parts (b) and (c) reliable? Why or why not?
- ▶ 2. A set of $n = 8$ ordered pairs has the following sums:
- $$\Sigma x = 111 \quad \Sigma x^2 = 1,869 \quad \Sigma y = 618$$
- $$\Sigma y^2 = 49,374 \quad \Sigma xy = 7,860$$
- Find the line of best fit.
 - Predict the value of y when $x = 8$.
 - Predict the value of x when $y = 70$.
 - Find the coefficient of linear correlation.
 - Are the predictions in parts (b) and (c) reliable? Why or why not?
3. A set of $n = 5$ ordered pairs has the following sums:
- $$\Sigma x = 37 \quad \Sigma x^2 = 299 \quad \Sigma y = 38$$
- $$\Sigma y^2 = 310 \quad \Sigma xy = 279$$
- Find the line of best fit.
 - Predict the value of y when $x = 5$.
 - Predict the value of x when $y = 7$.
 - Find the coefficient of linear correlation.
 - Are the predictions in parts (b) and (c) reliable? Why or why not?
- ▶ 4. Given the ordered pairs (4, 40), (6, 37), (8, 34), and (10, 31):
- Find and interpret the coefficient of linear correlation.
 - Find the line of best fit.
 - Plot the given ordered pairs and sketch the graph of the line of best fit on the same coordinate system.
- ▶ 5. Given the ordered pairs (5, 5), (7, 10), (8, 11), (10, 15), and (13, 16):
- Plot the ordered pairs. Do the ordered pairs exhibit a linear trend?
 - Find the line of best fit.
 - Predict the value of y when $x = 9$.
 - Plot the given ordered pairs and sketch the graph of the line of best fit on the same coordinate system.
 - Find the coefficient of linear correlation.
 - Is the prediction in part (c) reliable? Why or why not?
- ▶ 6. Given the ordered pairs (5, 20), (6, 15), (10, 14), (12, 15), and (13, 10):
- Plot the ordered pairs. Do the ordered pairs exhibit a linear trend?
 - Find the line of best fit.
 - Predict the value of y when $x = 8$.
 - Plot the given ordered pairs and sketch the graph of the line of best fit on the same coordinate system.
 - Find the coefficient of linear correlation.
 - Is the prediction in part (c) reliable? Why or why not?
- ▶ 7. Given the ordered pairs (2, 6), (3, 12), (6, 15), (7, 4), (10, 6), and (11, 12):
- Plot the ordered pairs. Do the ordered pairs exhibit a linear trend?
 - Find the line of best fit.
 - Predict the value of y when $x = 8$.
 - Plot the given ordered pairs and sketch the graph of the line of best fit on the same coordinate system.
 - Find the coefficient of linear correlation.
 - Is the prediction in part (c) reliable? Why or why not?
8. The unemployment rate and the amount of emergency food assistance (food made available to hunger relief organizations such as food banks and soup kitchens) provided by the federal government in the United States in certain years are given in Figure 4.139.

Year	Unemployment Rate	Emergency Food Assistance (in millions)
2000	4.0%	\$225
2003	6.0	456
2004	5.5	420
2005	5.1	373
2006	4.6	300
2007	4.6	256

FIGURE 4.139 Unemployment rates and emergency food assistance. Sources: Bureau of Labor Statistics (U.S. Department of Labor); Food and Nutrition Service (U.S. Department of Agriculture).

- Letting $x =$ the unemployment rate and $y =$ the amount of emergency food assistance, plot the data. Do the data exhibit a linear trend?
- Find the line of best fit.

▶ Selected exercises available online at www.webassign.net/brookscole

- c. Predict the amount of emergency food assistance when the unemployment rate is 5.0%.
 - d. Predict the unemployment rate when the amount of emergency food assistance is \$350 million.
 - e. Find the coefficient of correlation.
 - f. Are the predictions in parts (a) and (b) reliable? Why or why not?
9. The average hourly earnings and the average tuition at public four-year institutions of higher education in the United States for 1997–2002 are given in Figure 4.140.

Year	Average Hourly Earnings	Average Tuition at Four-Year Institutions
1997	\$12.49	\$3,110
1998	13.00	3,229
1999	13.47	3,349
2000	14.00	3,501
2001	14.53	3,735
2002	14.95	4,059

FIGURE 4.140 Average hourly earnings and average tuition, 1997–2002. *Source:* Bureau of Labor Statistics and National Center for Education Statistics.

- a. Letting x = the average hourly earnings and y = the average tuition, plot the data. Do the data exhibit a linear trend?
 - b. Find the line of best fit.
 - c. Predict the average tuition when the average hourly earning is \$14.75.
 - d. Predict the average hourly earning when the average tuition is \$4,000.
 - e. Find the coefficient of correlation.
 - f. Are the predictions in parts (a) and (b) reliable? Why or why not?
- ▶ 10. The number of domestic and imported retail car sales (in hundreds of thousands) in the United States for 1999–2007 are given in Figure 4.141.
- a. Letting x = the number of domestic car sales and y = the number of imported car sales, plot the data. Do the data exhibit a linear trend?
 - b. Find the line of best fit.
 - c. Predict the number of imported car sales when there are 5,600,000 domestic car sales.
 - d. Predict the number of domestic car sales when there are 2,200,000 imported car sales.
 - e. Find the coefficient of correlation.
 - f. Are the predictions in parts (a) and (b) reliable? Why or why not?

Year	Domestic Car Sales	Imported Car Sales
1999	69.8	17.2
2000	68.3	20.2
2001	63.2	21.0
2002	58.8	22.3
2003	55.3	20.8
2004	53.6	21.5
2005	54.8	21.9
2006	54.4	23.4
2007	52.5	23.7

FIGURE 4.141 Domestic and imported retail car sales (hundred thousands), 1999–2007. *Source:* Ward's Commission.

- ▶ 11. The numbers of marriages and divorces (in millions) in the United States are given in Figure 4.142.

Year	1965	1970	1975	1980	1985	1990	2000
Marriages	1.800	2.158	2.152	2.413	2.425	2.448	2.329
Divorces	0.479	0.708	1.036	1.182	1.187	1.175	1.135

FIGURE 4.142 Number of marriages and divorces (millions). *Source:* National Center for Health Statistics.

- a. Letting x = the number of marriages and y = the number of divorces in a year, plot the data. Do the data exhibit a linear trend?
 - b. Find the line of best fit.
 - c. Predict the number of divorces in a year when there are 2,750,000 marriages.
 - d. Predict the number of marriages in a year when there are 1,500,000 divorces.
 - e. Find the coefficient of linear correlation.
 - f. Are the predictions in parts (c) and (d) reliable? Why or why not?
- ▶ 12. The median home price and average mortgage rate in the United States for 1999–2004 are given in Figure 4.143.
- a. Letting x = the median price of a home and y = the average mortgage rate, plot the data. Do the data exhibit a linear trend?
 - b. Find the line of best fit.
 - c. Predict the average mortgage rate if the median price of a home is \$165,000.
 - d. Predict the median home price if the average mortgage rate is 7.25%.

- e. Find the coefficient of correlation.
- f. Are the predictions in parts (a) and (b) reliable? Why or why not?

Year	Median Price (dollars)	Mortgage Rate (percent)
1999	133,300	7.33
2000	139,000	8.03
2001	147,800	7.03
2002	158,100	6.55
2003	180,200	5.74
2004	195,200	5.73

FIGURE 4.143 Median home price and average mortgage rate, 1999–2004. *Source:* National Association of Realtors.



Answer the following questions using complete sentences and your own words.

• **CONCEPT QUESTIONS**

- 13. What is a line of best fit? How do you find it?
- 14. How do you measure the strength of a linear trend?
- 15. What is a positive linear relation? Give an example.
- 16. What is a negative linear relation? Give an example.

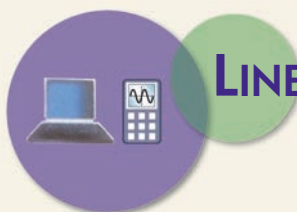
• **PROJECTS**

- 17. Measure the heights and weights of ten people. Let x = height and y = weight.
 - a. Plot the ordered pairs. Do the ordered pairs exhibit a linear trend?
 - b. Use the data to find the line of best fit.
 - c. Find the coefficient of linear correlation.
 - d. Will your line of best fit produce reliable predictions? Why or why not?



TECHNOLOGY AND LINEAR REGRESSION

In Example 2 of this section, we computed the slope and y -intercept of the line of best fit for the five ordered pairs (5, 14), (9, 17), (12, 16), (14, 18), and (17, 23). These calculations can be tedious when done by hand, even with only five data points. In the real world, there are always a large number of data points, and the calculations are always done with the aid of technology.



LINEAR REGRESSION ON A GRAPHING CALCULATOR

Graphing calculators can draw a scatter diagram, compute the slope and y -intercept of the line of best fit, and graph the line.

ON A TI-83/84:

- Put the calculator into statistics mode by pressing **STAT**.
- Set the calculator up for entering the data from Example 2 by selecting “Edit” from the “EDIT” menu, and the “List Screen” appears, as shown in Figure 4.144. If data already appear in a list (as they do in Figure 4.144) and you want to clear the list, use the arrow buttons to highlight the name of the list and press **CLEAR** **ENTER**.
- Use the arrow buttons and the **ENTER** button to enter the x -coordinates in list L_1 and the corresponding y -coordinates in list L_2 .

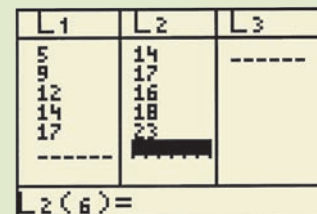


FIGURE 4.144 A TI-83/84’s list screen.

ON A CASIO:

- Put the calculator into statistics mode by pressing **MENU**, highlighting STAT, and pressing **EXE**.
- Use the arrow buttons and the **EXE** button to enter the x -coordinates from Example 2 in List 1 and the corresponding y -coordinates in List 2. When you are done, your screen should look like Figure 4.145. If data already appears in a list and you want to erase it, use the arrow keys to move to that list, press **F6** and then **F4** (i.e., **DEL-A**, which stands for “delete all”) and then **F1** (i.e., **YES**).

	List 1	List 2	List 3	List 4
1				
2	9	17		
3	12	16		
4	14	18		
5	17	23		
6				

FIGURE 4.145 A Casio's list screen.

Finding the Equation of the Line of Best Fit

Once the data have been entered, it is easy to find the equation. Recall from Example 2 that the equation is $y = 0.643192488x + 10.26760564$.

ON A TI-83/84:

- Press **2nd** **QUIT**.
- The TI-83/84 does not display the correlation coefficient unless you tell it to. To do so, press **2nd** **CATALOG**, scroll down, and select “DiagnosticOn.” When “DiagnosticOn” appears on the screen, press ENTER.
- Press **STAT**, scroll to the right, and select the “CALC” menu, and select “LinReg(ax + b)” from the “CALC” menu.
- When “LinReg(ax + b)” appears on the screen, press **ENTER**.
- The slope, the y -intercept and the correlation coefficient will appear on the screen, as shown in Figure 4.146. They are not labeled m , b , and r . Their labels are different and are explained in Figure 4.147.

```

LinReg
y=ax+b
a=.6431924883
b=10.26760563
r^2=.7797997424
r=.8830627058
  
```

FIGURE 4.146 Finding the equation of the line of best fit on a TI-83/84.

	TI-83/84 Labels
slope ma	
y -intercept b	b
correlation coefficient r	r

FIGURE 4.147 Graphing calculator line of best fit labels.

ON A CASIO:

- Press **CALC** (i.e., **F2**) to calculate the equation.
- Press **REG** (i.e., **F3**); “REG” stands for “regression.”
- Press **X** (i.e., **F1**) for linear regression. (See Figure 4.148.)

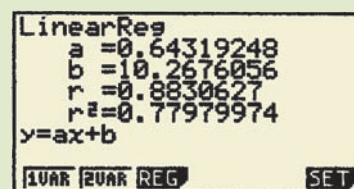


FIGURE 4.148 A Casio's equation of the line of best fit.

Drawing a Scatter Diagram and the Line of Best Fit

Once the equation has been found, the line can be graphed.

ON A TI-83/84:

- Press **Y=**, and clear any functions that may appear.
- Set the calculator up to draw a scatter diagram by pressing **2nd** **STAT PLOT** and selecting “Plot 1.” Turn the plot on and select the scatter icon.
- Tell the calculator to put the data entered in list L_1 on the x -axis by selecting “ L_1 ” for “Xlist,” and to put the data entered in list L_2 on the y -axis by selecting “ L_2 ” for “Ylist.”
- Automatically set the range of the window and obtain a scatter diagram by pressing **ZOOM** and selecting option 9: “ZoomStat.”
- If you don't want the data points displayed on the line of best fit, press **2nd** **STAT PLOT** and turn off plot 1.
- Quit the statistics mode by pressing **2nd** **QUIT**.
- Enter the equation of the line of best fit by pressing **Y=** **VARS**, selecting “Statistics,” scrolling to the right to select the “EQ” menu, and selecting “RegEQ” (for regression equation).
- Automatically set the range of the window and obtain a scatter diagram by pressing **ZOOM** and selecting option 9: “ZoomStat.” (See Figure 4.149.)
- Press **TRACE** to read off the data points as well as points on the line of best fit. Use the up and down arrows to switch between data points and points on the line of best fit. Use the left and right arrows to move left and right on the graph.

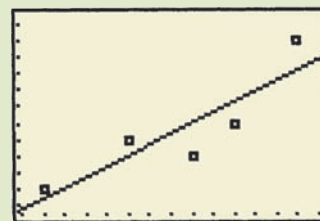


FIGURE 4.149 A TI-83/84's graph of the line of best fit and the scatter diagram.

ON A CASIO:

- Press **SHIFT** **QUIT** to return to List 1 and List 2.
- Press **GRPH** (i.e., **F1**).
- Press **SET** (i.e., **F6**).

- Make the resulting screen, which is labeled “StatGraph1,” read as follows:

Graph Type :Scatter
 Xlist :List1
 Ylist :List2
 Frequency :1

To make the screen read as described above,

- Use the down arrow button to scroll down to “Graph Type.”
- Press **Scat** (i.e., **F1**).
- In a similar manner, change “Xlist,” “Ylist,” and “Frequency” if necessary.
- Press **SHIFT** **QUIT** to return to List1 and List2.
- Press **GRPH** (i.e., **F1**).
- Press **GPH1** (i.e., **F1**) to obtain the scatter diagram.
- Press **X** (i.e., **F1**).
- Press **DRAW** (i.e., **F6**), and the calculator will display the scatter diagram as well as the line of best fit. (See Figure 4.150.)

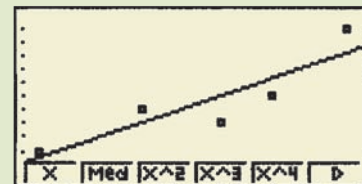


FIGURE 4.150 A Casio’s graph of the line of best fit and the scatter diagram.



LINEAR REGRESSION ON EXCEL

Entering the Data

We will use the data from Example 2. Start by entering the information as shown in Figure 4.151. Then save your spreadsheet.

	A	B	C
1	x-coordinate	y-coordinate	
2	5	14	
3	9	17	
4	12	16	
5	14	18	
6	17	23	
7			

FIGURE 4.151 The given data.

Using the Chart Wizard to Draw the Scatter Diagram

- Press the “Chart Wizard” button at the top of the spreadsheet. (It looks like a bar chart, and it might have a magic wand.)

- Select “XY (Scatter)” and choose the Chart sub-type that displays points without connecting lines.
- Press the “Next” button and a “Chart Source Data” box will appear.
- Click on the triangle to the right of “Data range.” This will cause the “Chart Source Data” box to shrink.
- Use your mouse to draw a box around all of the x -coordinates and y -coordinates. This will cause “= Sheet1!\$A\$2:\$B\$6” to appear in the “Chart Source Data” box.
- Press the triangle to the right of the “Chart Source Data” box and the box will expand.
- Press the “Next” button.
- Under “Chart Title” type an appropriate title. Our example lacks content, so we will use the title “Example 2.”
- After “Category (X)” type an appropriate title for the x -axis. Our example lacks content, so we will use the title “ x -axis.”
- After “Category (Y)” type an appropriate title for the y -axis.
- Click on the “Gridlines” tab and check the box under “value (X) axis,” next to “Major gridlines.”
- Click on the “Legend” tab and remove the check from the box next to “show legend.”
- Press the “Finish” button and the scatter diagram will appear.
- Save the spreadsheet.

After you have completed this step, your spreadsheet should include the scatter diagram shown in Figure 4.152.

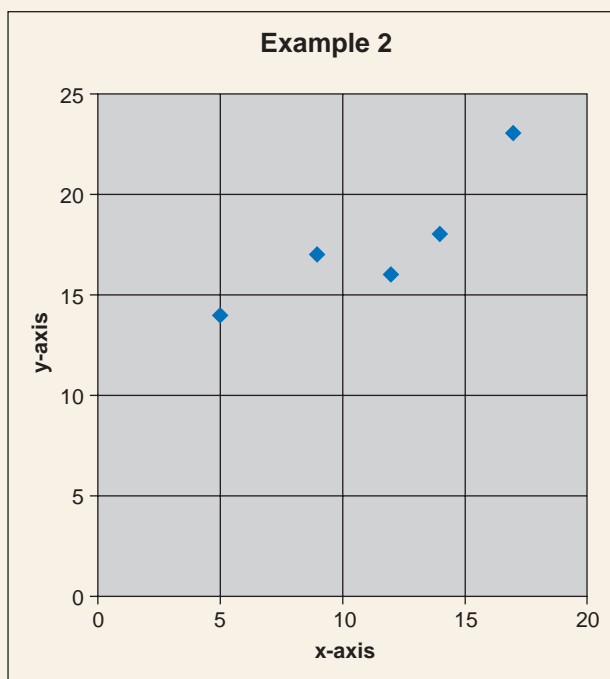


FIGURE 4.152 An Excel-generated scatter diagram.

Finding and Drawing the Line of Best Fit

- Click on the chart until it is surrounded by a thicker border.
- Use your mouse to select “Chart” at the very top of the screen. Then pull your mouse down until “Add Trendline . . .” is highlighted, and let go.
- Press the “Linear” button.

- Click on “Options”.
- Select “Display Equation on Chart”.
- Select “Display R-Squared value on Chart”.
- Press the “OK” button, and the scatter diagram, the line of best fit, the equation of the line of best fit, and r^2 (the square of the correlation coefficient) appear. If the equation is in the way, you can click on it and move it to a better location.
- Use the square root button on your calculator to find r , the correlation coefficient.
- Save the spreadsheet.

After completing this step, your spreadsheet should include the scatter diagram, the line of best fit, and the equation of the line of best fit, as shown in Figure 4.153.

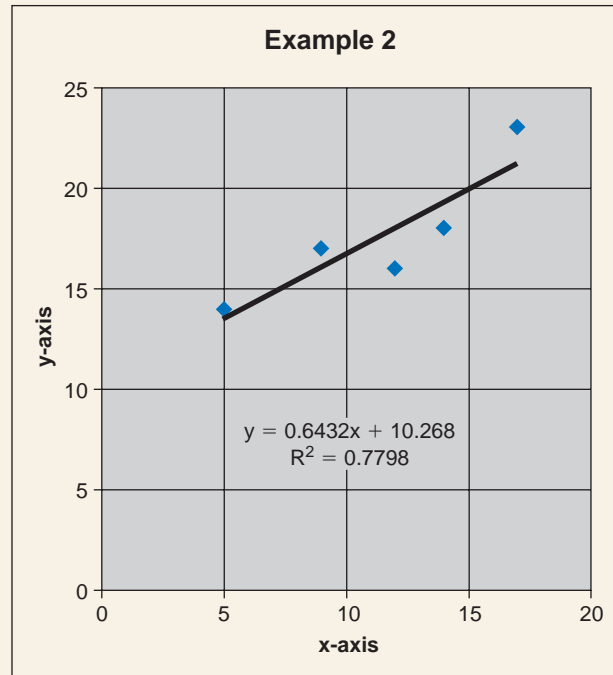


FIGURE 4.153 The scatter diagram with the line of best fit and its equation.

Printing the Scatter Diagram and Line of Best Fit

- Click on the diagram, until it is surrounded by a thicker border.
- Use your mouse to select “File” at the very top of the screen. Then pull your mouse down until “Print” is highlighted, and let go.
- Respond appropriately to the “Print” box that appears. Usually, it is sufficient to press a “Print” button or an “OK” button, but the actual way that you respond to the “Print” box depends on your printer.

EXERCISES

18. Throughout the twentieth century, the record time for the mile run steadily decreased, from 4 minutes 15.4 seconds in 1911 to 3 minutes 43.1 seconds in 1999. Some of the record times are given in Figure 4.154.
- Convert the given data to ordered pairs (x, y) , where x is the number of years since 1900 and y is the time in seconds.
 - Draw a scatter diagram, find the equation of the line of best fit, and find the correlation coefficient.

Year	1911	1923	1933	1942	1945	1954
Time (min:sec)	4:15.4	4:10.4	4:07.6	4:04.6	4:01.4	3:59.4

Year	1964	1967	1975	1980	1985	1999
Time (min:sec)	3:54.1	3:51.1	3:49.4	3:48.8	3:46.3	3:43.1

FIGURE 4.154 Record times for the mile run.

- Predict the record time for the mile run in the year 2020.
- Predict when the record time for the mile run will reach 3:30.
- Predict when the record time for the mile run will reach 0:30.

f. On what assumption are these predictions based? If this assumption is correct, how accurate are these predictions?

- Heart disease is much less of a problem in France than it is in the United States and other industrialized countries, despite a high consumption of saturated fats in France. One theory is that this lower heart disease rate is due to the fact that the French drink more wine. This so-called French paradox heightened an interest in moderate wine drinking. Figure 4.155 gives data on wine consumption and heart disease for nineteen industrialized countries, including France and the United States, in 1994.
 - Draw a scatter diagram, find the equation of line of best fit, and find the correlation for wine consumption versus annual deaths.
 - Do the data support the theory that wine consumption lowers the heart disease rate?

Country	Wine Consumption (liters per person)	Annual Deaths from Heart Disease per 100,000 People
Australia	2.5	211
Austria	3.9	167
Belgium	2.9	131
Canada	2.4	191
Denmark	2.9	220
Finland	0.8	297
France	9.1	71
Iceland	0.8	211
Ireland	0.6	300
Italy	7.9	107
Netherlands	1.8	266
New Zealand	1.9	266
Norway	0.8	227
Spain	6.5	86
Sweden	1.6	207
Switzerland	5.8	115
United Kingdom	1.3	285
United States	1.2	199
West Germany	2.7	172

FIGURE 4.155 Wine consumption and heart disease. *Source: New York Times, December 28, 1994.*

4

CHAPTER REVIEW

TERMS

bell-shaped curve
body table
categorical data
cells
coefficient of linear correlation
computerized spreadsheet
continuous variable
data point
density
descriptive statistics

deviation from the mean
discrete variable
error of the estimate
extreme value
frequency
frequency distribution
histogram
Inferential statistics
level of confidence
linear regression
linear trend
line of best fit
margin of error (MOE)

mathematical model
mean
measures of central tendency
measures of dispersion
median
mode
normal distribution
outlier
pie chart
population
population proportion
relative frequency

relative frequency density
sample
sample proportion
spreadsheet
standard deviation
standard normal distribution
statistics
summation notation
tail
variance
z-distribution
z-number

REVIEW EXERCISES

1. Find (a) the mean, (b) the median, (c) the mode, and (d) the standard deviation of the following set of raw data:

5 8 10 4 8 10 6 8 7 5

2. To study the composition of families in Winslow, Arizona, forty randomly selected married couples were surveyed to determine the number of children in each family. The following results were obtained:

3 1 0 4 1 3 2 2 0 2 0 2 2 1
4 3 1 1 3 4 2 1 3 0 1 0 2 5
1 2 3 0 0 1 2 3 1 2 0 2

- a. Organize the given data by creating a frequency distribution.
b. Find the mean number of children per family.
c. Find the median number of children per family.
d. Find the mode number of children per family.
e. Find the standard deviation of the number of children per family.
f. Construct a histogram using single-valued classes of data.
3. The frequency distribution in Figure 4.156 lists the number of hours per day that a randomly selected sample of teenagers spent watching television. Where possible, determine what percent of the teenagers

$x = \text{Hours per Day}$	Frequency
$0 \leq x < 2$	23
$2 \leq x < 4$	45
$4 \leq x < 6$	53
$6 \leq x < 8$	31
$8 \leq x \leq 10$	17

FIGURE 4.156 Time watching television.

spent the following number of hours watching television.

- a. less than 4 hours b. not less than 6 hours
c. at least 2 hours d. less than 2 hours
e. at least 4 hours but less than 8 hours
f. more than 3.5 hours
4. To study the efficiency of its new oil-changing system, a local service station monitored the amount of time it took to change the oil in customers' cars. The frequency distribution in Figure 4.157 summarizes the findings.

$x = \text{Time}$ (in minutes)	Number of Customers
$3 \leq x < 6$	18
$6 \leq x < 9$	42
$9 \leq x < 12$	64
$12 \leq x < 15$	35
$15 \leq x \leq 18$	12

FIGURE 4.157 Time to change oil.

- Find the mean number of minutes to change the oil in a car.
 - Find the standard deviation of the amount of time to change the oil in a car.
 - Construct a histogram to represent the data.
- If your scores on the first four exams (in this class) are 74, 65, 85, and 76, what score do you need on the next exam for your overall mean to be at least 80?
 - The mean salary of twelve men is \$37,000, and the mean salary of eight women is \$28,000. Find the mean salary of all twenty people.
 - Timo and Henke golfed five times during their vacation. Their scores are given in Figure 4.158.

Timo	103	99	107	93	92
Henke	101	92	83	96	111

FIGURE 4.158 Golf scores.

- Find the mean score of each golfer. Who has the lowest mean?
 - Find the standard deviation of each golfer's scores.
 - Who is the more consistent golfer? Why?
- Suzanne surveyed the prices for a quart of a certain brand of motor oil. The sample data, in dollars per quart, are summarized in Figure 4.159.

Price per Quart	Number of Stores
1.99	2
2.09	3
2.19	7
2.29	10
2.39	14
2.49	4

FIGURE 4.159 Price of motor oil.

- Find the mean and standard deviation of the prices.
 - What percent of the data lies within one standard deviation of the mean?
 - What percent of the data lies within two standard deviations of the mean?
 - What percent of the data lies within three standard deviations of the mean?
- Classify the following types of data as discrete, continuous, or neither.
 - weights of motorcycles
 - colors of motorcycles
 - number of motorcycles
 - ethnic background of students
 - number of students
 - amounts of time spent studying
 - What percent of the standard normal z -distribution lies in the following intervals?
 - between $z = 0$ and $z = 1.75$
 - between $z = -1.75$ and $z = 0$
 - between $z = -1.75$ and $z = 1.75$
 - A large group of data is normally distributed with mean 78 and standard deviation 7.
 - Find the intervals that represent one, two, and three standard deviations of the mean.
 - What percent of the data lies in each interval in part (a)?
 - Draw a sketch of the bell curve.
 - The time it takes a latex paint to dry is normally distributed. If the mean is $3\frac{1}{2}$ hours with a standard deviation of 45 minutes, find the probability that the drying time will be as follows.
 - less than 2 hours 15 minutes
 - between 3 and 4 hours

HINT: Convert everything to hours (or to minutes).
 - All incoming freshmen at a major university are given a diagnostic mathematics exam. The scores are normally distributed with a mean of 420 and a standard deviation of 45. If the student scores less than a certain score, he or she will have to take a review course. Find the cutoff score at which 34% of the students would have to take the review course.
 - Find the specified z -number.
 - $z_{0.4441}$
 - $z_{0.4500}$
 - $z_{0.1950}$
 - $z_{0.4975}$
 - A survey asked, "Do you think that the president is doing a good job?" Of the 1,200 Americans surveyed, 800 responded yes. For each of the following levels of confidence, find the sample proportion and the margin of error associated with the poll.
 - a 90% level of confidence
 - a 95% level of confidence

16. A survey asked, “Do you support capital punishment?” Of the 1,000 Americans surveyed, 750 responded no. For each of the following levels of confidence, find the sample proportion and the margin of error associated with the poll.
- an 80% level of confidence
 - a 98% level of confidence
17. A sample consisting of 580 men and 970 women was asked various questions pertaining to international affairs. For a 95% level of confidence, find the margin of error associated with the following samples.
- the male sample
 - the female sample
 - the combined sample
18. A poll pertaining to environmental concerns had the following footnote: “Based on a sample of 1,098 adults, the margin of error is plus or minus 2 percentage points.” Find the level of confidence of the poll.
19. You are planning a survey with a 94% level of confidence in your findings. How large should your sample be so that the margin of error is at most 2.5%?
20. A set of $n = 5$ ordered pairs has the following sums:

$$\begin{aligned}\sum x &= 66 & \sum x^2 &= 1,094 & \sum y &= 273 \\ \sum y^2 &= 16,911 & \sum xy &= 4,272\end{aligned}$$

- Find the line of best fit.
 - Predict the value of y when $x = 16$.
 - Predict the value of x when $y = 57$.
 - Find the coefficient of linear correlation.
 - Are the predictions in parts (b) and (c) reliable? Why or why not?
21. Given the ordered pairs (5, 38), (10, 30), (20, 33), (21, 25), (24, 18), and (30, 20),
- Plot the ordered pairs. Do the ordered pairs exhibit a linear trend?
 - Find the line of best fit.
 - Predict the value of y when $x = 15$.
 - Plot the given ordered pairs and sketch the graph of the line of best fit on the same coordinate system.
 - Find the coefficient of linear correlation.
 - Is the prediction in part (c) reliable? Why or why not?
22. The value of agricultural exports and imports in the United States for 1999–2007 are given in Figure 4.160.
- Letting $x =$ the value of agricultural exports and $y =$ the value of agricultural imports, plot the data. Do the data exhibit a linear trend?
 - Find the line of best fit.

Year	Agricultural Exports (billion dollars)	Agricultural Imports (billion dollars)
1999	48.4	37.7
2000	51.2	39.0
2001	53.7	39.4
2002	53.1	41.9
2003	59.5	47.3
2004	62.4	52.7
2005	62.5	57.7
2006	68.7	64.0
2007	82.2	70.1

FIGURE 4.160 Agricultural exports and imports (billion dollars), 1999–2007. Source: U.S. Department of Agriculture.

- Predict the value of agricultural imports when the value of agricultural exports is 75.0 billion dollars.
- Predict the value of agricultural exports when the value of agricultural imports is 60.0 billion dollars.
- Find the coefficient of correlation.
- Are the predictions in parts (a) and (b) reliable? Why or why not?



Answer the following questions using complete sentences and your own words.

• CONCEPT QUESTIONS

- What are the three measures of central tendency? Briefly explain the meaning of each.
- What does standard deviation measure?
- What is a normal distribution? What is the standard normal distribution?
- What is a margin of error? How is it related to a sample proportion?
- How do you measure the strength of a linear trend?

• HISTORY QUESTIONS

- What role did the following people play in the development of statistics?
 - George Gallup
 - Carl Friedrich Gauss



Jeffrey Coolidge/Stone/Getty Images

How you deal with money will have a big impact on the quality of your life. There's no question that you will borrow money. You probably already have a credit card loan or a student loan. Sooner or later, you'll borrow money for a car or a house. The question is whether or not you'll know enough that you can be wise about your borrowing.

Not everybody saves money, but your life will be less stressful and more successful if you do. You might now have a savings account at a bank or a money market account. Will you know enough about the mathematics of finance to be wise about your savings?

It's to your advantage to know the mathematics of finance. Right now, you probably know next to nothing about it. You might not even know what a money market account is or what an annuity is. You should know these things.



WHAT WE WILL DO IN THIS CHAPTER

BORROWING:

- Most loans require that you pay simple interest, so we will explore that.
- Many loans, including car loans and home loans, are amortized—that is, they require monthly payments—so we will explore amortized loans.

SAVING:

- Many investments, including savings accounts, certificates of deposit, and money market accounts, pay compound interest, so we will discuss how compound interest works.
- One of the best ways for an average person to save money is through an annuity, so we will explore that.

5.1 Simple Interest

OBJECTIVES

- Make simple interest calculations
- Determine a credit card finance charge
- Find the payment required by an add-on interest loan

Loans and investments are very similar financial transactions. Each involves:

- the flow of money from a source to another party
- the return of the money to its source
- the payment of a fee to the source for the use of the money

When you make a deposit in your savings account, you are making an investment, but the bank views it as a loan; you are lending the bank your money, which they will lend to another customer, perhaps to buy a house. When you borrow money to buy a car, you view the transaction as a loan, but the bank views it as an investment; the bank is investing its money in you in order to make a profit.

When an investor puts money into a savings account or buys a certificate of deposit (CD) or a Treasury bill, the investor expects to make a profit. The amount of money that is invested is called the **principal**. The profit is the **interest**. How much interest will be paid depends on the **interest rate** (usually expressed as a percent per year); the **term**, or length of time that the money is invested; and how the interest is calculated.

In this section, we'll explore simple interest in investments and short-term loans. **Simple interest** means that the amount of interest is calculated as a percent per year of the principal.

SIMPLE INTEREST FORMULA

The **simple interest** I on a principal P at an annual rate of interest r for a term of t years is

$$I = Prt$$

EXAMPLE 1

USING THE SIMPLE INTEREST FORMULA Tom and Betty buy a two-year CD that pays 5.1% simple interest from their bank for \$150,000. (Many banks pay simple interest on larger CDs and compound interest on smaller CDs.) They invest \$150,000, so the principal is $P = \$150,000$. The interest rate is $r = 5.1\% = 0.051$, and the term is $t = 2$ years.

- Find the interest that the investment earns.
- Find the value of the CD at the end of its term.

SOLUTION

- Using the Simple Interest Formula, we have

$$\begin{aligned} I &= Prt && \text{the Simple Interest Formula} \\ &= 150,000 \cdot 0.051 \cdot 2 && \text{substituting for } P, r, \text{ and } t \\ &= \$15,300 \end{aligned}$$

- b. Two years in the future, the bank will pay Tom and Betty

$$\$150,000 \text{ principal} + \$15,300 \text{ interest} = \$165,300$$

This is called the *future value* of the CD, because it is what the CD will be worth in the future.

Future Value

In Example 1, we found the future value by adding the principal to the interest. The **future value** FV is always the sum of the principal P plus the interest I . If we combine this fact with the Simple Interest Formula, we can get a formula for the future value:

$$\begin{aligned} FV &= P + I \\ &= P + Prt && \text{from the Simple Interest Formula} \\ &= P(1 + rt) && \text{factoring} \end{aligned}$$

SIMPLE INTEREST FUTURE VALUE FORMULA

The **future value** FV of a principal P at an annual rate of interest r for t years is

$$FV = P(1 + rt)$$

Capital Letters for Money

There is an important distinction between the variables FV , I , P , r , and t : FV , I , and P measure amounts of money, whereas r and t do not. For example, consider the interest rate r and the interest I . Frequently, people confuse these two. However, the interest rate r is a percentage, and the interest I is an amount of money; Tom and Betty's interest rate is $r = 5.1\%$, but their interest is $I = \$15,300$. To emphasize this distinction, we will always use capital letters for variables that measure amounts of money and lowercase letters for other variables. We hope this notation will help you avoid substituting $5.1\% = 0.051$ for I when it should be substituted for r .

Finding the Number of Days: "Through" versus "To"

Clearly, there is only one day from January 1 to January 2. If the answer weren't so obvious, we could find it by subtracting:

$$2 - 1 = 1 \text{ day}$$

This count of days includes January 1 but not January 2.

In contrast, there are *two* days from January 1 *through* January 2. The word “through” means to include both the beginning day (January 1) and the ending day (January 2). If the answer weren’t so obvious, we could find it by doing the “January 1 to January 2” calculation above and then adding the one ending day:

$$\underbrace{2 - 1}_{\text{the 1st to the 2nd}} + \underbrace{1}_{\text{adding the one ending day}} = 2$$

The number of days from August 2 to August 30 is

$$30 - 2 = 28 \text{ days}$$

This count includes August 2 but not August 30.

The number of days from August 2 *through* August 30 is

$$30 - 2 + 1 = 29 \text{ days}$$

This count includes August 2 and August 30.

Short Term Loans

One of the more common uses of simple interest is a short-term (such as a year or less) loan that requires a single lump sum payment at the end of the term. Businesses routinely obtain these loans to purchase equipment or inventory, to pay operating expenses, or to pay taxes. A **lump sum** payment is a single payment that pays off an entire loan. Some loans require smaller monthly payments rather than a single lump sum payment. We will investigate that type of loan later in this section and in Section 5.4.

EXAMPLE 2

USING THE SIMPLE INTEREST FUTURE VALUE FORMULA Espree Clothing borrowed \$185,000 at $7\frac{1}{4}\%$ from January 1 to February 28.

- Find the future value of the loan.
- Interpret the future value.

SOLUTION

- We are given:

$$P = 185,000$$

$$r = 7\frac{1}{4}\% = 0.0725$$

- Finding t :
 - January has 31 days.
 - February has 28 days.
 - There are $31 + 28 - 1 = 58$ days from January 1 to February 28.
 - $t = 58$ days

$$= 58 \text{ days} \cdot \frac{1 \text{ year}}{365 \text{ days}} = \frac{58}{365} \text{ years} \quad \text{using dimensional analysis (see Appendix E)}$$

- Finding FV :

$$FV = P(1 + rt)$$

$$= 185,000 \left(1 + 0.0725 \cdot \frac{58}{365} \right)$$

$$= 187,131.301 \dots$$

$$\approx \$187,131.30$$

the Simple Interest Future Value

Formula

substituting for P , r , and t

rounding to the nearest penny



$$185000 \times (1 + .0725 \times 58 \div 365) =$$

$$185000 \times (1 + .0725 \times 58 \div 365) \text{ ENTER}$$

- b. This means that Espree has agreed to make a lump sum payment to their lender \$187,168.05 on February 28.



As a rough check, notice that the answer is somewhat higher than the original \$185,000. This is as it should be. The future value includes the original \$185,000 plus interest.

Notice that in Example 2, we do not use $t = 2 \text{ months} = \frac{2}{12} \text{ year}$. If we did, we would get \$187,235.42, an inaccurate answer. Instead, we use the number of days, converted to years.

In Example 2, we naturally used 365 days per year, but some institutions traditionally count a year as 360 days and a month as thirty days (especially if that tradition works in their favor). This is a holdover from the days before calculators and computers—the numbers were simply easier to work with. Also, we used normal round-off rules to round \$187,131.301 . . . to \$187,131.30; some institutions round off some interest calculations in their favor. In this book, we will count a year as 365 days and use normal round-off rules (unless stated otherwise).

The issue of how a financial institution rounds off its interest calculations might seem unimportant. After all, we are talking about a difference of a fraction of a penny. However, consider one classic form of computer crime: the round-down fraud, performed on a computer system that processes a large number of accounts. Frequently, such systems use the normal round-off rules in their calculations and keep track of the difference between the theoretical account balance if no rounding off is done and the actual account balance with rounding off. Whenever that difference reaches or exceeds 1¢, the extra penny is deposited in (or withdrawn from) the account. A fraudulent computer programmer can write the program so that the extra penny is deposited in his or her own account. This fraud is difficult to detect because the accounts appear to be balanced. While each individual gain is small, the total gain can be quite large if a large number of accounts is processed regularly.

A written contract signed by the lender and the borrower is called a **loan agreement** or a **note**. The **maturity value** of the note (or just the **value** of the note) refers to the note's future value. Thus, the value of the note in Example 2 was \$187,168.05. This is what the note is worth to the lender in the future—that is, when the note matures.

National Debt

In almost every year since 1931, the U.S. federal budget called for deficit spending, that is, spending more money than is received. In 2008, the total U.S. federal debt was \$10,025 billion, and we paid about \$214 billion for interest on that debt. In 2009, the debt grew to \$11,343 billion, or about \$37,000 per person in the United States.

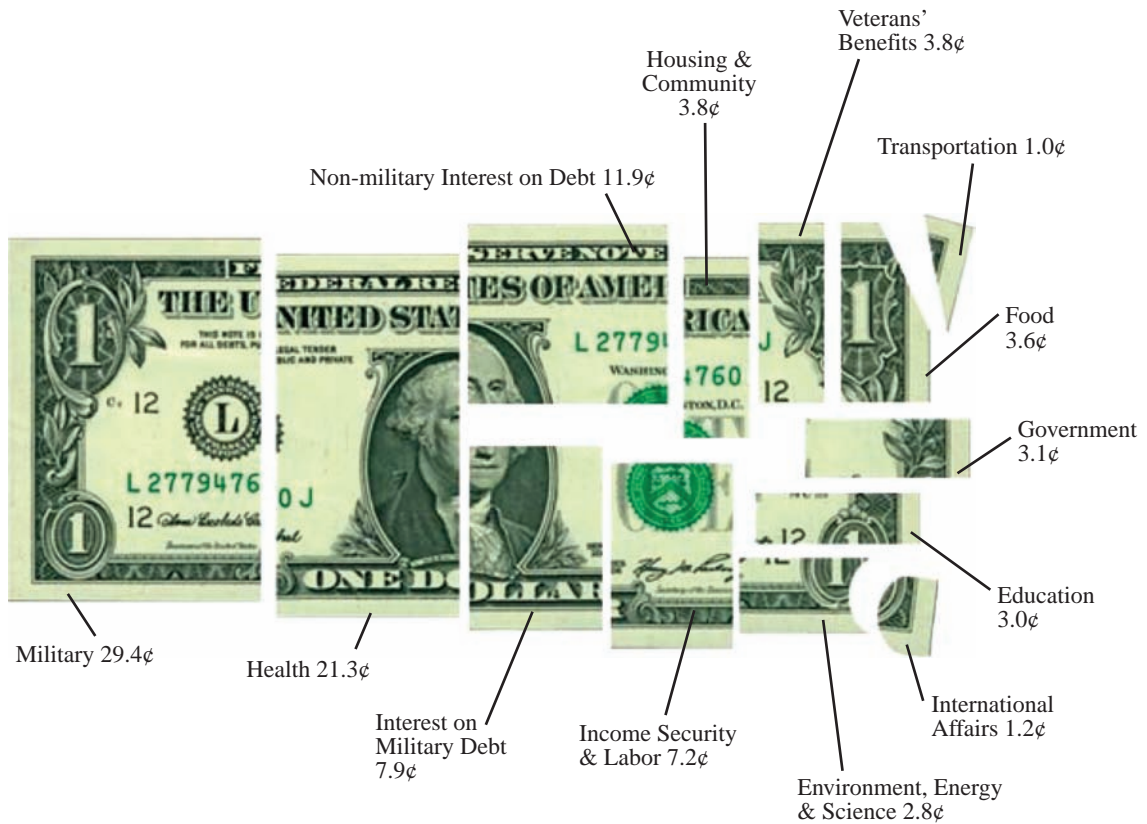


FIGURE 5.1 Where your 2008 federal income tax dollar was spent. *Source:* National Priorities Project.

EXAMPLE 3

THE NATIONAL DEBT Find the simple interest rate that was paid on the 2008 national debt.

SOLUTION

- We are given

$$\begin{aligned}
 I &= \$214 \text{ billion} \\
 P &= \$10,025 \text{ billion} \\
 t &= 1 \text{ year}
 \end{aligned}$$

- Finding r :

$$\begin{aligned}
 I &= Prt && \text{the Simple Interest Formula} \\
 \$214 \text{ billion} &= \$10,025 \text{ billion} \cdot r \cdot 1 \\
 r &= \$21 \text{ billion} / \$10,025 \text{ billion} \cdot r \cdot 1 && \text{substituting solving for } r \\
 &= 0.0213 \dots \approx 2.1\% && \text{rounding}
 \end{aligned}$$

Present Value

EXAMPLE 4

FINDING HOW MUCH TO INVEST NOW Find the amount of money that must be invested now at a $5\frac{3}{4}\%$ simple interest so that it will be worth \$1,000 in two years.

SOLUTION

We are asked to find the principal P that will generate a future value of \$1,000.

- We are given

$$FV = \$1,000$$

$$r = 5\frac{3}{4}\% = 0.0575$$

$$t = 2 \text{ years}$$

- Finding P :

$$FV = P(1 + rt)$$

$$1,000 = P(1 + 0.0575 \cdot 2)$$

$$P = \frac{1,000}{1 + 0.0575 \cdot 2}$$

$$= 896.86099 \dots$$

$$\approx \$896.86$$

the Simple Interest Future Value Formula
substituting

solving for P

rounding



$$1000 \div (1 + .0575 \times 2) =$$

$$1000 \div (1 + .0575 \times 2) \text{ ENTER}$$



As a rough check, notice that the answer is somewhat smaller than \$1,000. This is as it should be—the principal does not include interest, so it should be smaller than the future value of \$1,000.

In Example 4, the investment is worth \$1,000 two years *in the future*; that is, \$1000 is the investment's *future value*. But the same investment is worth \$896.86 *in the present*. For this reason, we say that \$896.86 is the investment's **present value**. In this case, this is the same thing as the principal; it is just called the *present value* to emphasize that this is its value in the present.

Add-on Interest

An **add-on interest loan** is an older type of loan that was common before calculators and computers, because calculations of such loans can easily be done by hand. With this type of loan, the total amount to be repaid is computed with the Simple Interest Future Value Formula. The payment is found by dividing that total amount by the number of payments.

Add-on interest loans have generally been replaced with the more modern amortized loans (see Section 5.4). However, they are not uncommon at auto lots that appeal to buyers who have poor credit histories.

EXAMPLE 5

AN ADD-ON INTEREST LOAN Chip Douglas's car died, and he must replace it right away. Centerville Auto Sales has a nine-year-old Ford that's "like new" for \$5,988. The sign in their window says, "Bad credit? No problem!" They offered Chip a 5% two-year add-on interest loan if he made a \$600 down payment. Find the monthly payment.

SOLUTION

The loan amount is $P = 5,988 - 600 = 5,388$. The total amount due is

$$\begin{aligned} FV &= P(1 + rt) && \text{the Simple Interest Future Value Formula} \\ &= 5,388(1 + 0.05 \cdot 2) && \text{substituting} \\ &= 5,926.80 \end{aligned}$$

The total amount due is spread out over twenty-four monthly payments, so the monthly payment is $\$5,926.8/24 = \246.95 . This means that Chip has to pay \$600 when he purchases the car and then \$246.95 a month for twenty-four months.

Credit Card Finance Charge

Credit cards have become part of the American way of life. Purchases made with a credit card are subject to a finance charge, but there is frequently a grace period, and no finance charge is assessed if full payment is received by the payment due date. One of the most common methods of calculating credit card interest is the **average daily balance** method. To find the average daily balance, the balance owed on the account is found for each day in the billing period, and the results are averaged. The finance charge consists of simple interest, charged on the result.

EXAMPLE 6

FINDING A CREDIT CARD FINANCE CHARGE The activity on Tom and Betty's Visa account for one billing is shown below. The billing period is October 15 through November 14, the previous balance was \$346.57, and the annual in rate is 21%.

October 21	payment	\$50.00
October 23	restaurant	\$42.14
November 7	clothing	\$18.55

- Find the average daily balance.
- Find the finance charge.

SOLUTION

- To find the average daily balance, we have to know the balance for each day in the billing period and the number of days at that balance, as shown in Figure 5.2. The average daily balance is then the weighted average of each daily balance, with each balance weighted to reflect the number of days at that balance.

Time Interval	Days	Daily Balance
October 15–20	$20 - 14 = 6$	\$346.57
October 21–22	$22 - 20 = 2$	$\$346.57 - \$50 = \$296.57$
October 23–November 6 (October has 31 days)	$31 - 22 = 9$ $9 + 6 = 15$	$\$296.57 + \$42.14 = \$338.71$
November 7–14	$14 - 6 = 8$	$\$338.71 + \$18.55 = \$357.26$

FIGURE 5.2 Preparing to find the average daily balance.

$$\begin{aligned} \text{Average daily balance} &= \frac{6 \cdot 346.57 + 2 \cdot 296.57 + 15 \cdot 338.71 + 8 \cdot 357.26}{6 + 2 + 15 + 8} \\ &= 342.29967 \dots \\ &\approx \$342.30 \end{aligned}$$



$$\left(6 \times 346.57 + 2 \times 296.57 + 15 \times 338.71 + 8 \times 357.26 \right) \div \left(6 + 2 + 15 + 8 \right) =$$

With a graphing calculator, type **ENTER** instead of **=**.

b. • We are given

$$P = \$342.29967 \dots$$

$$r = 21\% = 0.21$$

$$t = 31 \text{ days} = 31/365 \text{ years}$$

• Finding I :

$$I = Prt$$

$$= 342.29967 \dots \cdot 0.21 \cdot 31/365$$

$$= 6.1051258 \dots$$

$$\approx \$6.11$$

the Simple Interest Formula
substituting
multiplying
rounding

CREDIT CARD FINANCE CHARGE

To find the credit card finance charge with the average daily balance method, do the following:

1. Find the balance for each day in the billing period and the number of days at that balance.
2. The average daily balance is the weighted average of these daily balances, weighted to reflect the number of days at that balance.
3. The finance charge is simple interest applied to the average daily balance.

How Many Days?

Financial calculations usually involve computing the term t either in a whole number of years or in the number of days converted to years. This requires that you know the number of days in each month. As you can see from Figure 5.3, the months alternate between thirty-one days and thirty days, with two exceptions:

- February has twenty-eight days (twenty-nine in leap years).
- The alternation does not happen from July to August.

January	31 days	May	31 days	September	30 days
February	28 days	June	30 days	October	31 days
March	31 days	July	31 days	November	30 days
April	30 days	August	31 days	December	31 days

FIGURE 5.3 The number of days in a month.

HISTORICAL NOTE

THE HISTORY OF CREDIT CARDS

Credit cards and charge cards were first used in the United States in 1915, when Western Union issued a metal card to some of its regular customers. Holders of these cards were allowed to defer their payments and were assured of prompt and courteous service. Shortly thereafter, several gasoline companies, hotels, department stores, and railroads issued cards to their preferred customers.

The use of credit and charge cards virtually ceased during World War II, owing to government restrictions on credit. In 1950, a New York lawyer established the Diners' Club after being embarrassed when he lacked sufficient cash to pay a dinner bill. A year later, the club had billed more than \$1 million. Carte Blanche and the American Express card soon followed. These "travel and entertainment" cards were and are attractive to the public because they provide a convenient means of paying restaurant, hotel, and airline bills. They eliminate both the possibility of an out-of-town check being refused and the need to carry a large amount of cash.



Library of Congress, Prints and Photographs Division

The first bank card was issued in 1951 by Franklin National Bank in New York. Within a few years, 100 other banks had followed suit. Because these bank cards were issued by individual banks, they

were accepted for use only in relatively small geographical areas. In 1965, the California-based Bank of America began licensing other banks (both in the United States and abroad) to issue BankAmericards. Other banks formed similar groups, which allowed customers to use the cards out of state.

In 1970, the Bank of America transferred administration of its bank card program to a new company owned by the banks that issued the card. In 1977, the card was renamed Visa. MasterCard was originally created by several California banks (United California Bank, Wells Fargo, Crocker Bank, and the Bank of California) to compete with BankAmericard. It was then licensed to the First

National Bank of Louisville, Kentucky and the Marine Midland Bank of New York.

Credit cards are now a part of the American way of life. In 2004, we started to use credit cards, debit cards and other forms of electronic bill paying more than we use paper checks. Credit card debt has soared, especially among college students who are already in debt with college loans. Credit card issuers have been accused of targeting college students, who tend to make minimum payments and thus incur much more interest. See Exercises 45–48.



Library of Congress, Prints and Photographs Division
Early Diners' Club, Carte Blanche and American Express cards

5.1 EXERCISES

In Exercises 1–4, find the number of days.

1. a. September 1 to October 31 of the same year
b. September 1 through October 31 of the same year
2. a. July 1 to December 31 of the same year
b. July 1 through December 31 of the same year
- ▶ 3. a. April 1 to July 10 of the same year
b. April 1 through July 10 of the same year
4. a. March 10 to December 20 of the same year
b. March 10 through December 20 of the same year

In Exercises 5–10, find the simple interest of the given loan amount.

- ▶ 5. \$2,000 borrowed at 8% for three years
6. \$35,037 borrowed at 6% for two years
- ▶ 7. \$420 borrowed at $6\frac{3}{4}\%$ for 325 days
8. \$8,950 borrowed at $9\frac{1}{2}\%$ for 278 days
9. \$1,410 borrowed at $12\frac{1}{4}\%$ from September 1:
 - a. to October 31 of the same year
 - b. through October 31 of the same year

▶ Selected exercises available online at www.webassign.net/brookscole

- ▶ 10. \$5,682 borrowed at $11\frac{3}{4}\%$ from July 1:
 - a. to December 31 of the same year
 - b. through December 31 of the same year

In Exercises 11–14, find the future value of the given present value.

- ▶ 11. Present value of \$3,670 at $2\frac{3}{4}\%$ for seven years
- 12. Present value of \$4,719 at 14.1% for eleven years
- ▶ 13. Present value of \$12,430 at $5\frac{7}{8}\%$ for 660 days
- 14. Present value of \$172.39 at 6% for 700 days

In Exercises 15–20, find the maturity value of the given loan amount.

- ▶ 15. \$1,400 borrowed at $7\frac{1}{8}\%$ for three years
- 16. \$3,250 borrowed at $8\frac{1}{2}\%$ for four years
- 17. \$5,900 borrowed at $14\frac{1}{2}\%$ for 112 days
- ▶ 18. \$2,720 borrowed at $12\frac{3}{4}\%$ for 275 days
- 19. \$16,500 borrowed at $11\frac{7}{8}\%$ from April 1 through July 10 of the same year
- ▶ 20. \$2,234 borrowed at $12\frac{1}{8}\%$ from March 10 through December 20 of the same year

In Exercises 21–26 find the present value of the given future value.

- ▶ 21. Future value \$8,600 at $9\frac{1}{2}\%$ simple interest for three years
- 22. Future value \$420 at $5\frac{1}{2}\%$ simple interest for two years
- ▶ 23. Future value \$1,112 at $3\frac{5}{8}\%$ simple interest for 512 days
- ▶ 24. Future value \$5,750 at $4\frac{7}{8}\%$ simple interest for 630 days
- 25. Future value \$1,311 at $6\frac{1}{2}\%$ simple interest from February 10 to October 15 of the same year
- 26. Future value \$4,200 at $6\frac{3}{4}\%$ simple interest from April 12 to November 28 of the same year
- ▶ 27. If you borrow \$1,000 at 8.5% interest and the loan requires a lump sum payment of \$1,235.84, what is the term of the loan?
- ▶ 28. If you borrow \$1,700 at 5.25% interest and the loan requires a lump sum payment of \$1,863.12, what is the term of the loan?
- 29. The Square Wheel Bicycle store has found that they sell most of their bikes in the spring and early summer. On March 1, they borrowed \$226,500 to buy bicycles. They are confident that they can sell most of these bikes by August 1. Their loan is at $6\frac{7}{8}\%$ interest. What size lump sum payment would they have to make on August 1 to pay off the loan?
- ▶ 30. Ernie Bilko has a business idea. He wants to rent an abandoned gas station for just the months of November and December. He will convert the gas station into a drive-through Christmas wrapping station. Customers will drive in, drop off their gifts, return the next day, and pick up their wrapped gifts. He needs \$338,200 to rent the gas station, purchase

wrapping paper, hire workers, and advertise. If he borrows this amount at $6\frac{1}{2}\%$ interest for those two months, what size lump sum payment will he have to make to pay off the loan?

- 31. Alice Cohen buys a two-year-old Honda from U-Pay-Less-Cars for \$19,000. She puts \$500 down and finances the rest through the dealer at 13% add-on interest. If she agrees to make thirty-six monthly payments, find the size of each payment.
- ▶ 32. Sven Lundgren buys a three-year-old Chevrolet from Skunk Motors for \$14,600. He puts \$300 down and finances the rest through the dealer at 12.5% add-on interest. If he agrees to make twenty-four monthly payments, find the size of each payment.
- 33. Ray and Teresa Martinez buy a used car from Fowler's Wholesale 2U for \$6,700. They put \$500 down and finance the rest through the car lot at 9.8% add-on interest. If they make thirty-six monthly payments, find the size of each payment.
- ▶ 34. Dick Davis buys a five-year-old used Toyota from Pioneer Auto Sales for \$7,999. He puts \$400 down and finances the rest through the car lot at 7.7% add-on interest. He agrees to make thirty-six monthly payments. Find the size of each payment.

In Exercises 35–36, use the following information. When “trading up,” it is preferable to sell your old house before buying your new house because that allows you to use the proceeds from selling your old house to buy your new house. When circumstances do not allow this, the homeowner can take out a bridge loan.

- 35. Dale and Claudia have sold their house, but they will not get the proceeds from the sale for an estimated $2\frac{1}{2}$ months. The owner of the house they want to buy will not hold the house that long. Dale and Claudia have two choices: let their dream house go or take out a bridge loan. The bridge loan would be for \$110,000, at 7.75% simple interest, due in ninety days.
 - a. How big of a check would they have to write in 90 days?
 - b. How much interest would they pay for this loan?
- ▶ 36. Tina and Mike have sold their house, but they will not get the proceeds from the sale for an estimated 3 months. The owner of the house they want to buy will not hold the house that long. Tina and Mike have two choices: let their dream house go or take out a bridge loan. The bridge loan would be for \$85,000, at 8.5% simple interest, due in 120 days.
 - a. How big of a check would they have to write in 120 days?
 - b. How much interest would they pay for this loan?
- 37. The activity on Stuart Ratner's Visa account for one billing period is shown below. Find the average daily balance and the finance charge if the billing period is

April 11 through May 10, the previous balance was \$126.38, and the annual interest rate is 18%.

April 15	payment	\$15.00
April 22	DVD store	\$25.52
May 1	clothing	\$32.18

- ▶ **38.** The activity on Marny Zell's MasterCard account for one billing period is shown below. Find the average daily balance and the finance charge if the billing period is June 26 through July 25, the previous balance was \$396.68, and the annual interest rate is 19.5%.

June 30	payment	\$100.00
July 2	gasoline	\$36.19
July 10	restaurant	\$53.00

- ▶ **39.** The activity on Denise Hellings' Sears account for one billing period is shown below. Find the average daily balance and the finance charge if the billing period is March 1 through March 31, the previous balance was \$157.14, and the annual interest rate is 21%.

March 5	payment	\$25.00
March 17	tools	\$36.12

- ▶ **40.** The activity on Charlie Wilson's Visa account for one billing period is shown below. Find the average daily balance and the finance charge if the billing period is November 11 through December 10, the previous balance was \$642.38, and the annual interest rate is 20%.

November 15	payment	\$150
November 28	office supplies	\$23.82
December 1	toy store	\$312.58

- 41.** Donovan and Pam Hamilton bought a house from Edward Gurney for \$162,500. In lieu of a 10% down payment, Mr. Gurney accepted 5% down at the time of the sale and a promissory note from the Hamiltons for the remaining 5%, due in four years. The Hamiltons also agreed to make monthly interest payments to Mr. Gurney at 10% interest until the note expires. The Hamiltons obtained a loan from their bank for the remaining 90% of the purchase price. The bank in turn paid the sellers the remaining 90% of the purchase price, less a sales commission of 6% of the purchase price, paid to the sellers' and the buyers' real estate agents.

- Find the Hamiltons' down payment.
- Find the amount that the Hamiltons borrowed from their bank.
- Find the amount that the Hamiltons borrowed from Mr. Gurney.
- Find the Hamilton's monthly interest-only payment to Mr. Gurney.
- Find Mr. Gurney's total income from all aspects of the down payment (including the down payment, the amount borrowed under the promissory note, and the monthly payments required by the promissory note).

- Find Mr. Gurney's net income from the Hamiltons' bank.
- Find Mr. Gurney's total income from all aspects of the sale.

- ▶ **42.** George and Peggy Fulwider bought a house from Sally Sinclair for \$233,500. In lieu of a 10% down payment, Ms. Sinclair accepted 5% down at the time of the sale and a promissory note from the Fulwiders for the remaining 5%, due in four years. The Fulwiders also agreed to make monthly interest payments to Ms. Sinclair at 10% interest until the note expires. The Fulwiders obtained a loan from their bank for the remaining 90% of the purchase price. The bank in turn paid the sellers the remaining 90% of the purchase price, less a sales commission of 6% of the purchase price, paid to the sellers' and the buyers' real estate agents.

- Find the Fulwiders' down payment.
- Find the amount that the Fulwiders borrowed from their bank.
- Find the amount that the Fulwiders borrowed from Ms. Sinclair.
- Find the Fulwiders' monthly interest-only payment to Ms. Sinclair.
- Find Ms. Sinclair's total income from all aspects of the down payment (including the down payment, the amount borrowed under the promissory note, and the monthly payments required by the promissory note).
- Find Ms. Sinclair's net income from the Fulwiders' bank.
- Find Ms. Sinclair's total income from all aspects of the sale.

- ▶ **43.** The Obamas bought a house from the Bushes for \$389,400. In lieu of a 20% down payment, the Bushes accepted a 10% down payment at the time of the sale and a promissory note from the Obamas for the remaining 10%, due in 4 years. The Obamas also agreed to make monthly interest payments to the Bushes at 11% interest until the note expires. The Obamas obtained a loan for the remaining 80% of the purchase price from their bank. The bank in turn paid the sellers the remaining 80% of the purchase price, less a sales commission of 6% of the sales price paid to the sellers' and the buyers' real estate agents.

- Find the Obamas' down payment.
- Find the amount that the Obamas borrowed from their bank.
- Find the amount that the Obamas borrowed from the Bushes.
- Find the Obamas' monthly interest-only payment to the Bushes.
- Find the Bushes' total income from all aspects of the down payment (including the down payment, the amount borrowed under the promissory note, and the monthly payments required by the note).

- f. Find the Bushes' income from the Obamas' bank.
- g. Find the Bushes' total income from all aspects of the sale.
44. Sam Needham bought a house from Sheri Silva for \$238,300. In lieu of a 20% down payment, Ms. Silva accepted a 10% down payment at the time of the sale and a promissory note from Mr. Needham for the remaining 10%, due in four years. Mr. Needham also agreed to make monthly interest payments to Ms. Silva at 9% interest until the note expires. Mr. Needham obtained a loan for the remaining 80% of the purchase price from his bank. The bank in turn paid Ms. Silva the remaining 80% of the purchase price, less a sales commission (of 6% of the sales price) paid to the sellers' and the buyers' real estate agents.
- Find Mr. Needham's down payment.
 - Find the amount that Mr. Needham borrowed from his bank.
 - Find the amount that Mr. Needham borrowed from the Ms. Silva.
 - Find Mr. Needham's monthly interest-only payment to Ms. Silva.
 - Find Ms. Silva's total income from all aspects of the down payment (including the down payment, the amount borrowed under the promissory note, and the monthly payments required by the note).
 - Find Ms. Silva's income from Mr. Needham's bank.
 - Find Ms. Silva's total income from all aspects of the sale.
45. Your credit card has a balance of \$1,000. Its interest rate is 21%. You have stopped using the card, because you don't want to go any deeper into debt. Each month, you make the minimum required payment of \$20.
- During the January 10 through February 9 billing period, you pay the minimum required payment on January 25th. Find the average daily balance, the finance charge and the new balance. (The new balance includes the finance charge.)
 - During the February 10 through March 9 billing period, you pay the minimum required payment on February 25th. Find the average daily balance, the finance charge and the new balance. (The new balance includes the finance charge.)
 - During the March 10 through April 9 billing period, you pay the minimum required payment on March 25th. Find the average daily balance, the finance charge and the new balance. (The new balance includes the finance charge.)
 - Discuss the impact of making the minimum required payment, both on yourself and on the credit card issuer.
- ▶ 46. Your credit card has a balance of \$1,200. Its interest rate is 20.5%. You have stopped using the card, because you don't want to go any deeper into debt. Each month, you make the minimum required payment of \$24.
- During the September 10 through October 9 billing period, you pay the minimum required payment on September 25th. Find the average daily balance, the finance charge and the new balance. (The new balance includes the finance charge.)
 - During the October 10 through November 9 billing period, you pay the minimum required payment on October 25th. Find the average daily balance, the finance charge and the new balance. (The new balance includes the finance charge.)
 - During the November 10 through December 9 billing period, you pay the minimum required payment on November 25th. Find the average daily balance, the finance charge and the new balance. (The new balance includes the finance charge.)
 - Discuss the impact of making the minimum required payment, both on yourself and on the credit card issuer.
- ▶ 47. Your credit card has a balance of \$1,000. Its interest rate is 21%. You have stopped using the card, because you don't want to go any deeper into debt. Each month, you make the minimum required payment. Your credit card issuer recently changed their minimum required payment policy, in response to the Bankruptcy Abuse Prevention and Consumer Protection Act of 2005. As a result, your minimum required payment is \$40.
- During the January 10 through February 9 billing period, you pay the minimum required payment on January 25th. Find the average daily balance, the finance charge and the new balance. (The new balance includes the finance charge.)
 - During the February 10 through March 9 billing period, you pay the minimum required payment on February 25th. Find the average daily balance, the finance charge and the new balance. (The new balance includes the finance charge.)
 - During the March 10 through April 9 billing period, you pay the minimum required payment on March 25th. Find the average daily balance, the finance charge and the new balance. (The new balance includes the finance charge.)
 - Compare the results of parts (a) through (c) with those of Exercise 45. Discuss the impact of the credit card issuer's change in their minimum required payment policy.
- ▶ 48. Your credit card has a balance of \$1,200. Its interest rate is 20.5%. You have stopped using the card, because you don't want to go any deeper into debt. Each month, you make the minimum required payment. Your credit card issuer recently changed their minimum required payment policy, in response to the Bankruptcy Abuse Prevention and Consumer Protection Act of 2005. As a result, your minimum required payment is \$48.

- a. During the September 10 through October 9 billing period, you pay the minimum required payment on September 25th. Find the average daily balance, the finance charge and the new balance. (The new balance includes the finance charge.)
- b. During the October 10 through November 9 billing period, you pay the minimum required payment on October 25th. Find the average daily balance, the finance charge and the new balance. (The new balance includes the finance charge.)
- c. During the November 10 through December 9 billing period, you pay the minimum required payment on November 25th. Find the average daily balance, the finance charge and the new balance. (The new balance includes the finance charge.)
- d. Compare the results of parts (a) through (c) with those of Exercise 46. Discuss the impact of the credit card issuer's change in their minimum required payment policy.



Answer the following questions using complete sentences and your own words.

• CONCEPT QUESTIONS

49. Could Exercises 5–21 all be done with the Simple Interest Formula? If so, how? Could Exercises 5–21 all be done with the Simple Interest Future Value Formula? If so, how? Why do we have both formulas?
50. Which is always higher: future value or principal? Why?

• HISTORY QUESTIONS

51. Who offered the first credit card?
52. What was the first post–World War II credit card?
53. Who created the first post–World War II credit card?
54. What event prompted the creation of the first post–World War II credit card?
55. What was the first interstate bank card?

5.2 Compound Interest

OBJECTIVES

- Understand the difference between simple interest and compound interest
- Use the Compound Interest Formula
- Understand and compute the annual yield

Many forms of investment, including savings accounts, earn **compound interest**, in which interest is periodically paid on both the original principal and previous interest payments. This results in earnings that are significantly higher over a longer period of time. It is important that you understand this difference in order to make wise financial decisions. We'll explore compound interest in this section.

Compound Interest as Simple Interest, Repeated

EXAMPLE 1

UNDERSTANDING COMPOUND INTEREST Tom and Betty deposit \$1,000 into their new bank account. The account pays 8% interest compounded quarterly. This means that interest is computed and deposited every quarter of a year. Find the account balance after six months, using the Simple Interest Future Value formula to compute the balance at the end of each quarter.

SOLUTION

At the end of the first quarter, $P = \$1,000$, $r = 8\% = 0.08$, and $t =$ one quarter or $\frac{1}{4}$ year.

$$\begin{aligned}
 FV &= P(1 + rt) && \text{the Simple Interest Future Value Formula} \\
 &= 1,000(1 + 0.08 \cdot \frac{1}{4}) && \text{substituting} \\
 &= 1,000(1 + 0.02) \\
 &= \$1,020
 \end{aligned}$$

This means that there is \$1,020 in Tom and Betty's account at the end of the first quarter. It also means that the second quarter's interest will be paid on this new principal. So at the end of the second quarter, $P = \$1,020$ and $r = 0.08$. Note that $t = \frac{1}{4}$, not $\frac{2}{4}$, because we are computing interest for one quarter.

$$\begin{aligned}
 FV &= P(1 + rt) && \text{the Simple Interest Future Value Formula} \\
 &= 1,020(1 + 0.08 \cdot \frac{1}{4}) && \text{substituting} \\
 &= 1,020(1 + 0.02) \\
 &= \$1,040.40
 \end{aligned}$$

At the end of six months, the account balance is \$1,040.40.



As a rough check, notice that the future value is slightly higher than the principal, as it should be.

The Compound Interest Formula

This process would become tedious if we were computing the balance after twenty years. Because of this, compound interest problems are usually solved with their own formula.

The **compounding period** is the time period over which any one interest payment is calculated. In Example 1, the compounding period was a quarter of a year. For each quarter's calculation, we multiplied the annual rate of 8% (0.08) by the time 1 quarter ($\frac{1}{4}$ year) and got 2% (0.02). This 2% is the **quarterly rate** (or more generally, the *periodic rate*). A **periodic rate** is any rate that is prorated in this manner from an annual rate.

If i is the periodic interest rate, then the future value at the end of the first period is

$$FV = P(1 + i)$$

Because this is the account balance at the beginning of the second period, it becomes the new principal. The account balance at the end of the second period is

$$\begin{aligned}
 FV &= P(1 + i) \cdot (1 + i) && \text{substituting } P(1 + i) \text{ for } P \\
 &= P(1 + i)^2
 \end{aligned}$$

This means that $P(1 + i)^2$ is the account balance at the beginning of the third period, and the future value at the end of the third period is

$$\begin{aligned}
 FV &= [P(1 + i)^2] \cdot (1 + i) && \text{substituting } P(1 + i)^2 \text{ for } P \\
 &= P(1 + i)^3
 \end{aligned}$$

If we generalize these results, we get the Compound Interest Formula.

COMPOUND INTEREST FORMULA

If initial principal P earns compound interest at a periodic interest rate i for n periods, the future value is

$$FV = P(1 + i)^n$$

Notice that we have maintained our variables tradition: i and n do not measure amounts of money, so they are not capital letters. We now have three interest-related variables:

- r , the annual interest rate (not an amount of money)
- i , the periodic interest rate (not an amount of money)
- I , the interest itself (an amount of money)

EXAMPLE 2

SOLUTION

USING THE COMPOUND INTEREST FORMULA Use the Compound Interest Formula to recompute Tom and Betty's account balance from Example 1.

Their 8% interest is compounded quarterly, so each quarter they earn a quarter of $8\% = \frac{1}{4} \cdot 8\% = 2\% = 0.02$. Also, n counts the number of quarters, so $n = 2$.

$$\begin{aligned} FV &= P(1 + i)^n && \text{the Compound Interest Formula} \\ &= 1,000(1 + 0.02)^2 = \$1,040.40 && \text{substituting} \end{aligned}$$



$$1000 \times (1 + .02)^2 =$$

$$1000 (1 + .02)^2 \text{ ENTER}$$

Compound Interest Compared with Simple Interest

As we saw at the beginning of this section, compound interest is just simple interest, repeated. However, there can be profound differences in their results.

EXAMPLE 3

COMPOUND INTEREST OVER A LONG PERIOD OF TIME In 1777, it looked as though the Revolutionary War was about to be lost. George Washington's troops were camped at Valley Forge. They had minimal supplies, and the winter was brutal. According to a 1990 class action suit, Jacob DeHaven, a wealthy Pennsylvania merchant, saved Washington's troops and the revolutionary cause by loaning Washington \$450,000. The suit, filed by DeHaven's descendants, asked the government to repay the still-outstanding loan plus compound interest at the then-prevailing rate of 6%. (Source: *New York Times*, May 27, 1990.) How much did the government owe on the 1990 anniversary of the loan if the interest is compounded monthly?

SOLUTION

The principal is $P = \$450,000$. If 6% interest is paid each year, then $1/12$ th of 6% is paid each month, and $i = \frac{1}{12} \cdot 6\% = 0.06/12$. The term is

$$\begin{aligned} n &= 213 \text{ years} \\ &= 213 \text{ years} \cdot \frac{12 \text{ months}}{1 \text{ year}} && \text{using dimensional analysis} \\ &= 2,556 \text{ months} \end{aligned}$$

Using the Compound Interest Formula, we get

$$\begin{aligned}
 FV &= P(1 + i)^n && \text{the Compound Interest Formula} \\
 &= 450,000(1 + 0.06/12)^{2556} && \text{substituting} \\
 &= 1.547627234 \dots \times 10^{11} \\
 &\approx \$154,762,723,400 && \text{rounding}
 \end{aligned}$$



$$450000 \times (1 + .06 \div 12)^{2556} =$$

$$450000 (1 + .06 \div 12)^{2556} \text{ ENTER}$$

The DeHavens sued for this amount, but they also stated that they were willing to accept a more reasonable payment.



As a rough check, notice that the future value is much higher than the principal, as it should be because of the long time period.

Some sources have questioned the DeHavens' claim: "There is also no evidence to support the claim of the DeHaven family that their ancestor Jacob DeHaven lent George Washington \$450,000 in cash and supplies while the army was encamped at Valley Forge. This tradition first appeared in print in a history of the DeHaven family penned by Howard DeHaven Ross. Periodically, the descendants of Jacob DeHaven make attempts to get the "loan" repaid with interest. . . . This remarkably persistent tradition has been thoroughly debunked by Judith A. Meier, of the Historical Society of Montgomery County, whose genealogical research revealed that there were no DeHavens living in the immediate area until after 1790 and that Jacob DeHaven had never been rich enough to make such a fabulous loan." (*Source*: Lorette Treese, *Valley Forge: Making and Remaking a National Symbol*, University Park, PA: Pennsylvania University Press, 1995.)

EXAMPLE 4

COMPARING SIMPLE INTEREST WITH COMPOUND INTEREST OVER A LONG PERIOD OF TIME How much would the government have owed the DeHavens if the interest was simple interest?

SOLUTION

With simple interest, we use r and t , which are *annual* figures, rather than i and n , which are *periodic* figures. So $r = 6\% = 0.06$, and $t = 213$ years.

$$\begin{aligned}
 FV &= P(1 + rt) && \text{the Simple Interest Future Value Formula} \\
 &= 450,000(1 + 0.06 \cdot 213) && \text{substituting} \\
 &= \$6,201,000
 \end{aligned}$$



$$450000 \times (1 + .06 \times 213) =$$

$$450000 (1 + .06 \times 213) \text{ ENTER}$$

Simple interest would have required a payment of only \$6 million. This is a lot, but not in comparison with the \$155 billion payment required by compound interest.

In Examples 3 and 4, the future value with compound interest is almost 25,000 times the future value with simple interest. Over longer periods of time, compound interest is immensely more profitable to the investor than simple interest, because compound interest gives interest on interest. Similarly, compounding more frequently (daily rather than quarterly, for example) is more profitable to the investor.

The effects of the size of the time interval and the compounding period can be seen in Figure 5.4.

Type of Interest	Future Value of \$1000 at 10% Interest		
	After 1 Year	After 10 Years	After 30 Years
Simple interest	\$1100.00	\$2000.00	\$ 4,000.00
Compounded annually	\$1100.00	\$2593.74	\$17,449.40
Compounded quarterly	\$1103.81	\$2685.06	\$19,358.15
Compounded monthly	\$1104.71	\$2707.04	\$19,837.40
Compounded daily	\$1105.16	\$2717.91	\$20,077.29

FIGURE 5.4 Comparing simple interest and compound interest.

Finding the Interest and the Present Value

EXAMPLE 5

FINDING THE AMOUNT OF INTEREST EARNED Betty's boss paid her an unexpected bonus of \$2,500. Betty and her husband decided to save the money for their daughter's education. They deposited it in account that pays 10.3% interest compounded daily. Find the amount of interest that they would earn in fifteen years by finding the future value and subtracting the principal.

SOLUTION

Compounding daily, we have $P = 2500$, $i = 1/365$ th of 10.3% = $0.103/365$, and $n = 15$ years = 15 years \cdot 365 days/year = 5,475 days.

$$\begin{aligned}
 FV &= P(1 + i)^n && \text{the Compound Interest Formula} \\
 &= 2,500(1 + 0.103/365)^{5475} && \text{substituting} \\
 &= 11,717.374 \dots \\
 &\approx \$11,717.37
 \end{aligned}$$



$$2500 \times (1 + .103 \div 365)^{5475} =$$

$$2500 (1 + .103 \div 365)^{5475} \text{ ENTER}$$

Warning: If you compute $\frac{0.103}{365}$ separately, you will get a long decimal. Do not round off that decimal, because the resulting answer will be inaccurate. Doing the calculation all at once (as shown above) avoids this difficulty.

The principal is \$2,500, and the total of principal and interest is \$11,717.37. Thus, the interest is $\$11,717.37 - \$2500 = \$9,217.37$.

EXAMPLE 6

FINDING THE PRESENT VALUE Find the amount of money that must be invested now at $7\frac{3}{4}\%$ interest compounded annually so that it will be worth \$2,000 in 3 years.

SOLUTION

The question actually asks us to find the present value, or principal P , that will generate a future value of \$2,000. We have $FV = 2,000$, $i = 7\frac{3}{4}\% = 0.0775$, and $n = 3$.

$$\begin{aligned} FV &= P(1 + i)^n && \text{the Compound Interest Formula} \\ 2,000 &= P(1 + 0.0775)^3 && \text{substituting} \\ P &= \frac{2,000}{1.0775^3} && \text{solving for } P \\ &= 1,598.7412 && \text{rounding} \\ &\approx \$1,598.74 \end{aligned}$$



As a rough check, notice that the principal is somewhat lower than the future value, as it should be, because of the short time period.

Annual Yield



FIGURE 5.5 Most banks advertise their yields as well as their rates.

Which investment is more profitable: one that pays 5.8% compounded daily or one that pays 5.9% compounded quarterly? It is difficult to tell. Certainly, 5.9% is a better rate than 5.8%, but compounding daily is better than compounding quarterly. The two rates cannot be directly compared because of their different compounding frequencies. The way to tell which is the better investment is to find the annual yield of each.

The **annual yield** (also called the **annual percentage yield** or **APY**) of a compound interest deposit is the *simple interest rate* that has the same future value as the compound rate would have in one year. The annual yields of two different investments can be compared, because they are both simple interest rates. Annual yield provides the consumer with a uniform basis for comparison and banks display both their interest rates and their annual yields, as shown in Figure 5.5. The annual yield should be slightly higher than the compound rate, because compound interest is slightly more profitable than simple interest over a short period of time. The compound rate is sometimes called the **nominal rate** to distinguish it from the yield (here, *nominal* means “named” or “stated”).

To find the **annual yield** r of a given compound interest rate, you find the simple interest rate that makes the future value under simple interest the same as the future value under compound interest in one year.

$$FV(\text{simple interest}) = FV(\text{compound interest})$$

$$P(1 + r) = P(1 + i)^n$$

EXAMPLE 7

SOLUTION

FINDING THE ANNUAL YIELD Find the annual yield of \$2,500 deposited in an account in which it earns 10.3% interest compounded daily for 15 years.

Simple interest

$$P = 2,500$$

$r =$ unknown annual yield

$$t = 1 \text{ year}$$

(one year, not fifteen years—its *annual* yield)

$$FV(\text{simple interest}) = FV(\text{compounded monthly})$$

$$P(1 + rt) = P(1 + i)^n$$

$$2,500(1 + r \cdot 1) = 2,500(1 + 0.103/365)^{365}$$

$$(1 + r \cdot 1) = (1 + 0.103/365)^{365}$$

$$r = (1 + 0.103/365)^{365} - 1$$

$$r = 0.10847 \dots \approx 10.85\%$$

Compound interest

$$P = 2,500$$

$$i = \frac{1}{365} \text{th of } 10.3\% = \frac{0.103}{365}$$

$$n = 1 \text{ year} = 365 \text{ days}$$

**substituting
dividing by 2,500
solving for r
rounding**



As a rough check, notice that the annual yield is slightly higher than the compound rate, as it should be.

By tradition, we round this to the nearest hundredth of a percent, so the annual yield is 10.85%. This means that in one year's time, 10.3% compounded daily has the same effect as 10.85% simple interest. For any period of time longer than a year, 10.3% compounded daily will yield *more* interest than 10.85% simple interest would, because compound interest gives interest on interest.

Notice that in Example 7, the principal of \$2,500 canceled out. If the principal were two dollars or two million dollars, it would still cancel out, and the annual yield of 10.3% compounded daily would still be 10.85%. The principal does not matter in computing the annual yield. Also notice that the fifteen years did not enter into the calculation—*annual* yield is always based on a *one-year* period.

EXAMPLE 8

SOLUTION

FINDING THE ANNUAL YIELD Find the annual yield corresponding to a nominal rate of 8.4% compounded monthly.

We are told neither the principal nor the time, but (as discussed above) these variables do not affect the annual yield.

Compounding monthly, we have $i = 1/12$ of $8.4\% = \frac{0.084}{12}$, $n = 1$ year = 12 months, and $t = 1$ year.

$$FV(\text{simple interest}) = FV(\text{compounded monthly})$$

$$P(1 + rt) = P(1 + i)^n$$

$$(1 + rt) = (1 + i)^n$$

$$(1 + r \cdot 1) = (1 + .084/12)^{12}$$

$$r = (1 + .084/12)^{12} - 1$$

$$r = 0.08731 \dots \approx 8.73\%$$

**dividing by P
annual, so $t = 1$ year and
 $n = 12$ months
solving for r
rounding**



As a rough check, notice that the annual yield is slightly higher than the compound rate, as it should be.



TOPIC X

BENJAMIN FRANKLIN'S GIFT COMPOUND INTEREST IN THE REAL WORLD

In 1789, when Benjamin Franklin was 83, he added a codicil to his will. That codicil was meant to aid young people, in Boston (where Franklin grew up), and Philadelphia (where he had been President of the state of Pennsylvania) over a time span of two hundred years.

In his codicil, Franklin wrote of his experience as an apprentice printer, of the friends who had loaned him the money to set up his own printing business, and of the importance of craftsmen to a city. He then went on to say, "To this End I devote Two thousand Pounds Sterling, which I give, one thousand thereof to the Inhabitants of the Town of Boston, in Massachusetts, and the other thousand to the Inhabitants of the City of Philadelphia, in Trust to and for the Uses, Interests and Purposes hereinafter mentioned and declared." At that time, £1,000 was the equivalent of about \$4,500.

Franklin's plan called for each town to use the money as a loan fund for 100 years. The money would be loaned out to young tradesmen to help them start their own businesses. He calculated that at the end of 100 years, each fund would have grown to £131,000 (or about \$582,000 in



Stock Montage/Stock Montage/Getty Images

that at the end of 200 years, each of the two city's loan funds would have grown to £4,061,000 (or about \$7,000,000 in 1992 dollars). At that point, he called for the money to be given to the two cities and states.

Twenty-two tradesmen, including bricklayers, hairdressers, jewelers, and tanners, applied for loans from Franklin's fund in the first month after Franklin's death. The fund remained popular until the onset of the Industrial Revolution in the early 1800s, when young people stopped becoming tradesmen with their own shops. Instead, most went to work as mechanics in factories.

When Boston's fund reached its hundredth anniversary in 1891, its value was approximately \$391,000. One-fourth of that money continued to be used for a loan fund, as Franklin

1892 dollars). At that point, one-fourth of the money would continue to be used for a loan fund. The remainder would be used for public works.

He calculated

wished. The city of Boston decided to use the remainder to build a trade school, because it fit with Franklin's goal of helping young people.

Legal problems delayed the school's founding, but the Benjamin Franklin Institute of Technology was opened in Boston in 1908. At that point, the trade school part of the Boston fund had risen to \$432,367, and the loan fund part had grown to \$163,971.

The Institute continues to operate today. It has almost 400 students, 90% of whom receive financial aid. It awards Bachelor of Science degrees, Associate in Engineering degrees, and Associate in Science degrees. Ninety-eight percent of its graduates find work in their fields within six months of graduation.

Pennsylvania used its share of the money to fund the Franklin Institute of Philadelphia. Originally, the institute promoted the "mechanical arts." Now it houses a planetarium, an IMAX theater, and a science museum.

Exercises 57–62 explore some aspects of Franklin's bequest.

Source: The Benjamin Franklin Institute of Technology and the Franklin Institute of Philadelphia.

In Example 8, we found that 8.4% compounded monthly generates an annual yield of 8.73%. This means that 8.4% compounded monthly has the same effect as does 8.73% simple interest in one year's time. Furthermore, as Figure 5.6 indicates, 8.4% compounded monthly has the same effect as does 8.73% compounded annually for any time period.

For a Principal of \$1,000	After 1 Year	After 10 Years
<i>FV</i> at 8.4% compounded monthly	\$1,087.31	\$2,309.60
<i>FV</i> at 8.73% simple interest	\$1,087.30	\$1,873.00
<i>FV</i> at 8.73% compounded annually	\$1,087.30	\$2,309.37

FIGURE 5.6 What annual yield means.

Notice that all three rates have the same future value after one year (the 1¢ difference is due to rounding off the annual yield to 8.73%). However, after ten years, the simple interest has fallen way behind, while the 8.4% compounded monthly and 8.73% compounded annually remain the same (except for the round-off error). This always happens. The annual yield is the simple interest rate that has the same future value that the compound rate would have in one year. It is also the annually compounded rate that has the same future value that the nominal rate would have after any amount of time.

An annual yield formula does exist, but the annual yield can be calculated efficiently without it, as was shown above. The formula is developed in the exercises. See Exercise 51.

5.2 EXERCISES

In Exercises 1–6, find the periodic rate that corresponds to the given compound rate, if the rate is compounded (a) quarterly, (b) monthly, (c) daily, (d) biweekly (every two weeks), and (e) semimonthly (twice a month). Do not round off the periodic rate.

- ▶ 1. 12% 2. 6%
 3. 3.1% 4. 6.8%
 5. 9.7% ▶ 6. 10.1%

In Exercises 7–10, find the number of periods that corresponds to the given time span, if a period is (a) a quarter of a year, (b) a month, and (c) a day. (Ignore leap years.)

- ▶ 7. $8\frac{1}{2}$ years ▶ 8. $9\frac{3}{4}$ years
 ▶ 9. 30 years 10. 45 years

In Exercises 11–16, (a) find and (b) interpret the future value of the given amount.

- ▶ 11. \$3,000 at 6% compounded annually for 15 years
 12. \$7,300 at 7% compounded annually for 13 years
 ▶ 13. \$5,200 at $6\frac{3}{4}$ % compounded quarterly for $8\frac{1}{2}$ years
 14. \$36,820 at $7\frac{7}{8}$ % compounded quarterly for 4 years
 15. \$1,960 at $4\frac{1}{8}$ % compounded daily for 17 years (ignore leap years)
 ▶ 16. \$12,350 at 6% compounded daily for 10 years and 182 days (ignore leap years)

In Exercises 17–20, (a) find and (b) interpret the annual yield corresponding to the given nominal rate.

- ▶ 17. 8% compounded monthly
 18. $5\frac{1}{2}$ % compounded quarterly
 19. $4\frac{1}{4}$ % compounded daily
 ▶ 20. $12\frac{5}{8}$ % compounded daily

In Exercises 21 and 22, find and interpret the annual yield corresponding to the given nominal rate.

- ▶ 21. 10% compounded (a) quarterly, (b) monthly, and (c) daily

22. $12\frac{1}{2}$ % compounded (a) quarterly, (b) monthly, and (c) daily

In Exercises 23–26, (a) find and (b) interpret the present value that will generate the given future value.

- ▶ 23. \$1,000 at 8% compounded annually for 7 years
 24. \$9,280 at $9\frac{3}{4}$ % compounded monthly for 2 years and 3 months
 ▶ 25. \$3,758 at $11\frac{7}{8}$ % compounded monthly for 17 years and 7 months
 ▶ 26. \$4,459 at $10\frac{3}{4}$ % compounded quarterly for 4 years
 27. \$10,000 is deposited in an account in which it earns 10% interest compounded monthly. No principal or interest is withdrawn from the account. Instead, both continue to earn interest over time. Find the account balance after six months:
 a. using the simple interest future value formula to compute the balance at the end of each compounding period.
 b. using the compound interest formula.
 ▶ 28. \$20,000 is deposited in an account in which it earns 9.5% interest compounded quarterly. No principal or interest is withdrawn from the account. Instead, both continue to earn interest over time. Find the account balance after one year:
 a. using the simple interest future value formula to compute the balance at the end of each compounding period.
 b. using the compound interest formula.
 29. \$15,000 is deposited in an account in which it earns 6% interest compounded annually. No principal or interest is withdrawn from the account. Instead, both continue to earn interest over time. Find the account balance after three years:
 a. using the simple interest future value formula to compute the balance at the end of each compounding period.
 b. using the compound interest formula.

▶ Selected exercises available online at www.webassign.net/brookscole

30. \$30,000 is deposited in an account in which it earns 10% interest compounded annually. No principal or interest is withdrawn from the account. Instead, both continue to earn interest over time. Find the account balance after four years:
- using the simple interest future value formula to compute the balance at the end of each compounding period.
 - using the compound interest formula.
- ▶ 31. Donald Trumptobe decided to build his own dynasty. He is considering specifying in his will that at his death, \$10,000 would be deposited into a special account that would earn a guaranteed 6% interest compounded daily. This money could not be touched for 100 years, at which point it would be divided among his heirs. Find the future value.
- ▶ 32. How much would Donald Trumptobe in Exercise 31 have to have deposited so that his heirs would have \$50,000,000 or more in 100 years if his money earns 7% compounded monthly?
- ▶ 33. Donald Trumptobe in Exercise 31 predicts that in 100 years, he will have four generations of offspring (that is, children, grandchildren, great-grandchildren, and great-great-grandchildren). He estimates that each person will have two children. How much will he have to have deposited so that each of his great-great-grandchildren would have \$1,000,000 or more in 100 years if his money earns 7.5% compounded monthly?
34. Donald Trumptobe in Exercise 31 predicts that in 100 years, he will have four generations of offspring (that is, children, grandchildren, great-grandchildren, and great-great-grandchildren). He estimates that each person will have two children. How much will he have to have deposited so that each of his great-great-grandchildren would have \$1,000,000 or more in 100 years if his money earns 9.25% compounded daily?
- ▶ 35. When Jason Levy was born, his grandparents deposited \$3,000 into a special account for Jason's college education. The account earned $6\frac{1}{2}\%$ interest compounded daily.
- How much will be in the account when Jason is 18?
 - If, on turning 18, Jason arranges for the monthly interest to be sent to him, how much will he receive each thirty-day month?
 - How much would be in the account when Jason turns 18 if his grandparents started Jason's savings account on his tenth birthday?
36. When Alana Cooper was born, her grandparents deposited \$5,000 into a special account for Alana's college education. The account earned $7\frac{1}{4}\%$ interest compounded daily.
- How much will be in the account when Alana is 18?
 - If, on turning 18, Alana arranges for the monthly interest to be sent to her, how much will she receive each thirty-day month?
 - How much would be in the account when Alana turns 18 if his grandparents started Alana's savings account on her tenth birthday?
- For Exercises 37–40, note the following information: An Individual Retirement Account (IRA) is an account in which the saver does not pay income tax on the amount deposited but is not allowed to withdraw the money until retirement. (The saver pays income tax at that point, but his or her tax bracket is much lower then.)*
37. At age 27, Lauren Johnson deposited \$1,000 into an IRA, in which it earns $7\frac{1}{8}\%$ compounded monthly.
- What will it be worth when she retires at 65?
 - How much would the IRA be worth if Lauren didn't set it up until she was 35?
- ▶ 38. At age 36, Dick Shoemaker deposited \$2,000 into an IRA, in which it earns $8\frac{1}{8}\%$ compounded semiannually.
- What will it be worth when he retires at 65?
 - How much would the IRA be worth if Dick didn't set it up until he was 48?
- ▶ 39. Marlene Silva wishes to have an IRA that will be worth \$100,000 when she retires at age 65.
- How much must she deposit at age 35 at $8\frac{3}{8}\%$ compounded daily?
 - If, at age 65, she arranges for the monthly interest to be sent to her, how much will she receive each thirty-day month?
40. David Murtha wishes to have an IRA that will be worth \$150,000 when he retires at age 65.
- How much must he deposit at age 26 at $6\frac{1}{8}\%$ compounded daily?
 - If, at age 65, he arranges for the monthly interest to be sent to him, how much will he receive each thirty-day month?
- For Exercises 41–46, note the following information: A certificate of deposit (CD) is an agreement between a bank and a saver in which the bank guarantees an interest rate and the saver commits to leaving his or her deposit in the account for an agreed-upon period of time.*
- ▶ 41. First National Bank offers two-year CDs at 9.12% compounded daily, and Citywide Savings offers two-year CDs at 9.13% compounded quarterly. Compute the annual yield for each institution and determine which is more advantageous for the consumer.
- ▶ 42. National Trust Savings offers five-year CDs at 8.25% compounded daily, and Bank of the Future offers five-year CDs at 8.28% compounded annually. Compute the annual yield for each institution, and determine which is more advantageous for the consumer.
43. Verify the annual yield for the five-year certificate quoted in the bank sign in Figure 5.7 on page 352, using interest that is compounded daily and:
- 365-day years
 - 360-day years
 - Various combinations of 360-day and 365-day years

Product	Minimum	Interest Rate %	Annual Percentage Yield %
PASSBOOK SAVINGS		1.470	1.500
STATEMENT SAVINGS		1.470	1.500
INSURED Money Market Account		1.000	1.020
COMMERCIAL Money Market		0.750	0.760
3 MONTH Certificate		0.980	1.000
6 MONTH Certificate		1.030	1.050
12 MONTH Certificate		1.180	1.200
2 YEAR Certificate		1.320	1.350
30 MONTH Certificate		1.950	2.000
3 YEAR Certificate		1.570	1.600
4 YEAR Certificate		2.190	2.250
5 YEAR Certificate		3.390	3.500

*Ask us for further information about these accounts. Penalty for early withdrawal. Truth in savings disclosures available upon request.

© Robert Bremner / Photo Edit

FIGURE 5.7 A bank sign for Exercise 43.

- ▶ 44. Verify the yield for the one-year CDs quoted in Figure 5.8.

FIGURE 5.8 A savings bank advertisement for Exercise 44.

45. Verify the yield for the two-year CDs quoted in Figure 5.9.

FIGURE 5.9 A savings bank advertisement for Exercise 45.

46. Verify the yield for the six-month CDs quoted in the savings bank advertisement on page 347.

- ▶ 47. Recently, Bank of the West offered six-month CDs at 5.0% compounded monthly.
- Find the annual yield of one of these CDs.
 - How much would a \$1,000 CD be worth at maturity?
 - How much interest would you earn?
 - What percent of the original \$1,000 is this interest?
 - The answer to part (d) is not the same as that of part (a). Why?
 - The answer to part (d) is close to, but not exactly half, that of part (a). Why?
- (Source: Bank of the West.)

- ▶ 48. Recently, Bank of the West offered six-month CDs at 5.0% interest compounded monthly and one-year CDs at 5.20% interest compounded monthly. Maria Ruiz bought a six-month \$2,000 CD, even though she knew she would not need the money for at least a year, because it was predicted that interest rates would rise.

- Find the future value of Maria's CD.
- Six months later, Maria's CD has come to term, and in the intervening time, interest rates have risen. She reinvests the principal and interest from her first CD in a second six-month CD that pays 5.31% interest compounded monthly. Find the future value of Maria's second CD.
- Would Maria have been better off if she had bought a one-year CD instead of two six-month CDs?
- If Maria's second CD pays 5.46% interest compounded monthly, rather than 5.31%, would she be better off with the two six-month CDs or the one-year CD?

(Source: Bank of the West.)

- ▶ 49. CD interest rates vary significantly with time. They hit a historical high in 1981, when the average rate was 16.7%. In 2000, it was 8.1%. In 2009, it was 2%. Find the interest earned by a \$1000 two-year CD with interest compounded quarterly:
- in 1981
 - in 2000
 - in 2009
50. Use the data in Exercise 49 to find the yield of a \$10,000 ten-year CD with interest compounded daily:
- in 1981
 - in 2000
 - in 2009
51. Develop a formula for the annual yield of a compound interest rate.
HINT: Follow the procedure given in Example 8, but use the letters i and n in the place of numbers.)

In Exercises 52–56, use the formula found in Exercise 51 to compute the annual yield corresponding to the given nominal rate.

- ▶ 52. $9\frac{1}{2}\%$ compounded monthly
53. $7\frac{1}{4}\%$ compounded quarterly
54. $12\frac{3}{8}\%$ compounded daily

55. $5\frac{5}{8}\%$ compounded (a) semiannually, (b) quarterly, (c) monthly, (d) daily, (e) biweekly, and (f) semimonthly.
56. $10\frac{1}{2}\%$ compounded (a) semiannually, (b) quarterly, (c) monthly, (d) daily, (e) biweekly, and (f) semimonthly.

Exercises 57–62 refer to Benjamin Franklin's gift, discussed on page 349.

- ▶ 57. Would Benjamin Franklin have used simple or compound interest in projecting the 100-year future value of his bequest? Why? What interest rate did he use in calculating the 100-year future value? Use pounds, not dollars, in your calculation.
- ▶ 58. Would Benjamin Franklin have used simple or compound interest in projecting the 200-year future value of his bequest? Why? What interest rate did he use in calculating the 200-year future value? Use pounds, not dollars, in your calculation.
59. What interest rate did Franklin's Boston bequest actually earn in the first hundred years? Use dollars, not pounds, in your calculation.
- ▶ 60. What interest rate did Franklin's Boston bequest actually earn from 1891 to 1908? Use dollars, not pounds, in your calculation.
61. After 100 years, Franklin's total bequest to Boston was worth much more than his original bequest. The future value was what percentage of the original bequest? Use dollars, not pounds, in your calculation.
- ▶ 62. In 1908, Franklin's total bequest to Boston was worth much more than his original bequest. The future value was what percentage of the original bequest? Use dollars, not pounds, in your calculation.
69. If money earns compound interest, why must the future value be *slightly* higher than the principal, after a short amount of time? Why must the future value be *much* higher than the principal after a long amount of time?
70. *Money* magazine and other financial publications regularly list the top-paying money-market funds, the top-paying bond funds, and the top-paying CDs and their yields. Why do they list yields rather than interest rates and compounding periods?
71. Equal amounts are invested in two different accounts. One account pays simple interest, and the other pays compound interest at the same rate. When will the future values of the two accounts be the same?
72. Suppose you invest some money in a new account that pays 5% interest compounded annually, and you do not make any further deposits into or withdrawals from that account. Which of the following must be true?
- The account grows by the same dollar amount in the second year as it did in the first year.
 - The account grows by a larger dollar amount in the second year as it did in the first year.
 - The account grows by a smaller dollar amount in the second year as it did in the first year. Why?



73. Go to the web sites of four different banks.
- a. For each bank, try to determine the following:
 - The interest rate
 - The compounding frequency
 - The annual yield (also called the *annual percentage yield*, or A.P.Y.) for CDs for two different terms. (Use the same terms for each bank.) Some banks might not give all of the above information, but all will give the annual yield, as required by federal law.
 - b. If a bank omits either the interest rate or the compounding frequency, calculate the omitted information.
 - c. If a bank omits both the interest rate and the compounding frequency, assume that the compounding frequency is daily and calculate the interest rate.
 - d. If a bank omits none of the information, verify the annual yield.
 - e. Which bank offers the best deal? Why?

Some useful links for this web project are listed on the text web site: www.cengage.com/math/johnson

• PROJECT

74. Suppose you have \$1,000 invested at 5% annual interest and you do not make any further deposits into that account. Let x measure years after you



Answer the following questions using complete sentences and your own words.

• CONCEPT QUESTIONS

63. Explain how compound interest is based on simple interest.
64. Why is there no work involved in finding the annual yield of a given simple interest rate?
65. Why is there no work involved in finding the annual yield of a given compound interest rate when that rate is compounded annually?
66. Which should be higher: the annual yield of a given rate compounded quarterly or compounded monthly? Explain why, *without* performing any calculations or referring to any formulas.
67. Why should the annual yield of a given compound interest rate be higher than the compound rate? Why should it be only slightly higher? Explain why, *without* performing any calculations or referring to any formulas.
68. Explain the difference between simple interest and compound interest.

invested the money, and let y measure the future value of the account. Draw a graph that shows the relationship between x and y , for $0 \leq x \leq 5$, if the interest rate is

- simple interest
- compounded annually
- compounded daily

d. Discuss the differences between the three graphs. In your discussion, address the following:

- The difference in their shapes
- Where they coincide
- Which graph is above the others
- Which is below the others



DOUBLING TIME WITH A TI'S TVM APPLICATION

Simple interest is a very straightforward concept. If an account earns 5% simple interest, then 5% of the principal is paid for each year that principal is in the account. In one year, the account earns 5% interest; in two years, it earns 10% interest; in three years, it earns 15% interest, and so on.

It is not nearly so easy to get an intuitive grasp of compound interest. If an account earns 5% interest compounded daily, then it does not earn only 5% interest in one year, and it does not earn only 10% interest in two years.

Annual yield is one way of gaining an intuitive grasp of compound interest. If an account earns 5% interest compounded daily, then it will earn 5.13% interest in 1 year (because the annual yield is 5.13%), but it does not earn merely $2 \cdot 5.13\% = 10.26\%$ interest in two years.

Doubling time is another way of gaining an intuitive grasp of compound interest. **Doubling time** is the amount of time it takes for an account to double in value; that is, it's the amount of time it takes for the future value to become twice the principal. To find the doubling time for an account that earns 5% interest compounded daily, substitute $2P$ for the future value and solve the resulting equation.

$$FV = P(1 + i)^n \quad \text{Compound Interest Formula}$$

$$2P = P \left(1 + \frac{0.05}{365} \right)^n \quad \text{substituting}$$

$$2 = \left(1 + \frac{0.05}{365} \right)^n \quad \text{dividing by } P$$

Solving this equation for n involves mathematics that will be covered in Section 10.0B. For now, we will use the TI-83/84's "Time Value of Money" (TVM) application.

EXAMPLE 9

SOLUTION

FINDING DOUBLING TIME Use a TI-83/84's TVM application to find the doubling time for an account that earns 5% interest compounded daily.

- Press **[APPS]**, select option 1: "Finance", and press **[ENTER]**.
- Select option 1: "TVM Solver", and press **[ENTER]**.
- Enter appropriate values for the variables:

- N is the number of compounding periods. This is the number we're trying to find. We temporarily enter 0. Later, we'll solve for the actual value of N .
- $I\%$ is the *annual* interest rate (not a periodic rate), as a percent, so enter 5 for $I\%$. Note that we do not convert to a decimal or a periodic rate.
- PV is the present value. The size of the present value doesn't matter, so we'll make it \$1. However, it's an outgoing amount of money (since we give it to the bank), so we

```

N=0
I%=5
PV=-1
PMT=0
FV=2
P/Y=365
C/Y=365
PMT:END BEGIN

```

FIGURE 5.10

Preparing the TVM screen.

```

N=5060.320984
I%=5
PV=-1
PMT=0
FV=2
P/Y=365
C/Y=365
PMT:END BEGIN

```

FIGURE 5.11

Solving for N.

enter -1 for PV. You must enter a negative number for any outgoing amount of money.

- PMT is the payment. There are no payments here, so we enter 0.
- FV is the future value. We're looking for the amount of time that it takes the present value to double, so enter 2. This is an incoming amount of money (since the bank gives it to us), so we use 2 rather than -2 .
- P/Y is the number of periods per year. The interest is compounded daily, so there are 365 periods per year. Enter 365 for P/Y.
- C/Y is automatically made to be the same as P/Y. In this text, we will never encounter a situation in which C/Y is different from P/Y.

See Figure 5.10.

4. To solve for N, use the arrow buttons to highlight the 0 that we entered for N. Then press **ALPHA** **SOLVE**. (Pressing **ALPHA** makes the **ENTER** button become the **SOLVE** button.) As a result, we find that N is 5060.320984. See Figure 5.11.

This is the number of *days* for the money to double (a period is a day, because of our P/Y entry). This means that it takes about $5060.320984/365 = 13.86 \dots \approx 13.9$ years for money invested at 5% interest compounded daily to double.

EXERCISES

- ▶ 75. If \$1,000 is deposited into an account that earns 5% interest compounded daily, the doubling time is approximately 5,061 days.
- Find the amount in the account after 5,061 days.
 - Find the amount in the account after $2 \cdot 5,061$ days.
 - Find the amount in the account after $3 \cdot 5,061$ days.
 - Find the amount in the account after $4 \cdot 5,061$ days.
 - What conclusion can you make?
- ▶ 76. Do the following. (Give the number of periods and the number of years, rounded to the nearest hundredth.)
- Find the doubling time corresponding to 5% interest compounded annually.
 - Find the doubling time corresponding to 5% interest compounded quarterly.
 - Find the doubling time corresponding to 5% interest compounded monthly.
 - Find the doubling time corresponding to 5% interest compounded daily.
 - Discuss the effect of the compounding period on doubling time.
77. Do the following. (Give the number of periods and the number of years, rounded to the nearest hundredth.)
- Find the doubling time corresponding to 6% interest compounded annually.
 - Find the doubling time corresponding to 7% interest compounded annually.
 - Find the doubling time corresponding to 10% interest compounded annually.
 - Discuss the effect of the interest rate on doubling time.
- ▶ 78. If you invest \$10,000 at 8.125% interest compounded daily, how long will it take for you to accumulate \$15,000? How long will it take for you to accumulate \$100,000? (Give the number of periods and the number of years, rounded to the nearest hundredth.)
79. If you invest \$15,000 at $9\frac{3}{8}\%$ interest compounded daily, how long will it take for you to accumulate \$25,000? How long will it take for you to accumulate \$100,000? (Give the number of periods and the number of years, rounded to the nearest hundredth.)
80. If you invest \$20,000 at $6\frac{1}{4}\%$ interest compounded daily, how long will it take for you to accumulate \$30,000? How long will it take for you to accumulate \$100,000? (Give the number of periods and the number of years, rounded to the nearest hundredth.)

5.3

Annuities

OBJECTIVES

- Understand what an annuity is
- Use the Annuity Formulas
- Determine how to use an annuity to save for retirement.

Many people have long-term financial goals and limited means with which to accomplish them. Your goal might be to save \$3,000 over the next four years for your college education, to save \$10,000 over the next ten years for the down payment on a home, to save \$30,000 over the next eighteen years to finance your new baby's college education, or to save \$300,000 over the next forty years for your retirement. It seems incredible, but each of these goals can be achieved by saving only \$50 a month (if interest rates are favorable)! All you need to do is start an annuity. We'll explore annuities in this section.

An **annuity** is simply a sequence of equal, regular payments into an account in which each payment receives compound interest. Because most annuities involve relatively small periodic payments, they are affordable for the average person. Over longer periods of time, the payments themselves start to amount to a significant sum, but it is really the power of compound interest that makes annuities so amazing. If you pay \$50 a month into an annuity for the next forty years, then your total payment is

$$\frac{\$50}{\text{month}} \cdot \frac{12 \text{ months}}{\text{year}} \cdot 40 \text{ years} = \$24,000$$

A **Christmas Club** is an annuity that is set up to save for Christmas shopping. A Christmas Club participant makes regular equal deposits, and the deposits and the resulting interest are released to the participant in December when the money is needed. Christmas Clubs are different from other annuities in that they span a short amount of time—a year at most—and thus earn only a small amount of interest. (People set them up to be sure that they are putting money aside rather than to generate interest.) Our first few examples will deal with Christmas Clubs, because their short time span makes it possible to see how an annuity actually works.

Annuities as Compound Interest, Repeated

EXAMPLE 1

UNDERSTANDING ANNUITIES On August 12, Patty Leitner joined a Christmas Club through her bank. For the next three months, she would deposit \$200 at the beginning of each month. The money would earn $8\frac{3}{4}\%$ interest compounded monthly, and on December 1, she could withdraw her money for shopping. Use the compound interest formula to find the future value of the account.

SOLUTION

We are given $P = 200$ and $r = \frac{1}{12}$ th of $8\frac{3}{4}\% = \frac{0.0875}{12}$.

First, calculate the future value of the first payment (made on September 1). Use $n = 3$ because it will receive interest during September, October, and November.

$$\begin{aligned} FV &= P(1 + i)^n && \text{the Compound Interest Formula} \\ &= 200\left(1 + \frac{0.0875}{12}\right)^3 && \text{substituting} \\ &= 204.40698 \approx \$204.41 && \text{rounding} \end{aligned}$$

Next, calculate the future value of the second payment (made on October 1). Use $n = 2$ because it will receive interest during October and November.

$$\begin{aligned} FV &= P(1 + i)^n && \text{the Compound Interest Formula} \\ &= 200\left(1 + \frac{0.0875}{12}\right)^2 && \text{substituting} \\ &= 202.9273 \approx \$202.93 && \text{rounding} \end{aligned}$$

Next, calculate the future value of the third payment (made on November 1). Use $n = 1$ because it will receive interest during November.

$$\begin{aligned} FV &= P(1 + i)^n && \text{the Compound Interest Formula} \\ &= 200\left(1 + \frac{0.0875}{12}\right)^1 && \text{substituting} \\ &= 201.45833 \approx \$201.46 && \text{rounding} \end{aligned}$$

The payment schedule and interest earned are illustrated in Figure 5.12.

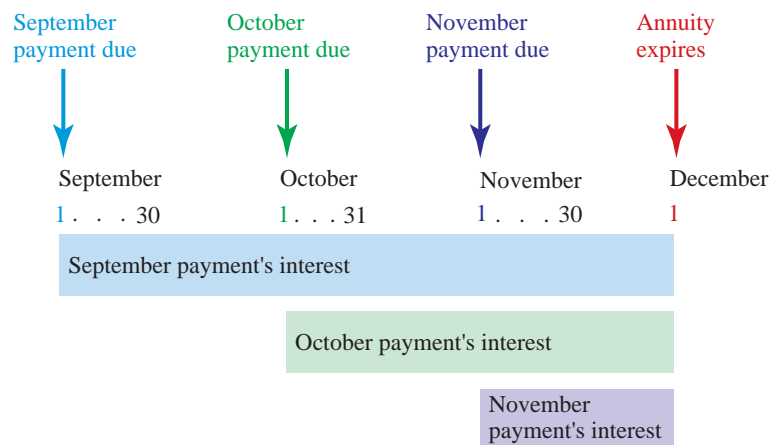


FIGURE 5.12 Patty's payment schedule and interest earning periods.

The future value of Patty's annuity is the sum of the future values of each payment:

$$\begin{aligned} FV &\approx \$204.41 + \$202.93 + \$201.46 \\ &= \$608.80 \end{aligned}$$

Patty's deposits will total \$600.00. She will earn \$8.80 interest on her deposits.

The **payment period** of an annuity is the time between payments; in Example 1, the payment period was one month. The **term** is the time from the beginning of the first payment period to the end of the last; the term of Patty's Christmas Club was three months. When an annuity has **expired** (that is, when its term is over), the entire account or any portion of it may be withdrawn. Most annuities are **simple**,

that is, their compounding period is the same as their payment period (for example, if payments are made monthly, then interest is compounded monthly). In this book, we will work only with simple annuities.

Ordinary Annuities and Annuities Due

An **annuity due** is one in which each payment is due at the beginning of its time period. Patty's annuity in Example 1 was an annuity due, because the payments were due at the *beginning* of each month. An **ordinary annuity** is an annuity for which each payment is due at the end of its time period. As the name implies, this form of annuity is more typical. As we will see in the next example, the difference is one of timing.

EXAMPLE 2

UNDERSTANDING THE DIFFERENCE BETWEEN AN ORDINARY ANNUITY AND AN ANNUITY DUE Dan Bach also joined a Christmas Club through his bank. His was just like Patty's except that his payments were due at the end of each month, and his first payment was due September 30. Use the Compound Interest Formula to find the future value of the account.

SOLUTION

This is an *ordinary* annuity because payments are due at the *end* of each month. Interest is compounded monthly. From Example 1, we know that $P = 200$, $i = \frac{1}{12}$ of $8\frac{3}{4}\% = \frac{0.0875}{12}$.

To calculate the future value of the first payment (made on September 30), use $n = 2$. This payment will receive interest during October and November.

$$\begin{aligned} FV &= P(1 + i)^n && \text{the Compound Interest Formula} \\ &= 200\left(1 + \frac{0.0875}{12}\right)^2 && \text{substituting} \\ &= 202.9273 \approx \$202.93 && \text{rounding} \end{aligned}$$

To calculate the future value of the second payment (made on October 31), use $n = 1$. This payment will receive interest during November.

$$\begin{aligned} FV &= P(1 + i)^n && \text{the Compound Interest Formula} \\ &= 200\left(1 + \frac{0.0875}{12}\right)^1 && \text{substituting} \\ &= 201.45833 \approx \$201.46 && \text{rounding} \end{aligned}$$

To calculate the future value of the second payment (made on November 30), note that no interest is earned, because the payment is due November 30 and the annuity expires December 1. Therefore,

$$FV = \$200$$

Dan's payment schedule and interest payments are illustrated in Figure 5.13.

The future value of Dan's annuity is the sum of the future values of each payment:

$$\begin{aligned} FV &\approx \$200 + \$201.46 + \$202.93 \\ &= \$604.39 \end{aligned}$$

Dan earned \$4.39 interest on his deposits.

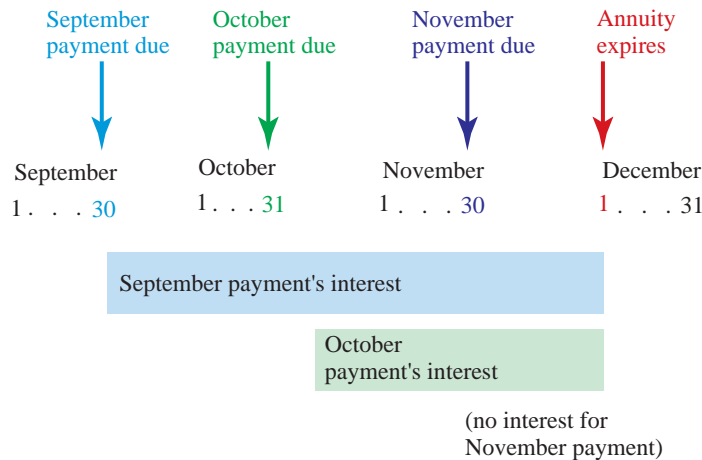


FIGURE 5.13 Dan's payment schedule and interest earning periods.

In Examples 1 and 2, why did Patty earn more interest than Dan? The reason is that each of her payments was made a month earlier and therefore received an extra month's interest. In fact, we could find the future value of Patty's account by giving Dan's future value one more month's interest.

$$\begin{aligned} \text{Patty's } FV &= \text{Dan's } FV \cdot (1 + i)^1 \\ \$608.80 &= \$604.39 \cdot \left(1 + \frac{0.0875}{12}\right)^1 \end{aligned}$$

More generally, we can find the future value of an annuity due by giving an ordinary annuity's future value one more month's interest.

$$FV(\text{due}) = FV(\text{ordinary}) \cdot (1 + i)^1$$

The difference between an ordinary annuity and an annuity due is strictly a timing difference, because any ordinary annuity in effect will become an annuity due if you leave all funds in the account for one extra period.

The Annuity Formulas

The procedure followed in Examples 1 and 2 reflects what actually happens with annuities, and it works fine for a small number of payments. However, most annuities are long-term, and the procedure would become tedious if we were computing the future value after forty years. Because of this, long-term annuities are calculated with their own formula.

For an *ordinary* annuity with payment $pymt$, a periodic rate i , and a term of n payments, the first payment receives interest for $n - 1$ periods. The payment is made at the end of the first period, so it received no interest for that one period. Its future value is

$$FV(\text{first } pymt) = pymt(1 + i)^{n-1}$$

The last payment receives no interest (under the annuity), because it is due at the end of the last period and it expires the next day. Its future value is

$$FV(\text{last } pymt) = pymt$$

The next-to-last payment receives one period's interest, so its future value is

$$FV(\text{next-to-last } pymt) = pymt(1 + i)^1$$

The future value of the annuity is the sum of all of these future values of individual payments:

$$FV = pymt + pymt(1 + i)^1 + pymt(1 + i)^2 + \cdots + pymt(1 + i)^{n-1}$$

To get a short-cut formula from all this, we will multiply each side of this equation by $(1 + i)$ and then subtract the original equation from the result. This leads to a lot of cancelling.

$$FV(1 + i) = \cancel{pymt(1 + i)} + \cancel{pymt(1 + i)^2} + \cdots + \cancel{pymt(1 + i)^{n-1}} + pymt(1 + i)^n$$

$$\text{minus: } FV = pymt + \cancel{pymt(1 + i)} + \cancel{pymt(1 + i)^2} + \cdots + \cancel{pymt(1 + i)^{n-1}}$$

$$\text{equals: } FV(1 + i) - FV = pymt(1 + i)^n - pymt \quad \text{subtracting}$$

$$FV(1 + i - 1) = pymt[(1 + i)^n - 1] \quad \text{factoring}$$

$$FV(i) = pymt[(1 + i)^n - 1]$$

$$FV = pymt \frac{(1 + i)^n - 1}{i} \quad \text{dividing}$$

This is the future value of the ordinary annuity.

ORDINARY ANNUITY FORMULA

The future value FV of an ordinary annuity with payment size $pymt$, a periodic rate i , and a term of n payments is

$$FV(\text{ord}) = pymt \frac{(1 + i)^n - 1}{i}$$

As we saw in Examples 1 and 2, the future value of an annuity due is the future value of an ordinary annuity plus one more period's interest.

$$FV(\text{due}) = FV(\text{ord}) \cdot (1 + i)$$

This gives us a formula for the future value of an annuity due.

ANNUITY DUE FORMULA

The future value FV of an annuity due with payment size $pymt$, a periodic rate i , and a term of n payments is

$$\begin{aligned} FV(\text{due}) &= FV(\text{ord}) \cdot (1 + i) \\ &= pymt \frac{(1 + i)^n - 1}{i} (1 + i) \end{aligned}$$

Tax-Deferred Annuities

A **tax-deferred annuity (TDA)** is an annuity that is set up to save for retirement. Money is automatically deducted from the participant's paychecks until retirement, and the federal (and perhaps state) tax deduction is computed *after* the annuity payment has been deducted, resulting in significant tax savings. In some cases, the employer also makes a regular contribution to the annuity.

The following example involves a long-term annuity. Usually, the interest rate of a long-term annuity varies somewhat from year to year. In this case, calculations must be viewed as predictions, not guarantees.

EXAMPLE 3

USING AN ANNUITY TO SAVE FOR RETIREMENT Tom and Betty decided that they should start saving for retirement, so they set up a tax-deferred annuity. They arranged to have \$200 taken out of each of Tom's monthly checks, which will earn $8\frac{3}{4}\%$ interest. Because of the tax-deferring effect of the TDA, Tom's take-home pay went down by only \$115. Tom just had his thirtieth birthday, and his ordinary annuity will come to term when he is 65.

- Find the future value of the annuity.
- Find Tom's contribution and the interest portion.

SOLUTION

- a. This is an ordinary annuity, with $pymt = 200$, $i = \frac{1}{12}$ th of $8\frac{3}{4}\% = 0.0875/12$, and $n = 35 \text{ years} = 35 \text{ years} \cdot 12 \text{ months/year} = 420$ monthly payments.

$$\begin{aligned} FV(\text{ord}) &= pymt \frac{(1+i)^n - 1}{i} && \text{the Ordinary Annuity Formula} \\ &= 200 \frac{(1 + 0.0875/12)^{420} - 1}{0.0875/12} && \text{substituting} \\ &\approx \$552,539.96 && \text{rounding} \end{aligned}$$



Because $0.0875/12$ occurs twice in the calculation, compute it first and put it into your calculator's memory. Then type

(1 + RCL) y^x 420 - 1 = ÷ RCL × 200 =

In the above, "RCL" refers to recalling the stored number. The way to do this varies with different calculators.



Because $0.0875/12$ occurs twice in the calculation, compute it and store it by typing

.0875 ÷ 12 STO> ALPHA I

Then type

200 × ((1 + ALPHA I) ^ 420 - 1) ÷
ALPHA I ENTER

- b. The principal part of this \$552,539.96 is Tom's contribution, and the rest is interest.

- Tom's contribution is 420 payments of \$200 each = $420 \cdot \$200 = \$84,000$.
- The interest portion is then $\$552,539.96 - \$84,000 = \$468,539.96$.

In Example 3, the interest portion is almost six times as large as Tom's contribution! The magnitude of the earnings illustrates the amazing power of annuities and the effect of compound interest over a long period of time.

Sinking Funds

A **sinking fund** is an annuity in which the future value is a specific amount of money that will be used for a certain purpose, such as a child's education or the down payment on a home.

help the saver to understand his or her options; it refers to an alternative way of saving the same amount of money in the same time. It is called the *present value* because it refers to the single action that the saver can take *in the present* (i.e., at the beginning of the annuity's term) that would have the same effect as would the annuity.

EXAMPLE 5

FINDING THE PRESENT VALUE Find the present value of Tom and Betty's annuity.

SOLUTION

$$\begin{aligned}
 FV &= P(1 + i)^n && \text{the Compound Interest Formula} \\
 30,005.95588 &= P\left(1 + \frac{0.0925}{26}\right)^{468} && \text{substituting} \\
 P &= \frac{30,005.95588}{\left(1 + \frac{0.0925}{26}\right)^{468}} && \text{solving for } P \\
 &= 5,693.6451 \dots \approx \$5693.65 && \text{rounding}
 \end{aligned}$$

This means that Tom and Betty would have to deposit \$5,693.55 as a lump sum to save as much money as the annuity would yield. They chose an annuity over a lump sum deposit because they could not afford to tie up \$5,700 for eighteen years, but they could afford to deduct \$25 out of each paycheck.

EXAMPLE 6

FINDING THE PRESENT VALUE Find the present value of an ordinary annuity that has \$200 monthly payments for twenty-five years, where the account receives $10\frac{1}{2}\%$ interest.

SOLUTION

We could find the future value of the annuity and then find the lump sum deposit whose future value matches it, as we did in Example 5. However, it is simpler to do the calculation all at once. The key is to realize that the future value of the lump sum must equal the future value of the annuity:

Future value of lump sum = future value of annuity

$$P(1 + i)^n = pymt \frac{(1 + i)^n - 1}{i}$$

For both the lump sum and the annuity, $i = \frac{1}{12}$ of $10\frac{1}{2}\% = \frac{0.105}{12}$ and $n = 25$ years = 300 months. The annuity's payment is $pymt = \$200$.

$$\begin{aligned}
 P(1 + i)^n &= pymt \frac{(1 + i)^n - 1}{i} \\
 P(1 + 0.105/12)^{300} &= 200 \frac{(1 + 0.105/12)^{300} - 1}{0.105/12} && \text{substituting}
 \end{aligned}$$

First, calculate the right side, as with any annuity calculation. Then divide by $(1 + 0.105/12)^{300}$ to find P .

$$\begin{aligned}
 P &= 21182.363 \dots \\
 P &\approx \$21,182.36
 \end{aligned}$$

This means that one would have to make a lump sum deposit of more than \$21,000 to have as much money after twenty-five years as with monthly \$200 annuity payments.

PRESENT VALUE OF ANNUITY FORMULA

$$FV(\text{lump sum}) = FV(\text{annuity})$$

$$P(1 + i)^n = pymt \frac{(1 + i)^n - 1}{i}$$

The present value is the lump sum P .

5.3 EXERCISES

In Exercises 1–14, find the future value of the given annuity.

- ▶ 1. ordinary annuity, \$120 monthly payment, $5\frac{3}{4}\%$ interest, one year
- ▶ 2. ordinary annuity, \$175 monthly payment, $6\frac{1}{8}\%$ interest, eleven years
- 3. annuity due, \$100 monthly payment, $5\frac{7}{8}\%$ interest, four years
- ▶ 4. annuity due, \$150 monthly payment, $6\frac{1}{4}\%$ interest, thirteen years
- ▶ 5. On September 8, Bert Sarkis joined a Christmas Club. His bank will automatically deduct \$75 from his checking account at the end of each month and deposit it into his Christmas Club account, where it will earn 7% interest. The account comes to term on December 1. Find the following:
 - a. The future value of the account, using an annuity formula
 - b. The future value of the account, using the compound interest formula
 - c. Bert's total contribution to the account
 - d. The total interest
- 6. On August 19, Rachael Westlake joined a Christmas Club. Her bank will automatically deduct \$110 from her checking account at the end of each month and deposit it into her Christmas Club account, where it will earn $6\frac{7}{8}\%$ interest. The account comes to term on December 1. Find the following:
 - a. The future value of the account, using an annuity formula
 - b. The future value of the account, using the compound interest formula
 - c. Rachael's total contribution to the account
 - d. The total interest
- 7. On August 23, Ginny Deus joined a Christmas Club. Her bank will automatically deduct \$150 from her checking account at the beginning of each month and deposit it into her Christmas Club account, where it will earn $7\frac{1}{4}\%$ interest. The account comes to term on December 1. Find the following:
 - a. The future value of the account, using an annuity formula
 - b. The future value of the account, using the compound interest formula
 - c. Ginny's total contribution to the account
 - d. The total interest
- ▶ 8. On September 19, Lynn Knight joined a Christmas Club. Her bank will automatically deduct \$100 from her checking account at the beginning of each month and deposit it into her Christmas Club account, where it will earn 6% interest. The account comes to term on December 1. Find the following:
 - a. The future value of the account, using an annuity formula
 - b. The future value of the account, using the compound interest formula
 - c. Lynn's total contribution to the account
 - d. The total interest
- 9. Pat Gilbert recently set up a TDA to save for her retirement. She arranged to have \$175 taken out of each of her monthly checks; it will earn $10\frac{1}{2}\%$ interest. She just had her thirty-ninth birthday, and her ordinary annuity comes to term when she is 65. Find the following:
 - a. The future value of the account
 - b. Pat's total contribution to the account
 - c. The total interest
- ▶ 10. Dick Eckel recently set up a TDA to save for his retirement. He arranged to have \$110 taken out of each of his biweekly checks; it will earn $9\frac{7}{8}\%$ interest. He just had his twenty-ninth birthday, and his ordinary annuity comes to term when he is 65. Find the following:
 - a. The future value of the account
 - b. Dick's total contribution to the account
 - c. The total interest

▶ Selected exercises available online at www.webassign.net/brookscole

11. Sam Whitney recently set up a TDA to save for his retirement. He arranged to have \$290 taken out of each of his monthly checks; it will earn 11% interest. He just had his forty-fifth birthday, and his ordinary annuity comes to term when he is 65. Find the following:
 - a. The future value of the account
 - b. Sam’s total contribution to the account
 - c. The total interest
- ▶ 12. Art Dull recently set up a TDA to save for his retirement. He arranged to have \$50 taken out of each of his biweekly checks; it will earn it will earn $9\frac{1}{8}\%$ interest. He just had his 30th birthday, and his ordinary annuity comes to term when he is 65. Find the following:
 - a. The future value of the account
 - b. Art’s total contribution to the account
 - c. The total interest

In Exercises 13–18, (a) find and (b) interpret the present value of the given annuity.

13. The annuity in Exercise 1
14. The annuity in Exercise 2
15. The annuity in Exercise 5
- ▶ 16. The annuity in Exercise 6
17. The annuity in Exercise 9
- ▶ 18. The annuity in Exercise 10

In Exercises 19–24, find the monthly payment that will yield the given future value.

- ▶ 19. \$100,000 at $9\frac{1}{4}\%$ interest for thirty years; ordinary annuity
20. \$45,000 at $8\frac{7}{8}\%$ interest for twenty years; ordinary annuity
- ▶ 21. \$250,000 at $10\frac{1}{2}\%$ interest for forty years, ordinary annuity
22. \$183,000 at $8\frac{1}{4}\%$ interest for twenty-five years, ordinary annuity
23. \$250,000 at $10\frac{1}{2}\%$ interest for forty years, annuity due
24. \$183,000 at $8\frac{1}{4}\%$ interest for twenty-five years, annuity due
- ▶ 25. Mr. and Mrs. Gonzales set up a TDA to save for their retirement. They agreed to have \$100 deducted from each of Mrs. Gonzales’s biweekly paychecks, which will earn $8\frac{3}{8}\%$ interest.
 - a. Find the future value of their ordinary annuity if it comes to term after they retire in $35\frac{1}{2}$ years.
 - b. After retiring, the Gonzales family convert their annuity to a savings account, which earns 6.1% interest compounded monthly. At the end of each month, they withdraw \$650 for living expenses. Complete the chart in Figure 5.14 for their postretirement account.

Month Number	Account Balance at Beginning of the Month	Interest for the Month	Withdrawal	Account Balance at End of the Month
1				
2				
3				
4				
5				

FIGURE 5.14 Chart for Exercise 25.

26. Mr. and Mrs. Jackson set up a TDA to save for their retirement. They agreed to have \$125 deducted from each of Mrs. Jackson’s biweekly paychecks, which will earn $7\frac{5}{8}\%$ interest.
 - a. Find the future value of their ordinary annuity, if it comes to term after they retire in $32\frac{1}{2}$ years.
 - b. After retiring, the Jacksons convert their annuity to a savings account, which earns 6.3% interest compounded monthly. At the end of each month, they withdraw \$700 for living expenses. Complete the chart in Figure 5.15 for their postretirement account.

Month Number	Account Balance at Beginning of the Month	Interest for the Month	Withdrawal	Account Balance at End of the Month
1				
2				
3				
4				
5				

FIGURE 5.15 Chart for Exercise 26.

- ▶ 27. Jeanne and Harold Kimura want to set up a TDA that will generate sufficient interest at maturity to meet their living expenses, which they project to be \$950 per month.
 - a. Find the amount needed at maturity to generate \$950 per month interest if they can get $6\frac{1}{2}\%$ interest compounded monthly.
 - b. Find the monthly payment that they would have to put into an ordinary annuity to obtain the future value found in part (a) if their money earns $8\frac{1}{4}\%$ and the term is thirty years.
- ▶ 28. Susan and Bill Stamp want to set up a TDA that will generate sufficient interest at maturity to meet their living expenses, which they project to be \$1,200 per month.

- a. Find the amount needed at maturity to generate \$1,200 per month interest, if they can get $7\frac{1}{4}\%$ interest compounded monthly.
- b. Find the monthly payment that they would have to make into an ordinary annuity to obtain the future value found in part (a) if their money earns $9\frac{3}{4}\%$ and the term is twenty-five years.
- ▶ 29. In June 2004, Susan set up a TDA to save for retirement. She agreed to have \$200 deducted from each of her monthly paychecks. The annuity's interest rate was allowed to change once each year.
- a. In 2004, the interest rate was 1%. Find the account balance in June 2005.
- b. In 2005, the interest rate was 2.25%. Find the account balance in June 2006. To do this, think of the June 2005 account balance as a lump sum that earns compound interest.
- c. In 2006, the interest rate was 4.5%. Find the account balance in June 2007.
- Interest rate source: Mortgagex.com.*
30. In June 2007, Manuel set up a TDA to save for retirement. He agreed to have \$175 deducted from each of his monthly paychecks. The annuity's interest rate was allowed to change once each year.
- a. In 2007, the interest rate was 5.25%. Find the account balance in June 2008.
- b. In 2008, the interest rate was 3.8%. Find the account balance in June 2009. To do this, think of the June 2008 account balance as a lump sum that earns compound interest.
- c. In 2009, the interest rate was 2.2%. Find the account balance in June 2010.
- Interest rate source: Mortgagex.com.*
- In Exercises 31–34, use the following information. An Individual Retirement Account (IRA) is an annuity that is set up to save for retirement. IRAs differ from TDAs in that an IRA allows the participant to contribute money whenever he or she wants, whereas a TDA requires the participant to have a specific amount deducted from each of his or her paychecks.*
- ▶ 31. When Shannon Pegnim was 14, she got an after-school job at a local pet shop. Her parents told her that if she put some of her earnings into an IRA, they would contribute an equal amount to her IRA. That year and every year thereafter, she deposited \$1,000 into her IRA. When she became 25 years old, her parents stopped contributing, but Shannon increased her annual deposit to \$2,000 and continued depositing that amount annually until she retired at age 65. Her IRA paid 8.5% interest. Find the following:
- a. The future value of the account
- b. Shannon's and her parents' total contributions to the account
- c. The total interest
- d. The future value of the account if Shannon waited until she was 19 before she started her IRA
- e. The future value of the account if Shannon waited until she was 24 before she started her IRA
- ▶ 32. When Bo McSwine was 16, he got an after-school job at his parents' barbecue restaurant. His parents told him that if he put some of his earnings into an IRA, they would contribute an equal amount to his IRA. That year and every year thereafter, he deposited \$900 into his IRA. When he became 21 years old, his parents stopped contributing, but Bo increased his annual deposit to \$1,800 and continued depositing that amount annually until he retired at age 65. His IRA paid 7.75% interest. Find the following:
- a. The future value of the account
- b. Bo's and his parents' total contributions to the account
- c. The total interest
- d. The future value of the account if Bo waited until he was 18 before he started his IRA
- e. The future value of the account if Bo waited until he was 25 before he started his IRA
- ▶ 33. If Shannon Pegnim from Exercise 31 started her IRA at age 35 rather than age 14, how big of an annual contribution would she have had to have made to have the same amount saved at age 65?
- ▶ 34. If Bo McSwine from Exercise 32 started his IRA at age 35 rather than age 16, how big of an annual contribution would he have had to have made to have the same amount saved at age 65?
- ▶ 35. Toni Torres wants to save \$1,200 in the next two years to use as a down payment on a new car. If her bank offers her 9% interest, what monthly payment would she need to make into an ordinary annuity to reach her goal?
- ▶ 36. Fred and Melissa Furth's daughter Sally will be a freshman in college in six years. To help cover their extra expenses, the Furths decide to set up a sinking fund of \$12,000. If the account pays 7.2% interest and they wish to make quarterly payments, find the size of each payment.
- ▶ 37. Anne Geyer buys some land in Utah. She agrees to pay the seller a lump sum of \$65,000 in five years. Until then, she will make monthly simple interest payments to the seller at 11% interest.
- a. Find the amount of each interest payment.
- b. Anne sets up a sinking fund to save the \$65,000. Find the size of her semiannual payments if her payments are due at the end of every six-month period and her money earns $8\frac{3}{8}\%$ interest.
- c. Prepare a table showing the amount in the sinking fund after each deposit.
- ▶ 38. Chrissy Fields buys some land in Oregon. She agrees to pay the seller a lump sum of \$120,000 in six years.

Until then, she will make monthly simple interest payments to the seller at 12% interest.

- a. Find the amount of each interest payment.
 - b. Chrissy sets up a sinking fund to save the \$120,000. Find the size of her semiannual payments if her money earns $10\frac{3}{4}\%$ interest.
 - c. Prepare a table showing the amount in the sinking fund after each deposit.
39. Develop a new formula for the present value of an ordinary annuity by solving the Present Value of Annuity Formula for P and simplifying.
 40. Use the formula developed in Exercise 39 to find the present value of the annuity in Exercise 2.
 41. Use the formula developed in Exercise 39 to find the present value of the annuity in Exercise 1.
 42. Use the formula developed in Exercise 39 to find the present value of the annuity in Exercise 6.
 43. Use the formula developed in Exercise 39 to find the present value of the annuity in Exercise 5.



Answer the following questions using complete sentences and your own words.

• CONCEPT QUESTIONS

44. Explain the difference between compound interest and an annuity.
45. Explain how an annuity is based on compound interest.
46. Describe the difference between an ordinary annuity and an annuity due.
47. Compare and contrast an annuity with a lump sum investment that receives compound interest. Be sure to discuss both the similarity and the difference between these two concepts, as well as the advantages and disadvantages of each.
48. Which is always greater: the present value of an annuity or the future value? Why?
49. If you want to retire on \$2,000 a month for twenty-five years, do you need to save \$2,000 a month for twenty-five years before retiring? Why or why not?
50. *For those who have completed Section 1.1:* Is the logic used in deriving the Ordinary Annuity Formula inductive or deductive? Why? Is the logic used in deriving the relationship

$$FV(\text{due}) = FV(\text{ordinary}) \cdot (1 + i)$$

inductive or deductive? Why?



51. Think of something that you would like to be able to afford but cannot—perhaps a car, boat, or motorcycle.

In this exercise, you will explore how to make that unaffordable dream a realistic goal.

- a. Just what is it that you would like to be able to afford?
 - b. Determine an appropriate but realistic date for making the purchase. Justify your date.
 - c. Do some research and determine how much your goal would cost currently. Cite your sources.
 - d. Do some web research and determine the current rate of inflation. Inflation rate information is readily available on the web. Cite your sources.
 - e. Use the results of parts (b), (c), and (d), as well as either the simple or compound interest formula, to predict how much your goal will cost in the future.
 - f. Discuss why you chose to use the simple or compound interest formula in part (e).
 - g. Go to the web sites of four different banks and determine the current interest rate available for an annuity of an appropriate term.
 - h. Determine the necessary annuity payment that will allow you to meet your goal. If necessary, alter the date from part (b).
52. Suppose you had a baby in 2005 and you decided that it would be wise to start saving for his or her college education. In this exercise, you will explore how to go about doing that.
 - a. According to the College Board, four-year public schools cost an average of \$5,491 per year in 2005–2006. Four-year private schools cost an average of \$21,235 per year. These costs include tuition and fees, room and board, books and supplies, and personal expenses. Furthermore, four-year public schools increased 7.1% from 2004–2005, and four-year private schools increased 5.9% from 2004–2005. Use this information, as well as either the simple or compound interest formula, to predict the cost of your child's college education in his or her freshman, sophomore, junior, and senior years for both public and private institutions. Discuss your assumptions and justify your work.
(Source: www.collegeboard.com/pay.)
 - b. Will you save for a public or a private institution? Why?
 - c. Discuss why you chose to use the simple or compound interest formula in part (a).
 - d. Go to the web sites of four different banks and determine the current interest rate available for an annuity of an appropriate term.
 - e. Determine the necessary annuity payment that will allow you to save the total amount from part (a) in time to meet your goal.

Some useful links for these web projects are listed on the text web site:

www.cengage.com/math/johnson



ANNUITIES WITH A TI'S TVM APPLICATION

EXAMPLE 7

FINDING HOW LONG IT TAKES Use a TI-83/84's TVM application to find the how long it takes for an ordinary annuity to have a balance of one million dollars if the monthly payments of \$600 earn 6% interest.

SOLUTION

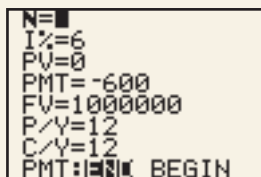


FIGURE 5.16

Preparing the TVM screen.

1. Press **APPS**, select option 1: "Finance", and press **ENTER**.
2. Select option 1: "TVM Solver", and press **ENTER**.
3. Enter appropriate values for the variables:
 - N is the number of compounding periods. This is the number we're trying to find, so we temporarily enter 0.
 - $I\%$ is the *annual* interest rate (not a periodic rate), as a percent, so enter 6 for $I\%$. Note that we do not convert to a decimal or a periodic rate.
 - PV is the present value. There is no present value here, so we enter 0.
 - PMT is the payment. The payment is \$600, but it's an outgoing amount of money (since we give it to the bank), so we enter -600 for PMT . You must enter a negative number for any outgoing amount of money.
 - FV is the future value. We're looking for the amount of time that it takes to have \$1,000,000, so enter 1000000. This is an incoming amount of money (since the bank gives it to us), so we use 1000000 rather than -1000000 .
 - P/Y is the number of periods per year. We make monthly payments, so there are twelve periods per year. Enter 12 for P/Y .
 - C/Y is automatically made to be the same as P/Y . In this text, we will never encounter a situation in which C/Y is different from P/Y . See Figure 5.16.
 - An ordinary annuity's payments are due at the end of each period so highlight "End"
4. To solve for N , use the arrow buttons to highlight the 0 that we entered for N . Then press **ALPHA** **SOLVE**. (Pressing **ALPHA** makes the **ENTER** button become the **SOLVE** button.) As a result, we find that N is 447.8343121.

This is the number of *months* it takes (a period is a month, because of our P/Y entry). This means that it takes about $447.8343121/12 = 37.31 \dots \approx 37.3$ years to have a million dollars.

EXERCISES

- ▶ 53. Redo Example 7 for an annuity due.
 - a. 2% interest
 - b. 4% interest
 - c. 6% interest
 - d. 10% interest
- ▶ 54. Find how long it takes for an ordinary annuity to have a balance of two million dollars if the monthly payments of \$1100 earn:

5.4

Amortized Loans

OBJECTIVES

- Understand what an amortized loan is
- Learn how to do home loan and car loan computations
- Learn how to make and use an amortization schedule

When you buy a house, a car, or a boat, chances are you will finance it with a simple interest amortized loan. A **simple interest amortized loan** is a loan in which each payment consists of principal and interest, and the interest is simple interest computed on the outstanding principal. We'll explore these loans in this section.

An **amortized loan** is a loan for which the loan amount, plus interest, is paid off with a series of regular equal payments. An add-on interest loan (discussed in Section 5.1) is an amortized loan. A simple interest amortized loan is also an amortized loan, but it is calculated differently from an add-on interest loan. You should be aware that the payments are smaller with a simple interest amortized loan than they are with an add-on interest loan (assuming, naturally, that the interest rates, loan amounts, and number of payments are the same).

A simple interest amortized loan is calculated as an ordinary annuity whose future value is the same as the loan amount's future value, under compound interest. Realize, though, that a simple interest amortized loan's payments are used to pay off a loan, whereas an annuity's payments are used to generate savings.

SIMPLE INTEREST AMORTIZED LOAN FORMULA

Future value of annuity = future value of loan amount

$$pymt \frac{(1 + i)^n - 1}{i} = P(1 + i)^n$$

where $pymt$ is the loan payment, i is the periodic interest rate, n is the number of periods, and P is the present value or loan amount.

Algebraically, this formula could be used to determine any one of the four unknowns ($pymt$, i , n , and P) if the other three are known. We will use it to find the payment when the annual interest rate, number of periods, and loan amount are known.

EXAMPLE 1

BUYING A CAR Tom and Betty decide that they need a more dependable car, now that they have a baby. They buy one for \$13,518.77. They make a \$1,000 down payment and finance the balance through a four-year simple interest amortized loan from their bank. They are charged 12% interest.

- Find their monthly payment.
- Find the total interest they will pay over the life of the loan.

SOLUTION

- a. Use the simple interest amortized loan formula, with a loan amount of $P = \$13,518.77 - \$1,000 = \$12,518.77$, a monthly interest rate of $12\%/12 = 1\% = 0.01$, and a term of $n = 4 \text{ years} = 48 \text{ months}$.

$$pymt \frac{(1 + i)^n - 1}{i} = P(1 + i)^n \quad \text{the Simple Interest Amortized Loan Formula}$$

$$pymt \frac{(1 + 0.01)^{48} - 1}{0.01} = 12,518.77(1 + 0.01)^{48} \quad \text{substituting}$$

To find $pymt$, we need to divide the right side by the fraction on the left side or, equivalently, multiply by its reciprocal. First, find the fraction on the left side, as with any annuity calculation. Then multiply the right side by the fraction's reciprocal.



(1 + .01) y^x 48 - 1 = ÷ .01 =

fraction on the left side

1/x × 12518.77 × (1 + .01) y^x 48 =

reciprocal fraction on the right side

(((1 + .01) ^ 48 - 1) ÷ .01)

fraction on the left side

x⁻¹ × 12518.77 (1 + .01) ^ 48 ENTER

reciprocal fraction on the right side

We get $pymt = 329.667... \approx \329.67 .



As a rough check, note that Tom and Betty borrowed \$12,518.77. If they paid no interest, each payment would be $\$12,518.77/48 \approx \260.81 . Since they must pay interest, each payment must be larger than \$260.81. Our calculation checks, because $\$329.67 > \260.81 .

- b. The total of their payments is $48 \cdot \$329.67 = \$15,824.16$. This includes both principal and interest. Of this, \$12,518.77 is principal, so $\$15,824.16 - \$12,518.77 = \$3,305.39$ is interest. Over the life of the loan, Tom and Betty are paying \$3,305.39 in interest.

Amortization Schedules

The simple interest amortized loan formula is an equation with the ordinary annuity formula on one side of the equal symbol and the compound interest formula on the other side. However, with such a loan, you do not pay compound interest. Instead, the interest portion of each payment is simple interest on the outstanding principal.

An **amortization schedule** is a list of several periods of payments, the principal and interest portions of those payments, and the **outstanding principal** (or **balance**) after each of those payments is made.

The data on the amortization schedule are important to the borrower for two reasons. The borrower needs to know the total interest paid, for tax purposes. Interest paid on a home loan is usually deductible from the borrower's income tax, and interest paid on a loan by a business is usually deductible. The borrower would also need the data on an amortization schedule if he or she is considering paying off the loan early. Such prepayment could save money because an advance payment would be all principal and would not include any interest; however, the lending institution may charge a **prepayment penalty** which would absorb some of the interest savings.

EXAMPLE 2

SOLUTION

PREPARING AN AMORTIZATION SCHEDULE Prepare an amortization schedule for the first two months of Tom and Betty's loan.

For any simple interest loan, the interest portion of each payment is simple interest on the outstanding principal, so use the Simple Interest Formula, $I = Prt$, to compute the interest. Recall that r and t are annual figures. For each payment, $r = 12\% = 0.12$ and $t = 1 \text{ month} = \frac{1}{12} \text{ year}$. For payment number 1, $P = \$12,418.77$ (the amount borrowed). The interest portion of payment number 1 is

$$\begin{aligned} I &= Prt && \text{the Simple Interest Formula} \\ &= 12,518.77 \cdot 0.12 \cdot \frac{1}{12} && \text{substituting} \\ &\approx \$125.19 && \text{rounding} \end{aligned}$$

The principal portion of payment number 1 is

$$\begin{aligned} &\text{Payment} - \text{interest portion} \\ &= \$329.67 - \$125.19 \\ &= \$204.48 \end{aligned}$$

The outstanding principal or balance is

$$\begin{aligned} &\text{Previous principal} - \text{principal portion} \\ &= \$12,518.77 - \$204.48 \\ &= \$12,314.29 \end{aligned}$$

All of the above information goes into the amortization schedule. See Figure 5.17.

Payment number	Principal portion	Interest portion	Total payment	Balance
0	—	—	—	\$12,518.77
1	\$204.48	\$125.19	\$329.67	\$12,314.29
2	\$206.53	\$123.14	\$329.67	\$12,107.76

FIGURE 5.17 An amortization schedule for the first two months of Tom and Betty's loan.

When it is time to make payment number 2, less money is owed. The outstanding principal is \$12,314.29, so this is the new value of P . The interest portion of payment number 2 is

$$\begin{aligned} I &= Prt && \text{the Simple Interest Formula} \\ &= 12,314.29 \cdot 0.12 \cdot \frac{1}{12} && \text{substituting} \\ &\approx \$123.14 && \text{rounding} \end{aligned}$$

The principal portion is

$$\begin{aligned} & \text{Payment} - \text{interest portion} \\ &= \$329.67 - \$123.14 \\ &= \$206.53 \end{aligned}$$

The outstanding principal or balance is

$$\begin{aligned} & \text{Previous principal} - \text{principal portion} \\ &= \$12,314.29 - \$206.53 \\ &= \$12,107.76 \end{aligned}$$

All of the above information is shown in the amortization schedule in Figure 5.17.

Notice how in Example 2, the principal portion increases and the interest portion decreases. This continues throughout the life of the loan, and the final payment is mostly principal. This happens because after each payment, the amount due is somewhat smaller, so the interest on the amount due is somewhat smaller also.

EXAMPLE 3

AMORTIZATION SCHEDULES AND WHY THE LAST PAYMENT IS DIFFERENT Comp-U-Rent needs to borrow \$60,000 to increase their inventory of rental computers. The company is confident that their expanded inventory will generate sufficient extra income to allow them to pay off the loan in a short amount of time, so they wish to borrow the money for only three months. First National Bank offers them a simple interest amortized loan at $8\frac{3}{4}\%$ interest.

- Find what their monthly payment would be with First National.
- Prepare an amortization schedule for the entire term of the loan.

SOLUTION

- $P = \$60,000$, $i = \frac{1}{12}$ of $8\frac{3}{4}\% = 0.0875/12$, and $n = 3$ months.

Future value of annuity = future value of loan amount

$$\begin{aligned} \text{pymt} \frac{(1+i)^n - 1}{i} &= P(1+i)^n \\ \text{pymt} \frac{\left(1 + \frac{0.0875}{12}\right)^3 - 1}{\frac{0.0875}{12}} &= 60000 \left(1 + \frac{0.0875}{12}\right)^3 \quad \text{substituting} \end{aligned}$$



First, compute $0.0875/12$ and store it in your calculator's memory. They type


$$\left(1 + \text{RCL} \right) \text{y}^x 3 - 1 = \div \text{RCL} =$$

fraction on the left side

$$\text{1/x} \times 60000 \times \left(1 + \text{RCL} \right) \text{y}^x 3 =$$

reciprocal

fraction on right side

 First compute and store 0.0875/12 by typing
 .0875 ÷ 12 **STO>** **ALPHA** **I** **ENTER**

Then type
 (((1 + **ALPHA** **I**) ^ 3 - 1) ÷ **ALPHA** **I**)

fraction on the left side

x⁻¹ **×** 60000 (1 + **ALPHA** **I**) ^ 3 **ENTER**

reciprocal fraction on the right side

See Figure 5.18 on page 374.

We get

$$pymt = 20,292.375 \dots \text{ solving for } pymt$$

$$\approx \$20,292.38 \text{ rounding}$$

See the “Total Payment” column in Figure 5.18 on page 374.



As a rough check, notice that Comp-U-Rent borrowed \$60,000, and each payment includes principal and interest, so the payment must be larger than $\frac{\$60,000}{3} = \$20,000$. Our calculation checks, because $\$20,292.38 > \$20,000$.

b. For each payment, $r = 8\frac{3}{4}\% = 0.0875$ and $t = 1 \text{ month} = \frac{1}{12} \text{ year}$.

For payment number 1, $P = \$60,000.00$ (the amount borrowed). The interest portion of payment number 1 is

$$I = Prt \quad \text{the Simple Interest Formula}$$

$$= 60,000.00 \cdot 0.0875 \cdot \frac{1}{12} \quad \text{substituting}$$

$$= \$437.50$$

See the “Interest portion” column in Figure 5.18.

The principal portion of payment number 1 is

$$\text{Payment} - \text{interest portion}$$

$$= \$20,292.38 - \$437.50 = \$19,854.88$$

See the “Principal portion” column in Figure 5.18.

The outstanding principal or balance is

$$\text{Previous principal} - \text{principal portion}$$

$$= \$60,000.00 - \$19,854.88 = \$40,145.12$$

For payment number 2, $P = \$40,145.12$ (the outstanding principal).

The interest portion of payment number 2 is

$$I = Prt \quad \text{the Simple Interest Formula}$$

$$= 40,145.12 \cdot 0.0875 \cdot \frac{1}{12} \quad \text{substituting}$$

$$\approx \$292.72 \quad \text{rounding}$$

The Principle portion of payment number 2 is

$$\text{Payment} - \text{interest portion}$$

$$= \$20,292.38 - \$292.72 = \$19,999.66$$

The outstanding principal or balance is

$$\$40,145.12 - \$19,999.66 = \$20,145.46$$

For payment number 3, $P = \$20,145.46$

The interest portion is

$$\begin{aligned} I &= PRT \\ &= \$20,145.46 \cdot 0.0875 \cdot \frac{1}{12} \\ &= \$146.89 \end{aligned}$$

The principal portion is

$$\text{Payment} - \text{interest portion} = \$20,292.38 - \$146.89 = \$20,145.49$$

The outstanding principal or balance is

$$\begin{aligned} \text{Previous principal} - \text{principal portion} \\ &= \$20,145.46 - \$20,145.49 = -\$0.03 \end{aligned}$$

Negative three cents cannot be correct. After the final payment is made, the amount due *must be* \$0.00. The discrepancy arises from the fact that we rounded off the payment size from \$20,292.375 . . . to \$20,292.38. If there were some way in which the borrower could make a monthly payment that is not rounded off, then the calculation above would have yielded an amount due of \$0.00. To repay the exact amount owed, we must compute the last payment differently.

The principal portion of payment number 3 *must be* \$20,145.46, because that is the balance, and this payment is the only chance to pay it. The payment must also include \$146.89 interest, as calculated above. The last payment is the sum of the principal due and the interest on that principal:

$$\$20,145.46 + \$146.89 = \$20,292.35$$

The closing balance is then \$20,145.46 - \$20,145.46 = \$0.00, as it should be. The amortization schedule is given in Figure 5.18.

Payment number	Principal portion	Interest portion	Total payment	Balance
0	—	—	—	\$60,000.00
1	\$19,854.88	\$437.50	\$20,292.38	\$40,145.12
2	\$19,999.66	\$292.72	\$20,292.38	\$20,145.46
3	\$20,145.46	\$146.89	\$20,292.35	\$0.00

FIGURE 5.18 The amortization schedule for Comp-U-Rent's loan.

AMORTIZATION SCHEDULE STEPS

For each payment, list the payment number, principal portion, interest portion, total payment, and balance.

For each payment:

1. Find the interest on the balance, using the simple interest formula.

For each payment except the last:

2. The principal portion is the payment minus the interest portion.
3. The new balance is the previous balance minus the principal portion.

For the last payment:

4. The principal portion is the previous balance.
5. The total payment is the sum of the principal portion and interest portion.

Figure 5.19 shows these steps in chart form.

Payment number	Principal portion	Interest portion	Total payment	Balance
0	—	—	—	Loan amount
First through next-to-last	Total payment minus interest portion	Simple interest on previous balance; use $I = Prt$	Use simple interest amortized loan formula	Previous balance minus this payment's principal portion
Last	Previous balance	Simple interest on previous balance; use $I = Prt$	Principal portion plus interest portion	\$0.00

FIGURE 5.19 Amortization schedule steps.

In preparing an amortization schedule, the new balance is the previous balance minus the principal portion. This means that only the principal portion of a payment goes toward paying off the loan. The interest portion is the lender's profit.

Prepaying a Loan

Sometimes a borrower needs to pay his or her loan off early, before the loan's term is over. This is called **prepaying a loan**. This often happens when people sell their houses. It also happens when interest rates have fallen since the borrower obtained his or her loan and the borrower decides to refinance the loan at a lower rate.

When you prepay a loan, you pay off the unpaid balance. You can find that unpaid balance by looking at an amortization schedule. If an amortization schedule has not already been prepared, the borrower can easily approximate the **unpaid balance** by subtracting the current value of the annuity from the current value of the loan. This approximation is usually off by at most a few pennies.

UNPAID BALANCE FORMULA

Unpaid balance = current value of loan amount – current value of annuity

$$\approx P(1 + i)^n - pymt \frac{(1 + i)^n - 1}{i}$$

where $pymt$ is the loan payment, i is the periodic interest rate, n is the number of periods *from the beginning of the loan to the present*, and P is the loan amount.

A common error in using this formula is to let n equal the number of periods in the entire life of the loan rather than the number of periods from the beginning of the loan until the time of prepayment. This results in an answer of 0. After all of the payments have been made, the unpaid balance should be 0.

EXAMPLE 4

PREPAYING A LOAN Ten years ago, Tom and Betty bought a house for \$140,000. They paid the sellers a 20% down payment and obtained a simple interest amortized loan for \$112,000 from their bank at $10\frac{3}{4}\%$ for thirty years. Their monthly payment is \$1,045.50.

Their old home was fine when there were just two of them. But with their new baby, they need more room. They are considering selling their existing home and buying a new home. This would involve paying off their home loan. They would use the income from the house's sale to do this. Find the cost of paying off the existing loan.

SOLUTION

The loan is a thirty-year loan, so if we were computing the payment, we would use $n = 30 \text{ years} \cdot 12 \text{ months/year} = 360$ months. However, we are computing the unpaid balance, not the payment, so n is not 360. Tom and Betty have made payments on their loan for 10 years = 120 months, so for this calculation, $n = 120$.

Unpaid balance

= current value of loan amount – current value of annuity

$$\approx P(1+i)^n - \text{pymt} \frac{(1+i)^n - 1}{i}$$

$$= 112,000 \left(1 + \frac{0.1075}{12} \right)^{120} - 1,045.50 \frac{\left(1 + \frac{0.1075}{12} \right)^{120} - 1}{\frac{0.1075}{12}}$$

$$= \$102,981.42$$



First, compute $0.1075/12$, and store it in the calculator's memory (as ALPHA I with a graphing calculator). Then type

112000 \times (1 + RCL) y^x 120 $-$ 1045.50 \times ((1 + RCL) y^x 120 $-$ 1) \div RCL =

112000 (1 + ALPHA I) $^$ 120 $-$ 1045.50 ((1 + ALPHA I) $^$ 120 $-$ 1) \div ALPHA I ENTER



As a rough check, note that Tom and Betty borrowed \$112,000, so their unpaid balance must be less. Our calculation checks, because $\$102,981.42 < \$112,000$.

The result of Example 4 means that after ten years of payments of over \$1000 a month on a loan of \$112,000, Tom and Betty still owe approximately \$102,981.42! They were shocked. It is extremely depressing for first-time home purchasers to discover how little of their beginning payments actually goes toward paying off the loan. At the beginning of the loan, you owe a lot of money, so most of each payment is interest, and very little is principal. Later, you do not owe as much, so most of each payment is principal and very little is interest.

5.4 EXERCISES

In the following exercises, all loans are simple interest amortized loans with monthly payments unless labeled otherwise.

In Exercises 1–6, find (a) the monthly payment and (b) the total interest for the given simple interest amortized loan.

- ▶ 1. \$5,000, $9\frac{1}{2}\%$, 4 years
2. \$8,200, $10\frac{1}{4}\%$, 6 years
3. \$10,000, $6\frac{1}{8}\%$, 5 years
4. \$20,000, $7\frac{3}{8}\%$, $5\frac{1}{2}$ years
- ▶ 5. \$155,000, $9\frac{1}{2}\%$, 30 years
6. \$289,000, $10\frac{3}{4}\%$, 35 years

- ▶ 7. Wade Ellis buys a new car for \$16,113.82. He puts 10% down and obtains a simple interest amortized loan for the rest at $11\frac{1}{2}\%$ interest for four years.
 - a. Find his monthly payment.
 - b. Find the total interest.
 - c. Prepare an amortization schedule for the first two months of the loan.
8. Guy dePrimo buys a new car for \$9,837.91. He puts 10% down and obtains a simple interest amortized loan for the rest at $8\frac{3}{8}\%$ interest for four years.
 - a. Find his monthly payment.
 - b. Find the total interest.
 - c. Prepare an amortization schedule for the first two months of the loan.
9. Chris Burditt bought a house for \$212,500. He put 20% down and obtains a simple interest amortized loan for the rest at $10\frac{7}{8}\%$ for thirty years.
 - a. Find his monthly payment.
 - b. Find the total interest.
 - c. Prepare an amortization schedule for the first two months of the loan.
 - d. Most lenders will approve a home loan only if the total of all the borrower's monthly payments, including the home loan payment, is no more than 38% of the borrower's monthly income. How much must Chris make to qualify for the loan?
- ▶ 10. Shirley Trembley bought a house for \$187,600. She put 20% down and obtains a simple interest amortized loan for the rest at $6\frac{3}{8}\%$ for thirty years.
 - a. Find her monthly payment.
 - b. Find the total interest.
 - c. Prepare an amortization schedule for the first two months of the loan.
 - d. Most lenders will approve a home loan only if the total of all the borrower's monthly payments, including the home loan payment, is no more than 38% of the borrower's monthly income. How much must Shirley make to qualify for the loan?

- ▶ 11. Dennis Lamenti wants to buy a new car that costs \$15,829.32. He has two possible loans in mind. One loan is through the car dealer; it is a four-year add-on interest loan at $7\frac{3}{4}\%$ and requires a down payment of \$1,000. The second is through his bank; it is a four-year simple interest amortized loan at $7\frac{3}{4}\%$ and requires a down payment of \$1,000.
 - a. Find the monthly payment for each loan.
 - b. Find the total interest paid for each loan.
 - c. Which loan should Dennis choose? Why?
- ▶ 12. Barry Wood wants to buy a used car that costs \$4,000. He has two possible loans in mind. One loan is through the car dealer; it is a three-year add-on interest loan at 6% and requires a down payment of \$300. The second is through his credit union; it is a three-year simple interest amortized loan at 9.5% and requires a 10% down payment.
 - a. Find the monthly payment for each loan.
 - b. Find the total interest paid for each loan.
 - c. Which loan should Barry choose? Why?
- ▶ 13. Investigate the effect of the term on simple interest amortized auto loans by finding the monthly payment and the total interest for a loan of \$11,000 at $9\frac{7}{8}\%$ interest if the term is
 - a. three years
 - b. four years
 - c. five years
- ▶ 14. Investigate the effect of the interest rate on simple interest amortized auto loans by finding the monthly payment and the total interest for a four-year loan of \$12,000 at
 - a. 8.5%
 - b. 8.75%
 - c. 9%
 - d. 10%
- ▶ 15. Investigate the effect of the interest rate on home loans by finding the monthly payment and the total interest for a thirty-year loan of \$100,000 at
 - a. 6%
 - b. 7%
 - c. 8%
 - d. 9%
 - e. 10%
 - f. 11%
- ▶ 16. Some lenders offer fifteen-year home loans. Investigate the effect of the term on home loans by finding the monthly payment and total interest for a loan of \$100,000 at 10% if the term is
 - a. thirty years
 - b. fifteen years
- ▶ 17. Some lenders offer loans with biweekly payments rather than monthly payments. Investigate the effect of this on home loans by finding the payment and total interest on a thirty-year loan of \$100,000 at 10% interest if payments are made (a) monthly and (b) biweekly.

18. Some lenders are now offering loans with biweekly payments rather than monthly payments. Investigate the effect of this on home loans by finding the payment and total interest on a thirty-year loan of \$150,000 at 8% interest if payments are made (a) monthly and (b) biweekly.
19. Verify (a) the monthly payments and (b) the interest savings in the savings and loan advertisement in Figure 5.20.

Continental Savings of America's

ReFiCenter

is Now Featuring...

9.875%

RATE • ANNUAL PERCENTAGE RATE

15 YEAR FIXED RATE

HOME LOAN FINANCING

Continental Savings of America's No-Point 15 Year Fully Amortized 9.875% Home Loan may be the best home loan on the market today. We call our 15 Year Home Loan "Equity Builders", because for a minimum increase in monthly payments you are paying off your loan much quicker and therefore saving a great deal on interest costs.

By amortizing your loan over 15 Years vs. 30 Years, you save \$137,251.38 over the life of the loan. Continental Savings of America's current 30 Year Fixed Rate No-Point Home Loan is 10.50%.

Total Interest Cost over Life of Loan.

30 Year / 10.50%	\$329,306.14
Monthly Payment \$914.73	
15 Year / 9.875%	\$192,054.76
Monthly Payment \$1,066.97	
Your Interest Savings...	\$137,251.38

*180 equal monthly payments of \$1,066.97. Based on a loan amount of \$100,000. No prepayment penalty. Maximum LTV: 90% on purchases with P.M.L. 80% on refinance. All rates are subject to credit qualifications and Freddie Mac Loan limits for 1-4 units. Owner occupied dwellings only. Rates are subject to change limited offer.

Courtesy Continental Savings of America

FIGURE 5.20 Loan terms for Exercise 19.

20. The home loan in Exercise 19 presented two options. The thirty-year option required a smaller monthly payment. A consumer who chooses the thirty-year option could take the savings in the monthly payment that this option generates and invest that savings in an annuity. At the end of fifteen years, the annuity might be large enough to pay off the thirty-year loan. Determine whether this is a wise plan if the annuity's interest rate is 7.875%. (Disregard the tax ramifications of this approach.)
- ▶ 21. Pool-N-Patio World needs to borrow \$75,000 to increase its inventory for the upcoming summer season. The owner is confident that he will sell most, if not all, of the new inventory during the summer, so he wishes to borrow the money for only four months. His

bank has offered him a simple interest amortized loan at $7\frac{3}{4}\%$ interest.

- a. Find the size of the monthly bank payment.
 - b. Prepare an amortization schedule for all four months of the loan.
- ▶ 22. Slopes R Us needs to borrow \$120,000 to increase its inventory of ski equipment for the upcoming season. The owner is confident that she will sell most, if not all, of the new inventory during the winter, so she wishes to borrow the money for only five months. Her bank has offered her a simple interest amortized loan at $8\frac{7}{8}\%$ interest.
- a. Find the size of the monthly bank payment.
 - b. Prepare an amortization schedule for all five months of the loan.
23. The owner of Blue Bottle Coffee is opening a second store and needs to borrow \$93,000. Her success with her first store has made her confident that she will be able to pay off her loan quickly, so she wishes to borrow the money for only four months. Her bank has offered her a simple interest amortized loan at $9\frac{1}{8}\%$ interest.
- a. Find the size of the monthly bank payment.
 - b. Prepare an amortization schedule for all four months of the loan.
- ▶ 24. The Green Growery Nursery needs to borrow \$48,000 to increase its inventory for the upcoming summer season. The owner is confident that he will sell most, if not all, of the new plants during the summer, so he wishes to borrow the money for only four months. His bank has offered him a simple interest amortized loan at $9\frac{1}{4}\%$ interest.
- a. Find the size of the monthly bank payment.
 - b. Prepare an amortization schedule for all four months of the loan.

*For Exercises 25–28, note the following information. A **line of credit** is an agreement between a bank and a borrower by which the borrower can borrow any amount of money (up to a mutually agreed-upon maximum) at any time, simply by writing a check. Typically, monthly simple interest payments are required, and the borrower is free to make principal payments as frequently or infrequently as he or she wants. Usually, a line of credit is secured by the title to the borrower's house, and the interest paid to the bank by the borrower is deductible from the borrower's income taxes.*

25. Kevin and Roxanne Gahagan did not have sufficient cash to pay their income taxes. However, they had previously set up a line of credit with their bank. On April 15, they wrote a check to the Internal Revenue Service on their line of credit for \$6,243. The line's interest rate is 5.75%.
- a. Find the size of the required monthly interest payment.
 - b. The Gahagans decided that it would be in their best interests to get this loan paid off in eight months.

- Find the size of the monthly principal-plus-interest payment that would accomplish this. (*HINT*: In effect, the Gahagans are converting the loan to an amortized loan.)
- c. Prepare an amortization schedule for all eight months of the loan.
 - d. Find the amount of line of credit interest that the Gahagans could deduct from their taxes next year.
- ▶ 26. James and Danna Wright did not have sufficient cash to pay their income taxes. However, they had previously set up a line of credit with their bank. On April 15, they wrote a check to the Internal Revenue Service on their line of credit for \$10,288. The line's interest rate is 4.125%.
- a. Find the size of the required monthly interest payment.
 - b. The Wrights decided that it would be in their best interests to get this loan paid off in six months. Find the size of the monthly principal-plus-interest payment that would accomplish this. (*HINT*: In effect, the Wrights are converting the loan to an amortized loan.)
 - c. Prepare an amortization schedule for all six months of the loan.
 - d. Find the amount of line of credit interest that the Wrights could deduct from their taxes next year.
- ▶ 27. Homer Simpson had his bathroom remodeled. He did not have sufficient cash to pay for it. However, he had previously set up a line of credit with his bank. On July 12, he wrote a check to his contractor on his line of credit for \$12,982. The line's interest rate is 8.25%.
- a. Find the size of the required monthly interest payment.
 - b. Homer decided that it would be in his best interests to get this loan paid off in seven months. Find the size of the monthly principal-plus-interest payment that would accomplish this. (*HINT*: In effect, Simpson is converting the loan to an amortized loan.)
 - c. Prepare an amortization schedule for all seven months of the loan.
 - d. Find the amount of line of credit interest that Simpson could deduct from his taxes next year.
28. Harry Trask had his kitchen remodeled. He did not have sufficient cash to pay for it. However, he had previously set up a line of credit with his bank. On July 12, he wrote a check to his contractor on his line of credit for \$33,519. The line's interest rate is 7.625%.
- a. Find the size of the required monthly interest payment.
 - b. Harry decided that it would be in his best interests to get this loan paid off in seven months. Find the size of the monthly principal-plus-interest payment that would accomplish this. (*HINT*: In effect, Trask is converting the loan to an amortized loan.)
 - c. Prepare an amortization schedule for all seven months of the loan.
 - d. Find the amount of line of credit interest that Trask could deduct from his taxes next year.
- ▶ 29. Wade Ellis buys a car for \$16,113.82. He puts 10% down and obtains a simple interest amortized loan for the balance at $11\frac{1}{2}\%$ interest for four years. Three years and two months later, he sells his car. Find the unpaid balance on his loan.
30. Guy de Primo buys a car for \$9837.91. He puts 10% down and obtains a simple interest amortized loan for the balance at $10\frac{7}{8}\%$ interest for four years. Two years and six months later, he sells his car. Find the unpaid balance on his loan.
31. Gary Kersting buys a house for \$212,500. He puts 20% down and obtains a simple interest amortized loan for the balance at $10\frac{7}{8}\%$ interest for thirty years. Eight years and two months later, he sells his house. Find the unpaid balance on his loan.
- ▶ 32. Shirley Trembley buys a house for \$187,600. She puts 20% down and obtains a simple interest amortized loan for the balance at $11\frac{3}{8}\%$ interest for thirty years. Ten years and six months later, she sells her house. Find the unpaid balance on her loan.
33. Harry and Natalie Wolf have a three-year-old loan with which they purchased their house. Their interest rate is $13\frac{3}{8}\%$. Since they obtained this loan, interest rates have dropped, and they can now get a loan for $8\frac{7}{8}\%$ through their credit union. Because of this, the Wolfs are considering refinancing their home. Each loan is a thirty-year simple interest amortized loan, and neither has a prepayment penalty. The existing loan is for \$152,850, and the new loan would be for the current amount due on the old loan.
- a. Find their monthly payment with the existing loan.
 - b. Find the loan amount for their new loan.
 - c. Find the monthly payment with their new loan.
 - d. Find the total interest they will pay if they do *not* get a new loan.
 - e. Find the total interest they will pay if they *do* get a new loan.
 - f. Should the Wolfs refinance their home? Why or why not?
- ▶ 34. Russ and Roz Rosow have a ten-year-old loan with which they purchased their house. Their interest rate is $10\frac{5}{8}\%$. Since they obtained this loan, interest rates have dropped, and they can now get a loan for $9\frac{1}{4}\%$ through their credit union. Because of this, the Rosows are considering refinancing their home. Each loan is a 30-year simple interest amortized loan, and neither has a prepayment penalty. The existing loan is for \$112,000, and the new loan would be for the current amount due on the old loan.
- a. Find their monthly payment with the existing loan.
 - b. Find the loan amount for their new loan.

- c. Find the monthly payment with their new loan.
- d. Find the total interest they will pay if they do *not* get a new loan.
- e. Find the total interest they will pay if they *do* get a new loan.
- f. Should the Rosows refinance their home? Why or why not?
- ▶ 35. Michael and Lynn Sullivan have a ten-year-old loan for \$187,900 with which they purchased their house. They just sold their highly profitable import-export business and are considering paying off their home loan. Their loan is a thirty-year simple interest amortized loan at 10.5% interest and has no prepayment penalty.
- Find their monthly payment.
 - Find the unpaid balance of the loan.
 - Find the amount of interest they will save by prepaying.
 - The Sullivans decided that if they paid off their loan, they would deposit the equivalent of half their monthly payment into an annuity. If the ordinary annuity pays 9% interest, find its future value after 20 years.
 - The Sullivans decided that if they do not pay off their loan, they would deposit an amount equivalent to their unpaid balance into an account that pays $9\frac{3}{4}\%$ interest compounded monthly. Find the future value of this account after 20 years.
 - Should the Sullivans prepay their loan? Why or why not?
36. Charlie and Ellen Wilson have a twenty-five-year-old loan for \$47,000 with which they purchased their house. The Wilsons are retired and are living on a fixed income, so they are contemplating paying off their home loan. Their loan is a thirty-year simple interest amortized loan at $4\frac{1}{2}\%$ and has no prepayment penalty. They also have savings of \$73,000, which they have invested in a certificate of deposit currently paying $8\frac{1}{4}\%$ interest compounded monthly. Should they pay off their home loan? Why or why not?
37. Ray and Helen Lee bought a house for \$189,500. They put 10% down, borrowed 80% from their bank for thirty years at $11\frac{1}{2}\%$, and convinced the owner to take a second mortgage for the remaining 10%. That 10% is due in full in five years (this is called a *balloon payment*), and the Lees agree to make monthly interest-only payments to the seller at 12% simple interest in the interim.
- Find the Lees' down payment
 - Find the amount that the Lees borrowed from their bank.
 - Find the amount that the Lees borrowed from the seller.
 - Find the Lees' monthly payment to the bank.
 - Find the Lees' monthly interest payment to the seller.
- ▶ 38. Jack and Laurie Worthington bought a house for \$163,700. They put 10% down, borrowed 80% from their bank for thirty years at 12% interest, and convinced the owner to take a second mortgage for the remaining 10%. That 10% is due in full in five years, and the Worthingtons agree to make monthly interest-only payments to the seller at 12% simple interest in the interim.
- Find the Worthingtons' down payment
 - Find the amount that the Worthingtons borrowed from their bank.
 - Find the amount that the Worthingtons borrowed from the seller.
 - Find the Worthingtons' monthly payment to the bank.
 - Find the Worthingtons' monthly interest payment to the seller.
39. a. If the Lees in Exercise 37 save for their balloon payment with a sinking fund, find the size of the necessary monthly payment into that fund if their money earns 6% interest.
- Find the Lees' total monthly payment for the first five years.
 - Find the Lees' total monthly payment for the last twenty-five years.
- ▶ 40. a. If the Worthingtons in Exercise 38 save for their balloon payment with a sinking fund, find the size of the necessary monthly payment into that fund if their money earns 7% interest.
- Find the Worthingtons' total monthly payment for the first five years.
 - Find the Worthingtons' total monthly payment for the last twenty-five years.
41. In July 2005, Tom and Betty bought a house for \$275,400. They put 20% down and financed the rest with a thirty-year loan at the then-current rate of 6%. In 2007, the real estate market crashed. In August 2009, they had to sell their house. The best they could get was \$142,000. Was this enough to pay off the loan? If so, how much did they profit? If not, how much did they have to pay out of pocket to pay off the loan?
- ▶ 42. In August 2004, Bonnie Martin bought a house for \$395,000. She put 20% down and financed the rest with a thirty-year loan at the then-current rate of $5\frac{3}{4}\%$. In 2007, the real estate market crashed. In June 2009, she had to sell her house. The best she could get was \$238,000. Was this enough to pay off the loan? If so, how much did she profit? If not, how much did she have to pay out of pocket to pay off the loan?
43. Al-Noor Koorji bought a house for \$189,000. He put 20% down and financed the rest with a thirty-year loan at $5\frac{3}{4}\%$.

- a. Find his monthly payment.
- b. If he paid an extra \$100 per month, how early would his loan be paid off?

(*HINT*: This involves some trial-and-error work.)

- 44. The Franklins bought a house for \$265,000. They put 20% down and financed the rest with a thirty-year loan at $5\frac{3}{8}\%$.
- a. Find their monthly payment.
 - b. If they paid an extra \$85 per month, how early would their loan be paid off?

(*HINT*: This involves some trial-and-error work.)



Answer the following questions using complete sentences and your own words.

• CONCEPT QUESTIONS

45. In Exercise 11, an add-on interest amortized loan required larger payments than did a simple interest amortized loan at the same interest rate. What is there about the structure of an add-on interest loan that makes its payments larger than those of a simple interest amortized loan?
(*HINT*: The interest portion is a percentage of what quantity?)
46. Why are the computations for the last period of an amortization schedule different from those for all preceding periods?
47. Give two different situations in which a borrower would need the information contained in an amortization schedule.
48. A borrower would need to know the unpaid balance of a loan if the borrower were to prepay the loan. Give three different situations in which it might be in a borrower's best interests to prepay a loan.
(*HINT*: See some of the preceding exercises.)
49. If you double the period of an amortized loan, does your monthly payment halve? Why or why not?
50. If you double the loan amount of an amortized loan, does your monthly payment double? Why or why not?

• WEB PROJECTS

51. This is an exercise in buying a car. It involves choosing a car and selecting the car's financing. Write a paper describing all of the following points.
- a. You might not be in a position to buy a car now. If that is the case, fantasize about your future. What job do you have? How long have you had that job? What is your salary? If you are married, does your spouse work? Do you have a family? What needs will your car fulfill? Make your fantasy realistic. Briefly describe what has happened between the present and your future fantasy. (If you are in a

position to buy a car now, discuss these points on a more realistic level.)

- b. Go shopping for a car. Look at new cars, used cars, or both. Read newspaper and magazine articles about your choices (see, for example, *Consumer Reports*, *Motor Trend*, and *Road and Track*). Discuss in detail the car you selected and why you did so. How will your selection fulfill your (projected) needs? What do newspapers and magazines say about your selection? Why did you select a new or a used car?
 - c. Go to the web sites of four different banks. Get all of the information you need about a car loan. Perform all appropriate computations yourself—do not have the lenders tell you the payment size, and do not use web calculators. Summarize the appropriate data (including the down payment, payment size, interest rate, duration of loan, type of loan, and loan fees) in your paper, and discuss which loan you would choose. Explain how you would be able to afford your purchase.
52. This is an exercise in buying a home. It involves choosing a home and selecting the home's financing. Write a paper describing all of the following points.
- a. You might not be in a position to buy a home now. If this is the case, fantasize about your future. What job do you have? How long have you had that job? What is your salary? If you are married, does your spouse work? Do you have a family? What needs will your home fulfill? Make your future fantasy realistic. Briefly describe what has happened between the present and your future fantasy. (If you are in a position to buy a home now, discuss these points on a more realistic level.)
 - b. Go shopping for a home. Look at houses, condominiums, or both. Look at new homes, used homes, or both. Used homes can easily be visited by going to an "open house," where the owners are gone and the real estate agent allows interested parties to inspect the home. Open houses are probably listed in your local newspaper. Read appropriate newspaper and magazine articles (for example, in the real estate section of your local newspaper). Discuss in detail the home you selected and why you did so. How will your selection fulfill your (projected) needs? Why did you select a house or a condominium? Why did you select a new or a used home? Explain your choice of location, house size, and features of the home.
 - c. Go to the web sites of four different banks. Get all of the information you need about a home loan. Perform all appropriate computations yourself—do not have the lenders tell you the payment size, and do not use web calculators. Summarize the appropriate data in your paper and discuss which loan you would choose. Include in your discussion the down

payment, the duration of the loan, the interest rate, the payment size, and other terms of the loan.

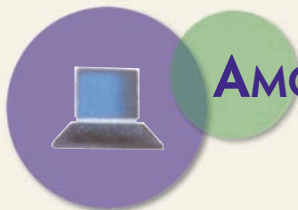
- d. Also discuss the real estate taxes (your instructor will provide you with information on the local tax rate) and the effect of your purchase on your income taxes (interest paid on a home loan is deductible from the borrower's income taxes).
- e. Most lenders will approve a home loan only if the total of all the borrower's monthly payments, including the home loan payment, real estate taxes, credit card payments, and car loan payments, is no more than 38% of the borrower's monthly income. Discuss your ability to qualify for the loan.

Some useful links for the above web projects are listed on the text web site: www.cengage.com/math/johnson

• PROJECTS

For Exercises 53–55, note the following information. An **adjustable-rate mortgage** (or ARM) is, as the name implies, a mortgage in which the interest rate is allowed to change. As a result, the payment changes too. At first, an ARM costs less than a fixed-rate mortgage—its initial interest rate is usually two or three percentage points lower. As time goes by, the rate is adjusted. As a result, it might or might not continue to hold this advantage.

53. Trustworthy Savings offers a thirty-year adjustable-rate mortgage with an initial rate of 5.375%. The rate and the required payment are adjusted annually. Future rates are set at 2.875 percentage points above the 11th District Federal Home Loan Bank's cost of funds. Currently, that cost of funds is 4.839%. The loan's rate is not allowed to rise more than two percentage points in any one adjustment, nor is it allowed to rise above 11.875%. Trustworthy Savings also offers a thirty-year fixed-rate mortgage with an interest rate of 7.5%.
 - a. Find the monthly payment for the fixed-rate mortgage on a loan amount of \$100,000.
 - b. Find the monthly payment for the ARM's first year on a loan amount of \$100,000.
 - c. How much would the borrower save in the mortgage's first year by choosing the adjustable rather than the fixed-rate mortgage?
 - d. Find the unpaid balance at the end of the ARM's first year.
 - e. Find the interest rate and the value of n for the ARM's second year if the 11th District Federal Home Loan Bank's cost of funds does not change during the loan's first year.
 - f. Find the monthly payment for the ARM's second year if the 11th District Federal Home Loan Bank's cost of funds does not change during the loan's first year.
 - g. How much would the borrower save in the mortgage's first two years by choosing the adjustable-rate mortgage rather than the fixed-rate mortgage if the cost of funds does not change?
 - h. Discuss the advantages and disadvantages of an adjustable-rate mortgage.
54. American Dream Savings Bank offers a thirty-year adjustable-rate mortgage with an initial rate of 4.25%. The rate and the required payment are adjusted annually. Future rates are set at three percentage points above the one-year Treasury bill rate, which is currently 5.42%. The loan's rate is not allowed to rise more than two percentage points in any one adjustment, nor is it allowed to rise above 10.25%. American Dream Savings Bank also offers a thirty-year fixed-rate mortgage with an interest rate of 7.5%.
 - a. Find the monthly payment for the fixed-rate mortgage on a loan amount of \$100,000.
 - b. Find the monthly payment for the ARM's first year on a loan amount of \$100,000.
 - c. How much would the borrower save in the mortgage's first year by choosing the adjustable rather than the fixed-rate mortgage?
 - d. Find the unpaid balance at the end of the ARM's first year.
 - e. Find the interest rate and the value of n for the ARM's second year if the one-year Treasury bill rate does not change during the loan's first year.
 - f. Find the monthly payment for the ARM's second year if the Treasury bill rate does not change during the loan's first year.
 - g. How much would the borrower save in the mortgage's first two years by choosing the adjustable-rate mortgage rather than the fixed-rate mortgage if the Treasury bill rate does not change?
 - h. Discuss the advantages and disadvantages of an adjustable-rate mortgage.
55. Bank Two offers a thirty-year adjustable-rate mortgage with an initial rate of 3.95%. This initial rate is in effect for the first six months of the loan, after which it is adjusted on a monthly basis. The monthly payment is adjusted annually. Future rates are set at 2.45 percentage points above the 11th District Federal Home Loan Bank's cost of funds. Currently, that cost of funds is 4.839%.
 - a. Find the monthly payment for the ARM's first year on a loan amount of \$100,000.
 - b. Find the unpaid balance at the end of the ARM's first six months.
 - c. Find the interest portion of the seventh payment if the cost of funds does not change.
 - d. Usually, the interest portion is smaller than the monthly payment, and the difference is subtracted from the unpaid balance. However, the interest portion found in part (c) is larger than the monthly payment found in part (a), and the difference is added to the unpaid balance found in part (b). Why would this difference be added to the unpaid balance? What effect will this have on the loan?
 - e. The situation described in part (d) is called **negative amortization**. Why?
 - f. What is there about the structure of Bank Two's loan that allows negative amortization?



AMORTIZATION SCHEDULES ON A COMPUTER

Interest paid on a home loan is deductible from the borrower's income taxes, and interest paid on a loan by a business is usually deductible. A borrower with either of these types of loans needs to know the total interest paid on the loan during the final year. The way to determine the total interest paid during a given year is to prepare an amortization schedule for that year. Typically, the lender provides the borrower with an amortization schedule, but it is not uncommon for this schedule to arrive after taxes are due. In this case, the borrower must either do the calculation personally or pay taxes without the benefit of the mortgage deduction and then file an amended set of tax forms after the amortization schedule has arrived.

Computing a year's amortization schedule is rather tedious, and neither a scientific calculator nor a graphing calculator offers relief. The best tool for the job is a computer, combined either with the Amortrix computer program (available with this book) or with a computerized spreadsheet. Each is discussed below.

Amortization Schedules and Amortrix

Amortrix is one of the features of the text web site (www.cengage.com/math/johnson). This software will enable you to quickly and easily compute an amortization schedule for any time period. We'll illustrate this process by using Amortrix to prepare an amortization schedule for a fifteen-year \$175,000 loan at 7.5% interest with monthly payments.

When you start Amortrix, a main menu appears. Click on the "Amortization Schedule" option. Once you're at the page labeled "Amortization Schedule," enter "175000" (without commas or dollar signs) for the Loan Amount (P), "7.5" (without a percent sign) for the Annual Interest Rate, and "180" for the Total Number of Payment Periods (n). See the top part of Figure 5.21. Click on "Calculate" and the software will create an amortization schedule. See the bottom part of Figure 5.21.

Amortrix: Loan Amortization -- WatchMe Ware Productions				
Loan Amount (P):	175000			
Annual Interest Rate:	7.5			
Payments Period:	Monthly (12)			
Total Number of Payment Periods (n):	180			
Calculate		Clear Values		
Payment Number	Principal Portion	Interest Portion	Total Payment	Balance
1	\$528.52	\$1093.75	\$1622.27	\$174471.48
2	\$531.82	\$1090.45	\$1622.27	\$173939.66
3	\$535.15	\$1087.12	\$1622.27	\$173404.51
4	\$538.49	\$1083.78	\$1622.27	\$172866.02
5	\$541.86	\$1080.41	\$1622.27	\$172324.16
6	\$545.24	\$1077.03	\$1622.27	\$171778.92
7	\$548.65	\$1073.62	\$1622.27	\$171230.27
8	\$552.08	\$1070.19	\$1622.27	\$170678.19
9	\$555.53	\$1066.74	\$1622.27	\$170122.66
10	\$559.00	\$1063.27	\$1622.27	\$169563.66
11	\$562.50	\$1059.77	\$1622.27	\$169001.16
12	\$566.01	\$1056.26	\$1622.27	\$168435.15
13	\$569.55	\$1052.72	\$1622.27	\$167865.60
14	\$573.11	\$1049.16	\$1622.27	\$167292.49
15	\$576.69	\$1045.58	\$1622.27	\$166715.80
16	\$580.30	\$1041.97	\$1622.27	\$166135.50
17	\$583.92	\$1038.35	\$1622.27	\$165551.58
18	\$587.57	\$1034.70	\$1622.27	\$164964.01
19	\$591.24	\$1031.03	\$1622.27	\$164372.77
Total Principal Paid		Total Interest Paid	Total Payments	
\$174999.44		\$117009.16	\$292008.60	

FIGURE 5.21 Part of an amortization schedule prepared with Amortrix.

The software will *not* correctly compute the last loan payment; it will compute the last payment in the same way it computes all other payments rather than in the way shown earlier in this section. You will have to correct this last payment if you use the computer to prepare an amortization schedule for a time period that includes the last payment.

Amortization Schedules and Excel

A **spreadsheet** is a large piece of paper marked off in rows and columns. Accountants use spreadsheets to organize numerical data and perform computations. A **computerized spreadsheet** such as Microsoft Excel is a computer program that mimics the appearance of a paper spreadsheet. It frees the user from performing any computations.

When you start a computerized spreadsheet, you see something that looks like a table waiting to be filled in. The rows are labeled with numbers and the columns are labeled with letters, as shown in Figure 5.22. The individual boxes are called **cells**. The cell in column A, row 1 is called cell A1.

	A	B	C	D	E	F	G
1	payment number	principal portion	interest portion	total payment	balance		
2	0				\$175,000.00	rate	7.50%
3						years	15
4						payments/yr	12
5							

FIGURE 5.22 The spreadsheet after step 1.

Excel is an ideal tool to use in creating an amortization schedule. We will illustrate this process by preparing the amortization schedule for a fifteen-year \$175,000 loan at 7.5% interest with monthly payments.

1. *Set up the spreadsheet.* Start by entering the information shown in Figure 5.22.
 - Adjust the columns' widths. To make column A wider, place your cursor on top of the line dividing the "A" and "B" labels. Then use your cursor to move that dividing line.
 - Format cell E2 as currency by highlighting that cell and pressing the "\$" button on the Excel ribbon at the top of the screen.
 - Be certain that you include the % symbol in cell G2.
 - Save the spreadsheet.

If you have difficulty creating a spreadsheet that looks like this, you may download it from our web site at www.cengage.com/math/johnson. See the file 5.4.xls and select the "step 1" sheet. Other sheets show later steps.

2. *Compute the monthly payment.* While you can do this with your calculator, as we discussed earlier in this section, you can also use the Excel "PMT" function to compute the monthly payment. This allows us to change the rate (in cell G2) or the loan amount (in cell E2) and Excel will automatically compute the new payment and change the entire spreadsheet accordingly. To use the "PMT" function, do the following:
 - Click on cell D3, where we will put the total payment.
 - Click on the *fx* button at the top of the screen.
 - In the resulting "Paste Function" box, select "Financial" under "Function Category", select "PMT" under "Function name", and press "OK." A "PMT" box will appear.

- Cell G2 holds 7.5%, the annual interest rate, and cell G4 holds 12, the number of periods/year. The periodic interest rate is $7.5\%/12 = G2/G4$. In the PMT box after “Rate”, type “ $\$G\$2/\$G\4 ”.
- The number of periods is $15 \cdot 12 = G3 \cdot G4$. In the PMT box after “Nper”, type “ $\$G\$3*\$G\4 .”
- After “PV”, type “ $\$E\2 ”. This is where we have stored the present value of the loan.
- Do not type anything after “Fv” or “type”.
- Press “OK” at the bottom of the “PMT” box, and “ $(\$1,622.27)$ ” should appear in cell D2. The parentheses mean that the number is negative.
- Save the spreadsheet.

Doing the above creates two problems that we must fix. The payment is negative. We can fix that with an extra minus sign. Also, the payment is not rounded to the nearest penny. We can fix that with the ROUND function.

In cell D2, Excel displays $(\$1622.27)$, but it stores $-1622.27163\dots$, and all calculations involving cell D2 will be done by using $-1622.27163\dots$. This will make the amortization schedule incorrect. To fix this, click on cell D3 and the long box at the top of the screen will have

$$=PMT(\$G\$2/\$G\$4,\$G\$3*\$G\$4,\$E\$2)$$

in it. Change this to

$$=ROUND(-PMT(\$G\$2/\$G\$4,\$G\$3*\$G\$4,\$E\$2),2)$$

↑ ↑ ↑
the round fuction a new minus sign part of the round function

The new parts here are “**ROUND(-**” at the beginning and “**,2)**” at the end. The “**ROUND (,2)**” part tells Excel to round to two decimal places, as a bank would. The minus sign after “**ROUND(**” makes the payment positive. Place your cursor where the new parts go, and add them.

3. Fill in row 3.

- We need to add a year each time we go down one row. Type “ $= A2 + 1$ ” in cell A3.
- Cell C3 should contain instructions on computing simple interest on the previous balance.

$$\begin{aligned} I &= P \cdot r \cdot t \\ &= \text{the previous balance} \cdot 0.075 \cdot 1/12 \\ &= E2 * \$G\$2 / \$G\$4 \end{aligned}$$

Move to cell C3 and type in “ $=ROUND(E2*\$G\$2/\$G\$4,2)$ ” and press “return”. The “**ROUND (,2)**” part rounds the interest to two decimal places, as a bank would.

- Cell B3 should contain instructions on computing the payment’s principal portion.

$$\begin{aligned} \text{Principal portion} &= \text{payment} - \text{interest portion} \\ &= D3 - C3 \end{aligned}$$

In cell B3, type “ $= D3 - C3$ ”.

- Cell E3 should contain instructions on computing the new balance.

$$\begin{aligned} \text{New balance} &= \text{previous balance} - \text{principal portion} \\ &= E2 - B3 \end{aligned}$$

In cell E3, type “ $= E2 - B3$ ”.

See Figure 5.23.

◇	A	B	C	D	E	F	G
1	payment number	principal portion	Interest portion	total payment	balance		
2	0				\$175,000.00	rate	7.50%
3	1	\$ 528.52	\$1,093.75	\$1,622.27	\$174,471.48	years	15
4						payments/yr	12
5							

FIGURE 5.23 The spreadsheet after completing step 3.

4. *Fill in rows 4 and beyond.* All of the remaining payments' computations are just like payment 1 computations (except for the last payment), so all we have to do is copy the payment 1 instructions in row 3 and paste them in rows 4 and beyond. After completion of this step, the top and bottom of your spreadsheet should look like that in Figure 5.24.

◇	A	B	C	D	E	F	G
1	payment number	principal portion	interest portion	total payment	balance		
2	0				\$175,000.00	rate	7.50%
3	1	\$ 528.52	\$1,093.75	\$1,622.27	\$174,471.48	years	15
4	2	\$ 531.82	\$1,090.45	\$1,622.27	\$173,939.66	payments/yr	12
5	3	\$ 535.15	\$1,087.12	\$1,622.27	\$173,404.51		
6	4	\$ 538.49	\$1,083.78	\$1,622.27	\$172,866.02		

The top of the spreadsheet, and ...

178	176	\$ 1,572.51	\$49.76	\$1,622.27	\$ 6,389.50
179	177	\$ 1,582.34	\$39.93	\$1,622.27	\$ 4,807.16
180	178	\$ 1,592.23	\$30.04	\$1,622.27	\$ 3,214.93
181	179	\$ 1,602.18	\$20.09	\$1,622.27	\$ 1,612.75
182	180	\$ 1,612.19	\$10.08	\$1,622.27	\$ 0.56
183					

... the bottom of the spreadsheet, after completing step 4

FIGURE 5.24 The final spreadsheet.

5. *Fix the last payment.* Clearly, the ending balance of \$0.56 in cell E182 is not correct—we cannot owe \$0.56 or any other amount after making our last payment. See Exercise 56.
6. *Find the interest paid.* Typing “SUM(C3:C6)” in a cell will result in the sum of cells C4, C4, C5, and C6 appearing in that cell. Using this function makes it easy to find the total interest paid for any time period.

EXERCISES

In Exercises 56–62, first use Amortrix or Excel to create an amortization schedule for a fifteen-year \$175,000 loan at 7.5% interest with monthly payments. (Figure 5.24 shows the first and last lines of this amortization schedule, created with Excel.) You will alter parts of this amortization schedule to answer Exercises 56–62.

56.
 - a. The very last line of your schedule is incorrect, because the calculations for the last line of any amortization schedule are done differently from those of all other lines. What should the last line be?
 - b. If you use Excel, state exactly what you should type in row 182 to fix the last payment.

57. Use Excel or Amortrix to find the total interest paid in
- the loan's first year and
 - the loan's last year if the first month of the loan is January.
 - If you use Excel, state exactly what you type in parts (a) and (b) to find the total interest paid.
58. Right before you signed your loan papers, the interest rate dropped from 7.5% to 7%.
- How does this affect the total payment?
 - How does this affect the total interest paid in the loan's first year if the first month of the loan is January?
 - How does this affect the total interest paid in the loan's last year?
59. Right before you signed your loan papers, you convinced the seller to accept a smaller price. As a result, your loan amount dropped from \$175,000 to \$165,000. (The interest rate remains at 7.5%)
- How does this affect the total payment?
 - How does this affect the total interest paid in the loan's first year if the first month of the loan is January?
 - How does this affect the total interest paid in the loan's last year?
60. You are considering a twenty-year loan. (The interest rate remains at 7.5% and the loan amount remains at \$175,000.)
- How does this affect the total payment?
 - How does this affect the total interest paid in the loan's first year if the first month of the loan is January?
 - How does this affect the total interest paid in the loan's fifteenth year?
 - Explain, without performing any calculations, why the total payment should go down and the total interest paid during the first year should go up.
61. You are considering a thirty-year loan. (The interest rate remains at 7.5%, and the loan amount remains at \$175,000.)
- How does this affect the total payment?
 - How does this affect the total interest paid in the loan's first year if the first month of the loan is January?
 - How does this affect the total interest paid in the loan's fifteenth year?
 - Explain, without performing any calculations, why the total payment should go down and the total interest paid during the first year should go up.
62. You are considering a loan with payments every two weeks. (The interest rate remains at 7.5%, the loan amount remains at \$175,000, and the term remains at fifteen years.)
- How does this affect the total payment?
 - How does this affect the total interest paid in the loan's first year if the first month of the loan is January?
 - How does this affect the total interest paid in the loan's last year?
 - Explain, without performing any calculations, why the total payment and the total interest paid during the first year should both go down.
- In Exercises 63–66:*
- Use Amortrix or Excel to prepare an amortization schedule for the given loan.
 - Find the amount that could be deducted from the borrower's taxable income (that is, find the total interest paid) in the loan's first year if the first payment was made in January. (Interest on a car loan is deductible only in some circumstances. For the purpose of these exercises, assume that this interest is deductible.)
 - Find the amount that could be deducted from the borrower's taxable income (that is, find the total interest paid) in the loan's last year.
 - If you sell your home or car after three years, how much will you still owe on the mortgage?
63. A five-year simple interest amortized car loan for \$32,600 at 12.25% interest
64. A four-year simple interest amortized car loan for \$26,200 at 5.75% interest
65. A fifteen-year simple interest amortized home loan for \$220,000 at 6.25% interest
66. A thirty-year simple interest amortized home loan for \$350,000 at 14.5% interest
67. Use Amortrix or Excel to do Exercise 29 on page 379. Is your answer the same as that of Exercise 29? If not, why not? Which answer is more accurate? Why?
68. Use Amortrix or Excel to do Exercise 30 on page 379. Is your answer the same as that of Exercise 30? If not, why not? Which answer is more accurate? Why?
69. Use Amortrix or Excel to do Exercise 31 on page 379. Is your answer the same as that of Exercise 31? If not, why not? Which answer is more accurate? Why?
70. Use Amortrix or Excel to do Exercise 32 on page 379. Is your answer the same as that of Exercise 32? If not, why not? Which answer is more accurate? Why?
71. *For Excel users only:* When interest rates are high, some lenders offer the following type of home loan. You borrow \$200,000 for thirty years. For the first five years, the interest rate is 5.75%, and the payments are calculated as if the interest rate were going to remain unchanged for the life of the loan. For the last twenty-five years, the interest rate is 9.25%.
- What are the monthly payments for the first five years?
 - What are the monthly payments for the last twenty-five years?
 - What is the total interest paid, during the loan's fifth year if the first month of the loan is January?
 - What is the total interest paid during the loan's sixth year?
 - Why would a lender offer such a loan?

5.5

Annual Percentage Rate with a TI's TVM Application

OBJECTIVES

- Understand what an annual percentage rate (APR) is
- Find an APR
- Use APRs to compare loans

A **simple interest loan** is any loan for which the interest portion of each payment is simple interest on the outstanding principal. A *simple interest amortized loan* fulfills this requirement; in fact, we compute an amortization schedule for a simple interest amortized loan by finding the simple interest on the outstanding principal.

The APR of an Add-on Interest Loan

Whenever a loan is not a simple interest loan, the Truth in Lending Act requires the lender to disclose the annual percentage rate to the borrower. The **annual percentage rate (APR) of an add-on interest loan** is the simple interest rate that makes the dollar amounts the same if the loan is recomputed as a simple interest amortized loan.

We'll explore APRs in this section. They are important because they allow someone who is shopping for a loan to compare loans with different terms and to determine which loan is better.

EXAMPLE 1

FINDING THE APR In Example 5 in Section 5.1, Chip Douglas bought a nine-year-old Ford that's "like new" from Centerville Auto Sales for \$5,988. He financed the purchase with the dealer's two-year, 5% add-on interest loan, which required a \$600 down payment and monthly payments of \$246.95 a month for twenty-four months. Use a TI-83/84's TVM application to find the APR of the loan.

SOLUTION

1. Press **[APPS]**, select option 1: "Finance", and press **[ENTER]**.
2. Select option 1: "TVM Solver", and press **ENTER**.
3. Enter appropriate values for the variables:
 - N is the number of payments. Enter 24 for N.
 - I% is the annual interest rate. This is the number we're trying to find, so temporarily enter 0. Later, we'll solve for the actual value of N.
 - PV is the present value, which is $\$5988 - \$600 = \$5388$.
 - PMT is the payment. The payment is \$246.95, but it's an outgoing amount of money (since we give it to the bank), so we enter -246.95 for PMT. You must enter a negative number for any outgoing amount of money.
 - FV is the future value. There is no future value here, so we enter 0.
 - P/Y is the number of periods per year. We make monthly payments, so there are twelve periods per year. Enter 12 for P/Y.
 - C/Y is automatically made to be the same as P/Y. In this text, we will never encounter a situation in which C/Y is different from P/Y. See Figure 5.25.
4. To solve for I%, use the arrow buttons to highlight the 0 that we entered for it earlier. Then press **[ALPHA]** **[SOLVE]**. (Pressing **[ALPHA]** makes the **[ENTER]** button become the

```
N=24
I%=0
PV=5388
PMT=-246.95
FV=0
P/Y=12
C/Y=12
PMT:END BEGIN
```

FIGURE 5.25

Preparing the TVM screen.

HISTORICAL NOTE

TRUTH IN LENDING ACT

The Truth in Lending Act was signed by President Lyndon Johnson in 1968. Its original intent was to promote credit shopping by requiring lenders to use uniform language and calculations and to make full disclosure of credit



Courtesy of Library of Congress

charges so that consumers could shop for the most favorable credit terms. The Truth in Lending Act is interpreted by the Federal Reserve Board's Regulation Z.

During the 1970s, Congress amended the act many times, and corresponding changes were made

in Regulation Z. These changes resulted in a huge increase in the length and complexity of the law. Its scope now goes beyond disclosure to grant significant legal rights to the borrower.

SOLVE button.) As a result, we find that I% is $9.323544\% \approx 9.32\%$, and the add-on interest loan has an APR of 9.32%. This means that the 5% add-on interest loan requires the same monthly payment that a 9.32% simple interest amortized loan would have!

The Truth in Lending Act, which requires a lender to divulge its loans' APR, allows a tolerance of one-eighth of 1% (0.125%) in the claimed APR. Thus, the dealership would be legally correct if it stated that the APR was between $9.323544\% - 0.125\% = 9.198544\%$ and $9.323544\% + 0.125\% = 9.448544\%$.

The APR of a Simple Interest Loan

Sometimes **finance charges** other than the interest portion of the monthly payment are associated with a loan; these charges must be paid when the loan agreement is signed. For example, a **point** is a finance charge that is equal to 1% of the loan amount; a **credit report fee** pays for a report on the borrower's credit history, including any late or missing payments and the size of all outstanding debts; an **appraisal fee** pays for the determination of the current market value of the property (the auto, boat, or home) to be purchased with the loan. The Truth in Lending Act requires the lender to inform the borrower of the total finance charge, which includes the interest, the points, and some of the fees (see Figure 5.26). Most of the finance

Costs Included in the Prepaid Finance Charges

2 points	\$2,415.32
prorated interest	\$1,090.10
prepaid mortgage insurance	\$ 434.50
loan fee	\$1,242.00
document preparation fee	\$ 80.00
tax service fee	\$ 22.50
processing fee	\$ 42.75
subtotal	\$5,327.17

Other Costs Not Included in the Finance Charge

appraisal fee	\$ 70.00
credit report	\$ 60.00
closing fee	\$ 670.00
title insurance	\$ 202.50
recording fee	\$ 20.00
notary fee	\$ 20.00
tax and insurance escrow	\$ 631.30
subtotal	\$1,673.80

FIGURE 5.26 Sample portion of a federal truth-in-lending disclosure statement (loan amount = \$120,765.90).

charges must be paid before the loan is awarded, so in essence the borrower must pay money now in order to get more money later. The law says that this means the lender is not really borrowing as much as he or she thinks. According to the law, the actual amount loaned is the loan amount minus all points and those fees included in the finance charge. The **APR of a simple interest amortized loan** is the rate that reconciles the payment and this actual loan amount.

EXAMPLE 2

FINDING THE APR OF A SIMPLE INTEREST LOAN Glen and Tanya Hansen bought a home for \$140,000. They paid the sellers a 20% down payment and obtained a simple interest amortized loan for the balance from their bank at $10\frac{3}{4}\%$ for thirty years. The bank in turn paid the sellers the remaining 80% of the purchase price, less a 6% sales commission paid to the sellers' and the buyers' real estate agents. (The transaction is illustrated in Figure 5.27.) The bank charged the Hansens 2 points plus fees totaling \$3,247.60; of these fees, \$1,012.00 were included in the finance charge.

- Find the size of the Hansens' monthly payment.
- Find the total interest paid.
- Compute the total finance charge.
- Find the APR

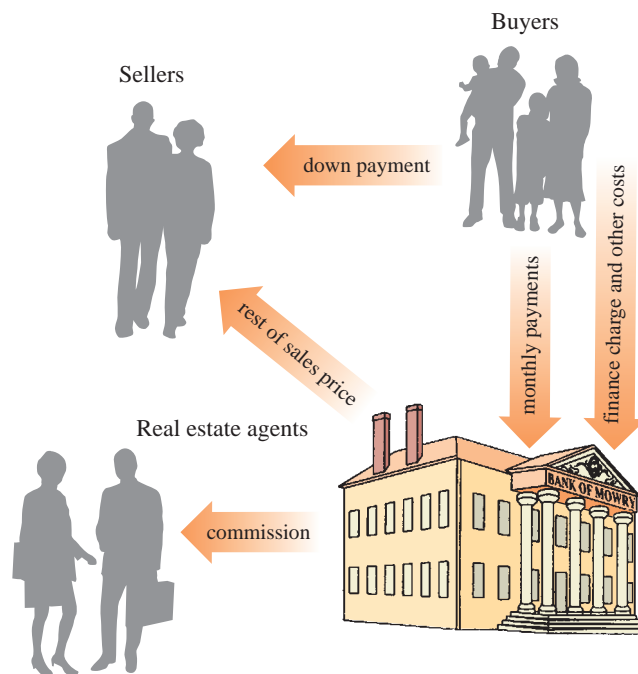


FIGURE 5.27 Buying a home.

SOLUTION

- Finding their monthly payment* We are given down payment = 20% of \$140,000 = \$28,000, P = loan amount = \$140,000 - \$28,000 = \$112,000, $i = \frac{1}{12}$ of $10\frac{3}{4}\% = \frac{0.1075}{12}$, and $n = 30$ years = 30 years \cdot (12 months)/(1 year) = 360 months.

Future value of annuity = future value of loan amount

$$pymt \frac{(1+i)^n - 1}{i} = P(1+i)^n$$

$$pymt \frac{\left(1 + \frac{0.1075}{12}\right)^{360} - 1}{\frac{0.1075}{12}} = 112,000 \left(1 + \frac{0.1075}{12}\right)^{360}$$

Computing the fraction on the left side and multiplying its reciprocal by the right side, we get

$$pymt = 1,045.4991 \approx \$1,045.50$$

- b. Finding the total interest paid** The total interest paid is the total amount paid minus the amount borrowed. The Hansens agreed to make 360 monthly payments of \$1045.50 each, for a total of $360 \cdot \$1,045.50 = \$376,380.00$. Of this, \$112,000 is principal; therefore, the total interest is

$$376,380 - 112,000 = \$264,380$$

- c. Computing their total finance charge**

$$2 \text{ points} = 2\% \text{ of } \$112,000 = \$ 2,240$$

$$\text{included fees} = \$ 1,012$$

$$\text{total interest paid} = \underline{\$264,380} \quad \text{from part (b)}$$

$$\text{total finance charge} = \$267,632$$

- d. Finding the APR** The APR is the simple interest rate that makes the dollar amounts the same if the loan is recomputed using the legal loan amount (loan amount less points and fees) in place of the actual loan amount. The legal loan amount is

$$P = \$112,000 - \$3252 = \$108,748$$

We'll find the APR with the TVM application.

1. Press **APPS**, select option 1: "Finance", and press **ENTER**.
2. Select option 1: "TVM Solver", and press **ENTER**.
3. Enter appropriate values for the variables.

- N is the number of payments. Enter 360.
- I% is the annual interest rate. This is the number we're trying to find, so temporarily enter 0. Later, we'll solve for the actual value of N.
- PV is the present value, which is \$108,748.
- PMT is the payment. The payment is \$1,045.50, but it's an outgoing amount of money (since we give it to the bank), so we enter -1045.5 for PMT. You must enter a negative number for any outgoing amount of money.
- FV is the future value. There is no future value here, so we enter 0.
- P/Y is the number of periods per year. We make monthly payments, so there are twelve periods per year. Enter 12 for P/Y.
- C/Y is automatically made to be the same as P/Y. In this text, we will never encounter a situation in which C/Y is different from P/Y. See Figure 5.28.

4. To solve for I%, use the arrow buttons to highlight the 0 that we entered for I%. Then press **ALPHA** **SOLVE**. (Pressing **ALPHA** makes the **ENTER** button become the **SOLVE** button.) As a result, we find that I% is 11.11993521% \approx 11.12%, and the loan has an APR of 11.12%. This means the $10\frac{3}{4}\%$ loan requires the same monthly payment that a 11.12% loan with no points or fees would have. In other words, the points and fees in effect increase the interest rate from $10\frac{3}{4}\%$ to 11.12%.

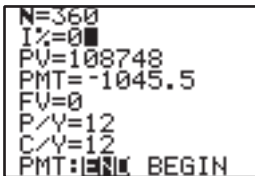


FIGURE 5.28

Preparing the TVM screen.

If you need to obtain a loan, it is not necessarily true that the lender with the lowest interest rate will give you the least expensive loan. One lender may charge more points or higher fees than does another lender. One lender may offer an add-on interest loan, while another lender offers a simple interest amortized loan. These differences can have a significant impact on the cost of a loan. If two lenders offer loans at the same interest rate but differ in any of these ways, that difference will be reflected in the APR. "Lowest interest rate" does *not* mean "least expensive loan," but "lowest APR," *does* mean "least expensive loan."

Estimating Prepaid Finance Charges

As we have seen, a borrower usually must pay an assortment of fees when obtaining a home loan. These fees can be quite substantial, and they can vary significantly from lender to lender. For example, the total fees in Figure 5.26 were \$7,000.97; the Hansens' total fees in Example 2 were \$5,487.60. When shopping for a home loan, a borrower should take these fees into consideration.

All fees must be disclosed to the borrower when he or she signs the loan papers. Furthermore, the borrower must be given an estimate of the fees when he or she applies for the loan. Before an application is made, the borrower can use a loan's APR to obtain a reasonable approximation of those fees that are included in the finance charge (for a fixed rate loan). This would allow the borrower to make a more educated decision in selecting a lender.

EXAMPLE 3

ESTIMATING PREPAID FINANCE CHARGES Felipe and Delores Lopez are thinking of buying a home for \$152,000. A potential lender advertises an 80%, thirty-year simple interest amortized loan at $7\frac{1}{2}\%$ interest with an APR of 7.95%.

- Find the size of the Lopezes' monthly payment.
- Use the APR to approximate the fees included in the finance charge.

SOLUTION

- Finding the monthly payment* The loan amount is 80% of \$152,000 = \$121,600, $i = \frac{1}{12}$ of 7.5% = $\frac{0.075}{12}$, and $n = 30$ years = 360 months.

Future value of annuity = future value of loan amount

$$pymt \frac{(1+i)^n - 1}{i} = P(1+i)^n$$

$$pymt \frac{\left(1 + \frac{0.075}{12}\right)^{360} - 1}{\frac{0.075}{12}} = 121,600 \left(1 + \frac{0.075}{12}\right)^{360}$$

Computing the fraction on the left side and multiplying its reciprocal by the right side, we get

$$pymt = 850.2448 \approx \$850.25$$

- Approximating the fees included in the finance charge* We approximate the fees by computing the legal loan amount (loan amount less points and fees), using the APR as the interest rate and the payment computed in part (a) as $pymt$. Therefore, $i = \frac{1}{12}$ of 7.95% = $\frac{0.0795}{12}$, $pymt = 850.25$.

Future value of annuity = future value of loan amount

$$pymt \frac{(1+i)^n - 1}{i} = P(1+i)^n$$

$$850.25 \frac{\left(1 + \frac{0.0795}{12}\right)^{360} - 1}{\frac{0.0795}{12}} = P \left(1 + \frac{0.0795}{12}\right)^{360}$$

Computing the left side, and dividing by $\left(1 + \frac{0.0795}{12}\right)^{360}$, we get

$$P = 116,427.6304 \approx \$116,427.63$$

This is the legal loan amount, that is, the loan amount less points and fees. The Lopezes were borrowing \$121,600, so that leaves $\$121,600 - \$116,427.63 = \$5,172.37$ in points and fees. This is an estimate of the points and fees included in the finance charge, such as a loan fee, a document preparation fee, and a processing fee; it does not include fees such as an appraisal fee, a credit report fee, and a title insurance fee.

5.5 EXERCISES

- ▶ 1. Wade Ellis buys a new car for \$16,113.82. He puts 10% down and obtains a simple interest amortized loan for the balance at $11\frac{1}{2}\%$ interest for four years. If loan fees included in the finance charge total \$814.14, find the APR.
- ▶ 2. Guy de Primo buys a new car for \$9,837.91. He puts 10% down and obtains a simple interest amortized loan for the balance at $10\frac{7}{8}\%$ interest for four years. If loan fees included in the finance charge total \$633.87, find the APR.
- ▶ 3. Chris Burditt bought a house for \$212,500. He put 20% down and obtained a simple interest amortized loan for the balance at $10\frac{7}{8}\%$ interest for thirty years. If Chris paid 2 points and \$4,728.60 in fees, \$1,318.10 of which are included in the finance charge, find the APR.
- ▶ 4. Shirley Trembley bought a house for \$187,600. She put 20% down and obtained a simple interest amortized loan for the balance at $11\frac{3}{8}\%$ for thirty years. If Shirley paid 2 points and \$3,427.00 in fees, \$1,102.70 of which are included in the finance charge, find the APR.
- ▶ 5. Jennifer Tonda wants to buy a used car that costs \$4,600. The used car dealer has offered her a four-year add-on interest loan that requires no down payment at 8% annual interest, with an APR of $14\frac{1}{4}\%$.
- Find the monthly payment.
 - Verify the APR.
6. Melody Shepherd wants to buy a used car that costs \$5,300. The used car dealer has offered her a four-year add-on interest loan that requires a \$200 down payment at 7% annual interest, with an APR of 10%.
- Find the monthly payment.
 - Verify the APR.
- ▶ 7. Anne Scanlan is buying a used car that costs \$10,340. The used car dealer has offered her a five-year add-on interest loan at 9.5% interest, with an APR of 9.9%. The loan requires a 10% down payment.
- Find the monthly payment.
 - Verify the APR.
- ▶ 8. Stan Loll bought a used car for \$9,800. The used car dealer offered him a four-year add-on interest loan at 7.8% interest, with an APR of 8.0%. The loan requires a 10% down payment.
- Find the monthly payment.
 - Verify the APR.
9. Susan Chin is shopping for a car loan. Her savings and loan offers her a simple interest amortized loan for four years at 9% interest. Her bank offers her a simple interest amortized loan for four years at 9.1% interest. Which is the less expensive loan?
10. Stephen Tamchin is shopping for a car loan. His credit union offers him a simple interest amortized loan for four years at 7.1% interest. His bank offers him a simple interest amortized loan for four years at 7.3% interest. Which is the less expensive loan?
11. Ruben Lopez is shopping for a home loan. Really Friendly Savings and Loan offers him a thirty-year simple interest amortized loan at 9.2% interest, with an APR of 9.87%. The Solid and Dependable Bank offers him a thirty-year simple interest amortized loan at 9.3% interest, with an APR of 9.80%. Which loan would have the lower payments? Which loan would be the least expensive, taking into consideration monthly payments, points, and fees? Justify your answers.
12. Keith Moon is shopping for a home loan. Sincerity Savings offers him a thirty-year simple interest amortized loan at 8.7% interest, with an APR of 9.12%. Pinstripe National Bank offers him a thirty-year simple interest amortized loan at 8.9% interest, with an APR of 8.9%. Which loan would have the lower payments? Which loan would be the least expensive, taking into consideration monthly payments, points, and fees? Justify your answers.
- ▶ 13. The Nguyens are thinking of buying a home for \$119,000. A potential lender advertises an 80%, thirty-year simple interest amortized loan at $8\frac{1}{4}\%$ interest, with an APR of 9.23%. Use the APR to approximate the fees included in the finance charge.
- ▶ 14. Ellen Taylor is thinking of buying a home for \$126,000. A potential lender advertises an 80%, thirty-year simple interest amortized loan at $10\frac{3}{4}\%$ interest, with an APR of 11.57%. Use the APR to approximate the fees included in the finance charge.
- ▶ 15. James Magee is thinking of buying a home for \$124,500. Bank of the Future advertises an 80%, thirty-year simple interest amortized loan at $9\frac{1}{4}\%$ interest, with an APR of 10.23%. R.T.C. Savings and Loan advertises an 80%, 30-year simple interest amortized loan at 9% interest with an APR of 10.16%.
- Find James's monthly payment if he borrows through Bank of the Future.
 - Find James's monthly payment if he borrows through R.T.C. Savings and Loan.
 - Use the APR to approximate the fees included in the finance charge by Bank of the Future.
 - Use the APR to approximate the fees included in the finance charge by R.T.C. Savings and Loan.
 - Discuss the advantages of each of the two loans. Who would be better off with the Bank of the Future loan? Who would be better off with the R.T.C. loan?
- ▶ 16. Holly Kresch is thinking of buying a home for \$263,800. State Bank advertises an 80%, thirty-year simple interest amortized loan at $6\frac{1}{4}\%$ interest, with an APR of 7.13%.

Boonville Savings and Loan advertises an 80%, thirty-year simple interest amortized loan at $6\frac{1}{2}\%$ interest with an APR of 7.27%.

- Find Holly's monthly payment if she borrows through State Bank.
- Find Holly's monthly payment if she borrows through Boonville Savings and Loan.
- Use the APR to approximate the fees included in the finance charge by State Bank.
- Use the APR to approximate the fees included in the finance charge by Boonville Savings and Loan.
- Discuss the advantages of the two loans. Who would be better off with the State Bank loan? Who would be better off with the Boonville loan?



Answer the following questions using complete sentences and your own words.

• CONCEPT QUESTIONS

- If the APR of a simple interest amortized home loan is equal to the loan's interest rate, what conclusions could you make about the loan's required fees and points?
- Compare and contrast the annual percentage rate of a loan with the annual yield of a compound interest rate. Be sure to discuss both the similarity and the difference between these two concepts.
- Substitute the dollar amounts from Example 1 into the simple interest amortized loan formulas. Then discuss why the resulting equation can't be solved with algebra.

5.6 Payout Annuities

OBJECTIVES

- Understand the difference between a payout annuity and an ordinary annuity
- Use the payout annuity formulas
- Determine how to use a payout annuity to save for retirement

If you save for your retirement with an IRA or an annuity or both, you'll do a lot better if you start early. This allows you to take advantage of the "magic" effect of compound interest over a long period of time.

You'll also do better if your annuity turns into a *payout annuity* when it comes to term. This allows your money to earn a higher rate of interest *after* you retire and start collecting from it. We'll explore payout annuities in this section.

The annuities that we have discussed are all savings instruments. In Section 5.3, we defined an annuity as a sequence of equal, regular payments into an account in which each payment receives compound interest. A saver who utilizes such an annuity can accumulate a sizable sum.

Annuities can be payout instruments rather than savings instruments. After you retire, you might wish to have part of your savings sent to you each month for living expenses. You might wish to receive equal, regular payments from an account where each payment has earned compound interest. Such an annuity is called a **payout annuity**. Payout annuities are also used to pay for a child's college education.

Calculating Short-Term Payout Annuities

EXAMPLE 1

UNDERSTANDING WHAT A PAYOUT ANNUITY IS On November 1, Debra Landre will make a deposit at her bank that will be used for a payout annuity. For the next three months, commencing on December 1, she will receive a payout of \$1,000 per month. Use the Compound Interest Formula to find how much money she must deposit on November 1 if her money earns 10% compounded monthly.

SOLUTION

First, calculate the principal necessary to receive \$1,000 on December 1. Use $FV = 1,000$ and $n = 1$ (interest is earned for one month).

$$FV = P(1 + i)^n$$

$$1,000 = P\left(1 + \frac{0.10}{12}\right)^1$$

$$P = 1,000 \div \left(1 + \frac{0.10}{12}\right)^1$$

$$= 991.7355 \dots \approx \$991.74$$

the Compound Interest Formula substituting
solving for P
rounding

Next, calculate the principal necessary to receive \$1,000 on January 1. Use $n = 2$ (interest is earned for two months).

$$FV = P(1 + i)^n$$

$$1,000 = P\left(1 + \frac{0.10}{12}\right)^2$$

$$P = 1,000 \div \left(1 + \frac{0.10}{12}\right)^2$$

$$= 983.5393 \dots \approx \$983.54$$

the Compound Interest Formula substituting
solving for P
rounding

Now calculate the principal necessary to receive \$1,000 on February 1. Use $n = 3$ (interest is earned for three months).

$$FV = P(1 + i)^n$$

$$1,000 = P\left(1 + \frac{0.10}{12}\right)^3$$

$$P = 1,000 \div \left(1 + \frac{0.10}{12}\right)^3$$

$$= 975.4109 \dots \approx \$975.41$$

the Compound Interest Formula substituting
solving for P
rounding

Debra must deposit the sum of the above three amounts if she is to receive three monthly payouts of \$1,000 each. Her total principal must be

$$\$991.74 + \$983.54 + \$975.41 = \$2,950.69$$



If Debra's principal received no interest, then she would need $3 \cdot \$1,000 = \$3,000$. Since her principal does receive interest, she needs slightly less than \$3,000.

Comparing Payout Annuities and Savings Annuities

Payout annuities and savings annuities are similar but not identical. It is important to understand their differences before we proceed. The following example is the “savings annuity” version of Example 1; that is, it is the savings annuity most similar to the payout annuity discussed in Example 1.

EXAMPLE 2

UNDERSTANDING THE DIFFERENCE BETWEEN A PAYOUT ANNUITY AND AN ORDINARY ANNUITY On November 1, Debra Landre set up an ordinary annuity with her bank. For the next three months, she will make a payment of \$1,000 per month. Each of those payments will receive compound interest. At the end of the three months (i.e., on February 1), she can withdraw her three \$1,000 payments plus the interest that they will have earned. Use the Compound Interest Formula to find how much money she can withdraw.

SOLUTION

Her first payment of \$1,000 would be due on November 30. It would earn interest for two months (December and January).

$$FV = P(1 + i)^n$$

the Compound Interest Formula

$$FV = 1,000 \left(1 + \frac{0.10}{12}\right)^2 \approx \$1,016.74$$

substituting

Her second payment of \$1,000 would be due on December 31. It would earn interest for one month (January).

$$FV = P(1 + i)^n$$

the Compound Interest Formula

$$FV = 1,000 \left(1 + \frac{0.10}{12}\right)^1 \approx \$1,008.33$$

substituting

Her third payment would be due on January 31. It would earn no interest (since the annuity expires February 1), so its future value is \$1,000. On February 1, Debra could withdraw

$$\$1,016.74 + \$1,008.33 + \$1,000 = \$3,025.07$$

Naturally, we could have found the future value of Debra's ordinary annuity more easily with the Ordinary Annuity Formula:

$$FV(\text{ordinary}) = pymt \frac{(1 + i)^n - 1}{i}$$

the Ordinary Annuity Formula

$$= 1,000 \frac{\left(1 + \frac{0.10}{12}\right)^3 - 1}{\frac{0.10}{12}}$$

substituting

$$\approx \$3,025.07$$

rounding

The point of the method shown in Example 2 is to illustrate the differences between a savings annuity and a payout annuity.

Calculating Long-Term Payout Annuities

The procedure used in Example 1 reflects what actually happens with payout annuities, and it works fine for a small number of payments. However, most annuities are long-term. In the case of a savings annuity, we do not need to calculate the future value of each individual payment, as we did in Example 2; instead, we can use the Ordinary Annuity Formula. We need such a formula for payout annuities. We can find a formula if we look more closely at how the payout annuity from Example 1 compares with the savings annuity from Example 2.

In Example 2, we found that the future value of an ordinary annuity with three \$1,000 payments is

$$FV = 1,000 \left(1 + \frac{0.10}{12}\right)^2 + 1,000 \left(1 + \frac{0.10}{12}\right)^1 + 1,000$$

In Example 1, we found that the total principal necessary to generate three \$1,000 payouts is

$$P = 1,000 \div \left(1 + \frac{0.10}{12}\right)^1 + 1,000 \div \left(1 + \frac{0.10}{12}\right)^2 + 1,000 \div \left(1 + \frac{0.10}{12}\right)^3$$

This can be rewritten, using exponent laws, as

$$P = 1,000 \left(1 + \frac{0.10}{12}\right)^{-1} + 1,000 \left(1 + \frac{0.10}{12}\right)^{-2} + 1,000 \left(1 + \frac{0.10}{12}\right)^{-3}$$

This is quite similar to the future value of the ordinary annuity—the only difference is the exponents. If we multiply each side by $\left(1 + \frac{0.10}{12}\right)^3$, even the exponents will match.

$$\begin{aligned} P \left(1 + \frac{0.10}{12}\right)^3 &= 1,000 \left(1 + \frac{0.10}{12}\right)^{-1} \left(1 + \frac{0.10}{12}\right)^3 \\ &\quad + 1,000 \left(1 + \frac{0.10}{12}\right)^{-2} \left(1 + \frac{0.10}{12}\right)^3 \\ &\quad + 1,000 \left(1 + \frac{0.10}{12}\right)^{-3} \left(1 + \frac{0.10}{12}\right)^3 \\ P \left(1 + \frac{0.10}{12}\right)^3 &= 1,000 \left(1 + \frac{0.10}{12}\right)^2 + 1,000 \left(1 + \frac{0.10}{12}\right)^1 + 1,000 \end{aligned}$$

The right side of the above equation is the future value of an ordinary annuity, so we can use the Ordinary Annuity Formula to rewrite it.

$$P \left(1 + \frac{0.10}{12}\right)^3 = 1,000 \frac{\left(1 + \frac{0.10}{12}\right)^3 - 1}{\frac{0.10}{12}}$$

If we generalize by replacing $\frac{0.10}{12}$ with i , 3 with n , and 1,000 with $pymt$, we have our Payout Annuity Formula.

PAYOUT ANNUITY FORMULA

$$P(1 + i)^n = pymt \frac{(1 + i)^n - 1}{i}$$

where P is the total principal necessary to generate n payouts, $pymt$ is the size of the payout, and i is the periodic interest rate.

We have seen this formula before. In Section 5.3, we used it to find the present value P of a savings annuity. In Section 5.4, we used it to find the payment of a simple interest amortized loan. In this section, we use it to find the required principal for a payout annuity. This is a versatile formula.

The following example involves a long-term annuity. Usually, the interest rate of a long-term annuity varies somewhat from year to year. In this case, calculations must be viewed as predictions, not guarantees.

EXAMPLE 3

FINDING THE NECESSARY DEPOSIT Fabiola Macias is about to retire, so she is setting up a payout annuity with her bank. She wishes to receive a payout of \$1,000 per month for the next twenty-five years. Use the Payout Annuity Formula to find how much money she must deposit if her money earns 10% compounded monthly.

SOLUTION

We are given that $pymt = 1,000$, $i = \frac{1}{12}$ of 10% = $\frac{0.10}{12}$, and $n = 25 \cdot 12 = 300$.

$$P(1 + i)^n = pymt \frac{(1 + i)^n - 1}{i} \quad \text{the Payout Annuity Formula}$$

$$P \left(1 + \frac{0.10}{12} \right)^{300} = 1,000 \frac{\left(1 + \frac{0.10}{12} \right)^{300} - 1}{\frac{0.10}{12}} \quad \text{substituting}$$

To find P , we need to calculate the right side and then divide by the $(1 + \frac{0.10}{12})^{300}$ from the left side.

We get $110,047.23005 \approx \$110,047.23$. This means that if Fabiola deposits \$110,047.23, she will receive monthly payouts of \$1,000 each for twenty-five years, or a total of $25 \cdot 12 \cdot 1,000 = \$300,000$.



If Fabiola's principal received no interest, then she would need $300 \cdot \$1,000 = \$300,000$. Since her principal does receive compound interest for a long time, she needs significantly less than \$300,000.

If Fabiola Macias in Example 3 were like most people, she would not have \$110,047.23 in savings when she retires, so she wouldn't be able to set up a payout annuity for herself. However, if she had set up a savings annuity thirty years before retirement, she could have saved that amount by making monthly payments of only \$48.68. This is something that almost anyone can afford, and it's a wonderful deal. Thirty years of monthly payments of \$48.68, while you are working, can generate twenty-five years of monthly payments of \$1,000 when you are retired. In the exercises, we will explore this combination of a savings annuity and a payout annuity.

Payout Annuities with Inflation

The only trouble with Fabiola's retirement payout annuity in Example 3 is that she is ignoring inflation. In twenty-five years, she will still be receiving \$1,000 a month, but her money won't buy as much as it does today. Fabiola would be better off if she allowed herself an annual **cost-of-living adjustment (COLA)**.

EXAMPLE 4

UNDERSTANDING WHAT A COLA IS After retiring, Fabiola Macias set up a payout annuity with her bank. For the next twenty-five years, she will receive payouts that start at \$1,000 per month and then receive an annual COLA of 3%. Find the size of her monthly payout for

- a. the first year
- b. the second year
- c. the third year
- d. the 25th year

SOLUTION

- a. During the first year, no adjustment is made, so she will receive \$1,000 per month.
 b. During the second year, her monthly payout of \$1,000 will increase 3%, so her new monthly payout will be

$$1,000 \cdot (1 + .03) = \$1,030$$

- c. During the third year, her monthly payout of \$1,030 will increase 3%, so her new monthly payout will be

$$1,030 \cdot (1 + .03) = \$1,060.90$$

Since the 1,030 in the above calculation came from computing $1,000 \cdot (1 + .03)$, we could rewrite this calculation as

$$1000 \cdot (1 + .03)^2 = \$1,060.90$$

- d. By the twenty-fifth year, she will have received twenty-four 3% increases, so her monthly payout will be

$$1,000 \cdot (1 + .03)^{24} = 2,032.7941 \dots \approx \$2,032.79$$

If a payout annuity is to have automatic annual cost-of-living adjustments, the following formula should be used to find the principal that must be deposited. The COLA is an annual one, so all other figures must also be annual figures; in particular, r is the *annual* interest rate, t is the duration of the annuity in *years*, and we use an *annual* payout.

ANNUAL PAYOUT ANNUITY WITH COLA FORMULA

A payout annuity of t years, where the payouts receive an annual COLA, requires a principal of

$$P = (pymt) \frac{1 - \left(\frac{1 + c}{1 + r} \right)^t}{r - c}$$

where $pymt$ is the annual payout for the first year, c is the annual COLA rate, and r is the annual rate at which interest is earned on the principal.

EXAMPLE 5

A PAYOUT ANNUITY WITH A COLA After retiring, Sam Needham set up a payout annuity with his bank. For the next twenty-five years, he will receive annual payouts that start at \$12,000 and then receive an annual COLA of 3%. Use the Annual Payout Annuity with COLA Formula to find how much money he must deposit if his money earns 10% interest per year.

SOLUTION

We are given that the annual payout is $pymt = 12,000$, $r = 10\% = 0.10$, $c = 3\% = 0.03$, and $t = 25$.

$$P = (pymt) \frac{1 - \left(\frac{1 + c}{1 + r} \right)^t}{r - c} \quad \text{the COLA Formula}$$

$$= (12,000) \frac{1 - \left(\frac{1 + 0.03}{1 + 0.10} \right)^{25}}{0.10 - 0.03} \quad \text{substituting}$$

$$= (12,000) \frac{1 - \left(\frac{1.03}{1.10} \right)^{25}}{0.10 - 0.03} \quad \text{simplifying}$$

$$= 138,300.4587 \approx \$138,300.46 \quad \text{rounding}$$



$$1 - (1.03 \div 1.10)^{25} = \div (.10 - .03) \times 12000 =$$

With a graphing calculator, press \wedge rather than y^x and **ENTER** rather than **=**.

This means that if Sam deposits \$138,300.46 now, he will receive

\$12,000 in one year

$$\$12,000 \cdot (1 + .03)^1 = \$12,360 \text{ in two years}$$

$$\$12,000 \cdot (1 + .03)^2 = \$12,730.80 \text{ in three years}$$

$$\$12,000 \cdot (1 + .03)^{24} = \$24,393.53 \text{ in twenty-five years}$$



If Sam's principal received no interest, he would need $25 \cdot \$12,000 = \$300,000$. Since his principal does receive compound interest for a long time, he needs significantly less than \$300,000.

Compare Sam's payout annuity in Example 5 with Fabiola's in Example 3. Sam receives annual payouts that start at \$12,000 and slowly increase to \$24,393. Fabiola receives exactly \$1,000 per month (or \$12,000 per year) for the same amount of time. Sam was required to deposit \$138,300, and Fabiola was required to deposit \$110,047.23.

5.6 EXERCISES

- ▶ 1. Cheryl Wilcox is planning for her retirement, so she is setting up a payout annuity with her bank. She wishes to receive a payout of \$1,200 per month for twenty years.
 - a. How much money must she deposit if her money earns 8% interest compounded monthly?
 - b. Find the total amount that Cheryl will receive from her payout annuity.
2. James Magee is planning for his retirement, so he is setting up a payout annuity with his bank. He wishes to receive a payout of \$1,100 per month for twenty-five years.
 - a. How much money must he deposit if his money earns 9% interest compounded monthly?
 - b. Find the total amount that James will receive from his payout annuity.
- ▶ 3. Dean Gooch is planning for his retirement, so he is setting up a payout annuity with his bank. He wishes to receive a payout of \$1,300 per month for twenty-five years.
 - a. How much money must he deposit if his money earns 7.3% interest compounded monthly?
 - b. Find the total amount that Dean will receive from his payout annuity.
- ▶ 4. Holly Krech is planning for her retirement, so she is setting up a payout annuity with her bank. She wishes to receive a payout of \$1,800 per month for twenty years.
 - a. How much money must she deposit if her money earns 7.8% interest compounded monthly?
 - b. Find the total amount that Holly will receive from her payout annuity.
- ▶ 5.
 - a. How large a monthly payment must Cheryl Wilcox (from Exercise 1) make if she saves for her payout annuity with an ordinary annuity, which she sets up thirty years before her retirement? (The two annuities pay the same interest rate.)
 - b. Find the total amount that Cheryl will pay into her ordinary annuity, and compare it with the total

▶ Selected exercises available online at www.webassign.net/brookscole

- amount that she will receive from her payout annuity.
6.
 - a. How large a monthly payment must James Magee (from Exercise 2) make if he saves for his payout annuity with an ordinary annuity, which he sets up twenty-five years before his retirement? (The two annuities pay the same interest rate.)
 - b. Find the total amount that James will pay into his ordinary annuity, and compare it with the total amount that he will receive from his payout annuity.
 - ▶ 7.
 - a. How large a monthly payment must Dean Gooch (from Exercise 3) make if he saves for his payout annuity with an ordinary annuity, which he sets up thirty years before his retirement? (The two annuities pay the same interest rate.)
 - b. How large a monthly payment must he make if he sets the ordinary annuity up twenty years before his retirement?
 - ▶ 8.
 - a. How large a monthly payment must Holly Krech (from Exercise 4) make if she saves for her payout annuity with an ordinary annuity, which she sets up thirty years before her retirement? (The two annuities pay the same interest rate.)
 - b. How large a monthly payment must she make if she sets the ordinary annuity up twenty years before her retirement?
 9. Lily Chang is planning for her retirement, so she is setting up a payout annuity with her bank. For twenty years, she wishes to receive annual payouts that start at \$14,000 and then receive an annual COLA of 4%.
 - a. How much money must she deposit if her money earns 8% interest per year?
 - b. How large will Lily's first annual payout be?
 - c. How large will Lily's second annual payout be?
 - d. How large will Lily's last annual payout be?
 - ▶ 10. Wally Brown is planning for his retirement, so he is setting up a payout annuity with his bank. For twenty-five years, he wishes to receive annual payouts that start at \$16,000 and then receive an annual COLA of 3.5%.
 - a. How much money must he deposit if his money earns 8.3% interest per year?
 - b. How large will Wally's first annual payout be?
 - c. How large will Wally's second annual payout be?
 - d. How large will Wally's last annual payout be?
 11. Oshri Karmon is planning for his retirement, so he is setting up a payout annuity with his bank. He is now 30 years old, and he will retire when he is 60. He wants to receive annual payouts for twenty-five years, and he wants those payouts to receive an annual COLA of 3.5%.
 - a. He wants his first payout to have the same purchasing power as does \$13,000 today. How big should that payout be if he assumes inflation of 3.5% per year?
 - b. How much money must he deposit when he is 60 if his money earns 7.2% interest per year?
 - c. How large a monthly payment must he make if he saves for his payout annuity with an ordinary annuity? (The two annuities pay the same interest rate.)
 - d. How large a monthly payment would he make if he waits until he is 40 before starting his ordinary annuity?
 - ▶ 12. Shelly Franks is planning for her retirement, so she is setting up a payout annuity with her bank. She is now 35 years old, and she will retire when she is 65. She wants to receive annual payouts for twenty years, and she wants those payouts to receive an annual COLA of 4%.
 - a. She wants her first payout to have the same purchasing power as does \$15,000 today. How big should that payout be if she assumes inflation of 4% per year?
 - b. How much money must she deposit when she is 65 if her money earns 8.3% interest per year?
 - c. How large a monthly payment must she make if she saves for her payout annuity with an ordinary annuity? (The two annuities pay the same interest rate.)
 - d. How large a monthly payment would she make if she waits until she is 40 before starting her ordinary annuity?

In Exercises 13–16, use the Annual Payout Annuity with COLA Formula to find the deposit necessary to receive monthly payouts with an annual cost-of-living adjustment. To use the formula, all figures must be annual figures, including the payout and the annual rate. You can adapt the formula for monthly payouts by using

- the future value of a one-year ordinary annuity in place of the annual payout, where $pymt$ is the monthly payout, and
 - the annual yield of the given compound interest rate in place of the annual rate r .
13. Fabiola Macias is about to retire, so she is setting up a payout annuity with her bank. She wishes to receive a monthly payout for the next twenty-five years, where the payout starts at \$1,000 per month and receives an annual COLA of 3%. Her money will earn 10% compounded monthly.
 - a. The annual payout is the future value of a one-year ordinary annuity. Find this future value.
 - b. The annual rate r is the annual yield of 10% interest compounded monthly. Find this annual yield. (*Do not round it off.*)
 - c. Use the Annual Payout Annuity with COLA Formula to find how much money she must deposit.
 - d. Fabiola could have saved for her payout annuity with an ordinary annuity. If she had started doing so thirty years ago, what would the required monthly payments have been? (The two annuities pay the same interest rate.)

- ▶ 14. Gary Kersting is about to retire, so he is setting up a payout annuity with his bank. He wishes to receive a monthly payout for the next twenty years, where the payout starts at \$1,300 per month and receives an annual COLA of 4%. His money will earn 8.7% compounded monthly.
- The annual payout is the future value of a one-year ordinary annuity. Find this future value.
 - The annual rate r is the annual yield of 8.7% interest compounded monthly. Find this annual yield. (*Do not* round it off.)
 - Use the Annual Payout Annuity with COLA Formula to find how much money he must deposit.
 - Gary could have saved for his payout annuity with an ordinary annuity. If he had started doing so twenty-five years ago, what would the required monthly payments have been? (The two annuities pay the same interest rate.)
15. Conrad von Shtup is about to retire, so he is setting up a payout annuity with his bank. He wishes to receive a monthly payout for the next twenty-three years, where the payout starts at \$1,400 per month and receives an annual COLA of 5%. His money will earn 8.9% compounded monthly.
- The annual payout is the future value of a one-year ordinary annuity. Find this future value.
 - The annual rate r is the annual yield of 8.9% interest compounded monthly. Find this annual yield. (*Do not* round it off.)
 - Use the Annual Payout Annuity with COLA Formula to find how much money he must deposit.
 - Conrad could have saved for his payout annuity with an ordinary annuity. If he had started doing so twenty years ago, what would the required monthly payments have been? (The two annuities pay the same interest rate.)
- ▶ 16. Mitch Martinez is about to retire, so he is setting up a payout annuity with his bank. He wishes to receive a monthly payout for the next thirty years, where the payout starts at \$1,250 per month and receives an annual COLA of 4%. His money will earn 7.8% compounded monthly.
- The annual payout is the future value of a one-year ordinary annuity. Find this future value.
 - The annual rate r is the annual yield of 7.8% interest compounded monthly. Find this annual yield.
 - Use the Annual Payout Annuity with COLA Formula to find how much money he must deposit.
 - Mitch could have saved for his payout annuity with an ordinary annuity. If he had started doing so twenty years ago, what would the required monthly payments have been? (The two annuities pay the same interest rate.)
17. Bob Pirtle won \$1 million in a state lottery. He was surprised to learn that he will not receive a check for

\$1 million. Rather, for twenty years, he will receive an annual check from the state for \$50,000. The state finances this series of checks by buying Bob a payout annuity. Find what the state pays for Bob's payout annuity if the interest rate is 8%.

- ▶ 18. John-Paul Ramin won \$2.3 million in a state lottery. He was surprised to learn that he will not receive a check for \$2.3 million. Rather, for twenty years, he will receive an annual check from the state for $\frac{1}{20}$ of his winnings. The state finances this series of checks by buying John-Paul a payout annuity. Find what the state pays for John-Paul's payout annuity if the interest rate is 7.2%.



Answer the following questions using complete sentences and your own words.

• CONCEPT QUESTIONS

19. Compare and contrast a savings annuity with a payout annuity. How do they differ in purpose? How do they differ in structure? How do their definitions differ?
HINT: Compare Examples 1 and 2.
20. Under what circumstances would a savings annuity and a payout annuity be combined?



WEB PROJECTS

21. This is an exercise in saving for your retirement. Write a paper describing all of the following points.
- Go on the web and find out what the annual rate of inflation has been for each of the last ten years. Use the average of these figures as a prediction of the future annual rate of inflation.
 - Estimate the total monthly expenses you would have if you were retired now. Include housing, food, and utilities.
 - Use parts (a) and (b) to predict your total monthly expenses when you retire, assuming that you retire at age 65.
 - Plan on financing your monthly expenses with a payout annuity. How much money must you deposit when you are 65 if your money earns 7.5% interest per year?
 - How large a monthly payment must you make if you save for your payout annuity with an ordinary annuity, starting now? (The two annuities pay the same interest rate.)
 - How large a monthly payment must you make if you save for your payout annuity with an ordinary annuity, starting ten years from now?
 - How large a monthly payment must you make if you save for your payout annuity with an ordinary annuity, starting twenty years from now?

Some useful links for this web project are listed on the text web site:

www.cengage.com/math/johnson

5

CHAPTER REVIEW



TERMS

add-on interest loan
adjustable rate mortgage
amortization schedule
amortized loan
annual percentage rate (APR) of an add-on interest loan
annual percentage rate (APR) of a simple interest amortized loan
annual yield
annuity

annuity due
average daily balance
balance
compound interest
compounding period
cost-of-living adjustment (COLA)
doubling time
expire
finance
finance charges
future value
Individual Retirement Account (IRA)
interest
interest rate

line of credit
loan agreement
lump sum
maturity value
negative amortization
nominal rate
note
ordinary annuity
outstanding principal
payment period
payout annuity
periodic rate
point
prepaying a loan
prepayment penalty

present value
present value of an annuity
principal
simple interest
simple interest amortized loan
simple interest loan
sinking fund
term
tax-deferred annuity (TDA)
Truth in Lending Act
unpaid balance
yield

FORMULAS

The **simple interest** I on a principal P at a interest rate r for t years is

$$I = Prt$$

and the **future value** is

$$FV = P(1 + rt)$$

After n compounding periods, the future value FV of an initial principal P earning **compound interest** at a periodic interest rate i for n periods is

$$FV = P(1 + i)^n$$

The **annual yield** of a given compound rate is the simple interest rate r that has the same future value as the compound rate in 1 year. To find it, solve

$$FV(\text{compound interest}) = FV(\text{simple interest})$$

$$P(1 + i)^n = P(1 + r)$$

for the simple interest rate r after making appropriate substitutions.

The future value FV of an **ordinary annuity** is

$$FV(\text{ordinary}) = pymt \frac{(1 + i)^n - 1}{i}$$

and the future value of an **annuity due** is

$$FV(\text{due}) = FV(\text{ordinary}) \cdot (1 + i)$$

$$= pymt \frac{(1 + i)^n - 1}{i} (1 + i)$$

The **present value of an ordinary annuity** is the lump sum P such that

$$FV(\text{lump sum}) = FV(\text{annuity})$$

$$P(1 + i)^n = pymt \frac{(1 + i)^n - 1}{i}$$

Simple interest amortized loan:

$$FV(\text{annuity}) = FV(\text{loan amount})$$

$$pymt \frac{(1 + i)^n - 1}{i} = P(1 + i)^n$$

Unpaid balance:

$$\text{Unpaid balance} = \text{current value of loan amount} - \text{current value of annuity}$$

$$\approx P(1 + i)^n - pymt \frac{(1 + i)^n - 1}{i}$$

where n is the number of periods *from the beginning of the loan to the present*.

Payout annuity: The total principal necessary to generate n payouts of size $pymt$ is P , where

$$P(1 + i)^n = pymt \frac{(1 + i)^n - 1}{i}$$

Annual payout annuity with COLA: A payout annuity with a term of t years, a first-year payout of $pymt$, and an annual COLA rate c , requires a principal of

$$P = (pymt) \frac{1 - \left(\frac{1 + c}{1 + r}\right)^t}{r - c}$$

STEPS

To find the **credit card finance charge** with the **average daily balance method**:

1. Find the balance for each day in the billing period and the number of days at that balance.

2. The average daily balance is the weighted average of these daily balances, weighted to reflect the number of days at that balance.
3. The finance charge is simple interest applied to the average daily balance.

To create an **amortization schedule**:

For each payment, list the payment number, principal portion, interest portion, total payment and balance.

For each payment:

1. Find the interest on the balance—use the simple interest formula.

For each payment except the last:

2. The principal portion is the payment minus the interest portion.
3. The new balance is the previous balance minus the principal portion.

For the last payment:

4. The principal portion is the previous balance.
5. The total payment is the sum of the principal portion and interest portion.

REVIEW EXERCISES

1. Find the interest earned by a deposit of \$8,140 at $9\frac{3}{4}\%$ simple interest for eleven years.
2. Find the interest earned by a deposit of \$10,620 at $8\frac{1}{2}\%$ simple interest for twenty-five years.
3. Find the future value of a deposit of \$12,288 at $4\frac{1}{4}\%$ simple interest for fifteen years.
4. Find the future value of a deposit of \$22,880 at $5\frac{3}{4}\%$ simple interest for thirty years.
5. Find the maturity value of a loan of \$3,550 borrowed at $12\frac{1}{2}\%$ simple interest for one year and two months.
6. Find the maturity value of a loan of \$12,250 borrowed at $5\frac{1}{2}\%$ simple interest for two years.
7. Find the present value of a future value of \$84,120 at $7\frac{1}{4}\%$ simple interest for twenty-five years.
8. Find the present value of a future value of \$10,250 at $5\frac{3}{4}\%$ simple interest for twenty years.
9. Find the future value of a deposit of \$8,140 at $9\frac{3}{4}\%$ interest compounded monthly for eleven years.
10. Find the future value of a deposit of \$7,250 at $5\frac{1}{4}\%$ interest compounded monthly for twenty years.
11. Find the interest earned by a deposit of \$7,990 at $4\frac{3}{4}\%$ interest compounded monthly for eleven years.
12. Find the interest earned by a deposit of \$22,250 at $9\frac{1}{4}\%$ interest compounded monthly for twenty years.
13. Find the present value of a future value of \$33,120 at $6\frac{1}{4}\%$ interest compounded daily for twenty-five years.
14. Find the present value of a future value of \$10,600 at $7\frac{7}{8}\%$ interest compounded daily for four years.
15. Find the annual yield corresponding to a nominal rate of 7% compounded daily.
16. Find the annual yield corresponding to a nominal rate of $8\frac{1}{2}\%$ compounded monthly.
17. Find the future value of a twenty-year ordinary annuity with monthly payments of \$230 at 6.25% interest.
18. Find the future value of a thirty-year ordinary annuity with biweekly payments of \$130 at 7.25% interest.
19. Find the future value of a twenty-year annuity due with monthly payments of \$450 at 8.25% interest.
20. Find the future value of a thirty-year annuity due with biweekly payments of \$240 at 6.75% interest.
21. Find (a) the monthly payment and (b) the total interest for a simple interest amortized loan of \$25,000 for five years at $9\frac{1}{2}\%$ interest.
22. Find (a) the monthly payment and (b) the total interest for a simple interest amortized loan of \$130,000 for twenty years at $8\frac{1}{4}\%$ interest.
23. The Square Wheel Bicycle store has found that they sell most of their bikes in the spring and early summer. On February 15, they borrowed \$351,500 to buy bicycles. They are confident that they can sell most of these bikes by September 1. Their loan is at $5\frac{7}{8}\%$ simple interest. What size lump sum payment would they have to make on September 1 to pay off the loan?

24. Mike Taylor buys a four-year old Ford from a car dealer for \$16,825. He puts 10% down and finances the rest through the dealer at 10.5% add-on interest. If he agrees to make thirty-six monthly payments, find the size of each payment.
25. The activity on Sue Washburn's MasterCard account for one billing period is shown below. Find the average daily balance and the finance charge if the billing period is August 26 through September 25, the previous balance was \$3,472.38, and the annual interest rate is $19\frac{1}{2}\%$.
- | | | |
|--------------|------------|----------|
| August 30 | payment | \$100.00 |
| September 2 | gasoline | \$34.12 |
| September 10 | restaurant | \$62.00 |
26. George and Martha Simpson bought a house from Sue Sanchez for \$205,500. In lieu of a 20% down payment, Ms. Sanchez accepted 5% down at the time of the sale and a promissory note from the Simpsons for the remaining 15%, due in eight years. The Simpsons also agreed to make monthly interest payments to Ms. Sanchez at 12% simple interest until the note expires. The Simpsons borrowed the remaining 80% of the purchase price from their bank. The bank paid that amount, less a commission of 6% of the purchase price, to Ms. Sanchez.
- Find the Simpsons' monthly interest-only payment to Ms. Sanchez.
 - Find Ms. Sanchez's total income from all aspects of the down payment.
 - Find Ms. Sanchez's total income from all aspects of the sale of the house, including the down payment.
27. Tien Ren Chiang wants to have an IRA that will be worth \$250,000 when he retires at age 65.
- How much must he deposit at age 25 at $8\frac{1}{8}\%$ compounded quarterly?
 - If he arranges for the monthly interest to be sent to him starting at age 65, how much would he receive each month? (Assume that he will continue to receive $8\frac{1}{8}\%$ interest, compounded monthly.)
28. Extremely Trustworthy Savings offers five-year CDs at 7.63% compounded annually, and Bank of the South offers five-year CDs at 7.59% compounded daily. Compute the annual yield for each institution and determine which offering is more advantageous for the consumer.
29. You are 32, and you have just set up an ordinary annuity to save for retirement. You make monthly payments of \$200 that earn $6\frac{1}{8}\%$ interest. Find the future value when you reach age 65.
30. Find and interpret the present value of the annuity in Exercise 29.
31. Find the future value of an annuity due with monthly payments of \$200 that earns $6\frac{1}{8}\%$ interest, after eleven years.
32. Matt and Leslie Silver want to set up a TDA that will generate sufficient interest on maturity to meet their living expenses, which they project to be \$1,300 per month.
- Find the amount needed at maturity to generate \$1,300 per month interest if they can get $8\frac{1}{4}\%$ interest compounded monthly.
 - Find the monthly payment they would have to make into an ordinary annuity to obtain the future value found in part (a) if their money earns $9\frac{3}{4}\%$ and the term is thirty years.
33. Mr. and Mrs. Liberatore set up a TDA to save for their retirement. They agreed to have \$100 deducted from each of Mrs. Liberatore's monthly paychecks, which will earn $6\frac{1}{8}\%$ interest.
- Find the future value of their ordinary annuity, if it comes to term after they retire in thirty years.
 - After retiring, the Liberatores convert their annuity to a savings account, which earns 5.75% interest compounded monthly. At the end of each month, they withdraw \$1,000 for living expenses. Complete the chart in Figure 5.29 for their postretirement account.

Month Number	Account Balance at Beginning of the Month	Interest for the Month	Withdrawal	Account Balance at End of the Month
1				
2				
3				
4				
5				

FIGURE 5.29 Chart for Exercise 33.

34. Delores Lopez buys some land in Nevada. She agrees to pay the seller a lump sum of \$235,000 in five years. Until then, she will make monthly simple interest payments to the seller at 10% interest.
- Find the amount of each interest payment.
 - Delores sets up a sinking fund to save the \$235,000. Find the size of her monthly payments if her payments are due at the end of every month and her money earns $9\frac{3}{8}\%$ interest.
 - Prepare a table showing the amount in the sinking fund after each of the first two deposits.
35. Maude Frickett bought a house for \$225,600. She put 20% down and obtains a simple interest amortized loan for the rest at $7\frac{3}{8}\%$ for thirty years.
- Find her monthly payment.
 - Find the total interest.
 - Prepare an amortization schedule for the first two months of the loan.

36. Navlet's Nursery needs to borrow \$228,000 to increase its inventory for the upcoming spring season. The owner is confident that he will sell most if not all of the new plants during the summer, so he wishes to borrow the money for only six months. His bank has offered him a simple interest amortized loan at $8\frac{1}{4}\%$ interest.
- Find the size of the monthly bank payment.
 - Prepare an amortization schedule for all six months of the loan.
37. Harry Carry had his kitchen remodeled. He did not have sufficient cash to pay for it. However, he had previously set up a line of credit with his bank. On May 16, he wrote a check to his contractor on his line of credit for \$41,519. The line's interest rate is $7\frac{3}{4}\%$.
- Find the size of the required monthly interest payment.
 - Harry decided that it would be in his best interests to get this loan paid off in seven months. Find the size of the monthly principal-plus-interest payment that would accomplish this.
(*HINT:* In effect, Carry is converting the loan to an amortized loan.)
 - Prepare an amortization schedule for all seven months of the loan.
 - Find the amount of line of credit interest that Carry could deduct from his taxes next year.
38. Ben Suico buys a car for \$13,487.31. He puts 10% down and obtains a simple interest amortized loan for the rest at $10\frac{7}{8}\%$ interest for five years.
- Find his monthly payment.
 - Find the total interest.
 - Prepare an amortization schedule for the first two months of the loan.
 - Mr. Suico decides to sell his car two years and six months after he bought it. Find the unpaid balance on his loan.
39. Scott Frei wants to buy a used car that costs \$6,200. The used car dealer has offered him a four-year add-on interest loan that requires a \$200 down payment at 9.9% annual interest with an APR of 10%.
- Find the monthly payment.
 - Verify the APR.
40. Miles Archer bought a house for \$112,660. He put 20% down and obtains a simple interest amortized loan for the rest at $9\frac{7}{8}\%$ for thirty years. If Miles paid two points and \$5,738.22 in fees, \$1,419.23 of which are included in the finance charge, find the APR.
41. Susan and Steven Tamchin are thinking of buying a home for \$198,000. A potential lender advertises an 80%, thirty-year simple interest amortized loan at $8\frac{1}{2}\%$ interest, with an APR of 9.02%.
- Find the size of the Tamchin's monthly payment.
 - Use the APR to approximate the fees included in the finance charge.
42. Fred Rodgers is planning for his retirement, so he is setting up a payout annuity with his bank. He wishes to receive a payout of \$1,700 per month for twenty-five years.
- How much money must he deposit if his money earns 6.1% interest compounded monthly?
 - How large a monthly payment would Fred have made if he had saved for his payout annuity with an ordinary annuity, set up thirty years before his retirement? (The two annuities pay the same interest rate.)
 - Find the total amount that Fred will pay into his ordinary annuity and the total amount that he will receive from his payout annuity.
43. Sue West is planning for her retirement, so she is setting up a payout annuity with her bank. She is now 30 years old, and she will retire when she is 60. She wants to receive annual payouts for twenty-five years, and she wants those payouts to have an annual COLA of 4.2%.
- She wants her first payout to have the same purchasing power as does \$17,000 today. How big should that payout be if she assumes inflation of 4.2% per year?
 - How much money must she deposit when she is sixty if her money earns 8.3% interest per year?
 - How large a monthly payment must she make if she saves for her payout annuity with an ordinary annuity? (The two annuities pay the same interest rate.)



Answer the following questions using complete sentences and your own words.

• CONCEPT QUESTIONS

- What is the difference between simple interest and compound interest?
- Describe a situation in which simple interest, rather than compound interest, would be expected.
- Describe a situation in which compound interest, rather than simple interest, would be expected.
- What is the difference between an account that earns compound interest and an annuity that earns compound interest?
- What is the difference between a simple interest amortized loan and an add-on interest loan?

• HISTORY QUESTIONS

- What does the Truth in Lending Act do for borrowers?
- Who offered the first credit card?
- How did the first credit card differ from the first post-World War II credit card?

VOTING AND APPORTIONMENT

6



Comstock/Jupiter Images

The right to vote for one's governing officials is the cornerstone of any democracy. In the United States, individual citizens have a voice in the selection of local mayors and city councils, county officials, state legislators, governors, members of the House of Representatives and the Senate, and even the ultimate office: the President of the United States. Elections of all types are commonplace, and their outcomes range from overwhelming landslides to razor-thin victories, legal suits, and tedious recounts.

When we vote to fill a political office, the outcome obviously depends on the number of votes cast for each candidate; when we vote to create laws, the numbers for and against a proposition determine its fate. However, determining the outcome of an election might not be as simple as some may think. The premises of "one

continued



WHAT WE WILL DO IN THIS CHAPTER

In this chapter, we will study systems of voting and methods of apportionment.

WE'LL DEVELOP AND EXPLORE VARIOUS SYSTEMS OF VOTING:

- Different systems of voting may produce different outcomes; one system may conclude that candidate A is the winner, whereas another system may proclaim that B wins. In any election, large or small, voters should understand the specific system of voting that is being used *before* they cast their votes!
- Ideally, any system of voting should produce fair results. What exactly is meant by the word *fair*? Is any system of voting ultimately fair in all situations?

WE'LL DEVELOP AND EXPLORE VARIOUS SYSTEMS OF APPORTIONMENT:

- Different methods of apportionment can lead to significantly different allocations of seats in a governing body. Some methods favor regions that have small populations, whereas others favor those with large. Who should decide which method will be used?

continued

407

WHAT WE WILL DO IN THIS CHAPTER — *continued*

- Ideally, any method of apportionment should produce fair results. However, in the pursuit of fairness, all methods of apportionment create paradoxes that undermine the fairness they seek! Which method is “best” to use?
- When the United States first became an independent nation, several of the “founding fathers” proposed various methods of apportionment for the House of Representatives. Which methods have actually been used? Which method is used today?

person, one vote” and “the candidate with the most votes wins” are neither universal nor absolute. In fact, several different voting systems have been used in democratic societies today and throughout history.

Elected governmental bodies range in scope from local to regional to national levels. That is, elected officials represent their constituencies in matters that affect others in a broader sense. For example, precincts within a city elect officials to represent the precinct on citywide matters, districts within a state elect officials to represent the district on statewide matters, and states elect officials to represent the individual states on nationwide matters. How many officials will a region receive to represent it in these matters? In some cases, each region may receive the same number of representative seats; in others, different regions may receive different numbers of seats. The determination of the number of representatives a region receives is known as apportionment.

6.1 Voting Systems

OBJECTIVES

- Develop and apply the plurality method of voting
- Develop and apply the plurality with elimination method of voting
- Develop and apply the instant runoff method of voting
- Develop and apply the Borda count method of voting
- Develop and apply the pairwise comparison method of voting
- Define and investigate the four fairness criteria for voting systems

Voting is an essential element in any democratic form of government. Whether it is the selection of leaders, the creation of laws and regulations, or deciding the outcome of issues ranging from mundane to volatile, decisions are made on the basis of the results of elections.



Matt Struchiner/Bloomberg via Getty Images

Voting is an essential element in any democratic form of government.

Once votes have been cast, they must be counted, and a winner must be declared. Many people assume that “the candidate with the most votes wins”; however, that is not always the case. Deciding who wins an election is ultimately determined by the type of voting system that is used. “The person with the most votes wins” is a commonly used system and is called the plurality method. However, the plurality method is not the only method that is used. In this section, we will examine several common voting systems: (1) the plurality method, (2) the plurality with elimination method, (3) the ranked-choice or instant run-off method (sometimes called the Australian method), (4) the Borda count method, and (5) the pairwise comparison method.

The Plurality Method

A common way to determine the outcome of an election is to declare the candidate with the most votes the winner. This is the basis of the **plurality method of voting**.

PLURALITY METHOD OF VOTING

Each person votes for his or her favorite candidate (or choice). The candidate (or choice) that receives the most votes is declared the winner. (In case of a tie, a special runoff election might be held.)

EXAMPLE 1

APPLYING THE PLURALITY METHOD OF VOTING: MAJORITY SCENARIO City Cab is a new taxi service that will soon begin operation in a major metropolitan area. The board of directors for City Cab must purchase a fleet of new vehicles, and the vehicles must all be the same color. Four colors are available: black, white, red, and green. The five directors vote for their choice of color and the results are given in Figure 6.1. Using the plurality method of voting, which color choice wins?

	Director 1	Director 2	Director 3	Director 4	Director 5
Choice	green	red	red	green	red

FIGURE 6.1 Results of voting for the color of City Cab vehicles.

SOLUTION

Tallying the results, we see that green received two votes and red received three. Because red received the most votes (3), red is declared the winner.

In Example 1, it should be noted that in addition to receiving the most votes, red also received a majority of the votes, that is, more than 50% of the votes were for red. As we can see, three out of five, or 60%, of the votes were for red, while 40% were not for red.

Although the plurality method is easy to apply, this method can produce an unusual winner, as we shall see in the next example.

EXAMPLE 2

APPLYING THE PLURALITY METHOD OF VOTING: NONMAJORITY SCENARIO Referring to City Cab in Example 1, suppose that the results of the election are those given in Figure 6.2. Using the plurality method of voting, which color choice wins?

	Director 1	Director 2	Director 3	Director 4	Director 5
Choice	green	red	red	black	white

FIGURE 6.2 Results of voting for the color of City Cab vehicles.

SOLUTION

Once again, because red received the most votes (two), red would be declared the winner. Although red is the winner, it should be noted that red did *not* receive a *majority* of the votes. Fewer than half of the votes, that is 40%, were for red, while significantly more than half (60%) of the directors did not want the color red. This type of “winner” can lead to very difficult situations in political elections. Imagine yourself as the “winner” of a local election when more than half of the voters do *not* support you.

The Plurality with Elimination Method

As was shown in Example 2, the plurality method of voting may produce a “winner” even though a majority of voters did not vote for the “winner.” This dilemma can be avoided in several ways. One common way is to eliminate the candidate with the fewest votes and then hold another election. This is the basis of the **plurality with elimination method**. Consequently, the voters who supported the eliminated candidate now vote for someone else; that is, they must vote for their second choice. Typically, this second election will produce a winner with a majority of support. However, if a majority is still not attained, the process is repeated until a candidate obtains a majority of the votes.

PLURALITY WITH ELIMINATION METHOD OF VOTING

Each person votes for his or her favorite candidate. If a candidate receives a majority of votes, that candidate is declared the winner. If no candidate receives a majority, then the candidate with the fewest votes is eliminated and a new election is held. This process continues until a candidate receives a majority of the votes.

EXAMPLE 3

APPLYING THE PLURALITY WITH ELIMINATION METHOD OF VOTING The managers of We Deliver (a local shipping business) are planning a party for their 60 employees. There are three possible locations for the party: the warehouse, the park, and the beach. The employees voted on which location they preferred, and the results are given in Figure 6.3.

	Warehouse	Park	Beach
Number of votes	23	19	18

FIGURE 6.3 Results of voting for party location.

- Using the plurality method of voting, which location wins?
- Using the plurality with elimination method of voting, which location wins?

SOLUTION

- Because the warehouse received the most votes (23), the party will be held at the warehouse.
- Although the warehouse received the most votes, it did not receive a majority; the warehouse received only $23/60$, or approximately 38.3%, of the vote. Because the beach received the fewest number of votes (18), it is eliminated, and a new election is held. The results of the second election are given in Figure 6.4.

	Warehouse	Park	Beach
Number of votes	25	35	0

FIGURE 6.4 Results of the second election.

Because the park received a majority of the votes ($35/60 \approx 58.3\%$), the party will be held at the park.

As is shown in Example 3, the outcome of an election may change depending on the type of voting system used. Therefore, it is imperative that the voters know ahead of time which system is being used!

As we have seen, the elimination of one candidate forces some people to shift their vote to their second choice in the new election. In the realm of real-world political elections, this method of elimination can be very expensive both for the local office of elections and for the candidates themselves, and it is overly time-consuming. Rather than holding a new election after eliminating the candidate with the fewest votes, a popular alternative is to have each voter rank each candidate during the first election. This is the basis of the ranked-choice or instant runoff method.

The Ranked-Choice or Instant Runoff Method

In Example 3, the employees of We Deliver had three choices for the location of their party. If each of the 60 employees was asked to rank these choices in order of preference, in how many different ways could the locations be ranked? Because the choices are being put in an order (that is, a person must make a first choice, a second choice, and a third choice), we conclude that *permutations* can be used to determine the number of different rankings (see Section 2.4). Specifically, there are

$${}_3P_3 = \frac{3!}{(3 - 3)!} = \frac{3!}{0!} = \frac{3 \cdot 2 \cdot 1}{1} = 6$$

different rankings (permutations) of the three locations. One such ranking is “warehouse, park, beach” or, more conveniently, WPB. Another such ranking is BPW (beach, park, warehouse). Figure 6.5 lists all possible permutations (or rankings) of the three locations.

1st choice	W	W	P	P	B	B
2nd choice	P	B	B	W	W	P
3rd choice	B	P	W	B	P	W

FIGURE 6.5 All permutations (or rankings) of “Warehouse, Park, Beach.”

After each employee ranked the locations and the ballots were tallied, it was determined that nine employees chose the WPB ranking, fourteen chose WBP, fifteen chose PBW, four chose PWB, two chose BWP, and sixteen chose BPW. These results are portrayed in the table shown in Figure 6.6.

		Number of Ballots Cast					
		9	14	15	4	2	16
Ranked Ballot	1st choice	W	W	P	P	B	B
	2nd choice	P	B	B	W	W	P
	3rd choice	B	P	W	B	P	W

FIGURE 6.6 Voter preference table.

Tables like that shown in Figure 6.6 are often called **voter preference tables**, that is, tables that list different rankings of the candidates along with the number of voters who chose each specific ranking. Tables of this nature are often used to depict the voting patterns in a ranked election.

Referring to the voter preference table in Figure 6.6, notice that the sum of the row of numbers is 60, that is, the sum of the row equals the number of people voting. In addition, to determine the number of first-choice votes for each candidate, go to the row labeled “1st choice” and observe that $9 + 14 = 23$ people had W (the warehouse) as their first choice, whereas $15 + 4 = 19$ people had P (the park) as their first choice, and $2 + 16 = 18$ had B (the beach) as their first choice. These subtotals agree with the values given in Example 3.

We now proceed as we did when using the plurality with elimination method; that is, we eliminate the candidate with the fewest first-choice votes (the beach). However, rather than holding a new election, we simply modify the original table by shifting the votes in the voter preference table upwards to fill the voids created by eliminating the beach, as shown in Figure 6.7.

		Number of Ballots Cast					
		9	14	15	4	2	16
Ranked Ballot	1st choice	W	W	P	P	B	B
	2nd choice	P	B	B	W	W [↑]	P [↑]
	3rd choice	B	P [↑]	W [↑]	B	P [↑]	W [↑]

FIGURE 6.7 Shifting votes in the voter preference table.

The modified voter preference table is shown in Figure 6.8.

		Number of Ballots Cast					
		9	14	15	4	2	16
1st choice	W	W	P	P	W	P	
2nd choice	P	P	W	W	P	W	

FIGURE 6.8 Modified voter preference table.

Referring to Figure 6.8, we now see that the Warehouse received $9 + 14 + 2 = 25$ first choice votes, whereas the Park received $15 + 4 + 16 = 35$ votes. Of course, these values agree with those given in Example 3.

The process described above is called the **ranked-choice or instant runoff method**. This voting system is commonly used in Australia. In March 2002, the voters of San Francisco, California, adopted the instant runoff method for the election of most political posts, including mayor, sheriff, district attorney, and many other high-visibility positions. It has been estimated that this voting method will save the City of San Francisco millions of dollars per election, and voters will not have to wait several weeks for a second election, as evidenced by the article on page 414 published in 2005.

The instant runoff method of voting is summarized as follows.

INSTANT RUNOFF METHOD OF VOTING

Each voter ranks all of the candidates; that is, each voter selects his or her first choice, second choice, third choice, and so on. If a candidate receives a majority of first-choice votes, that candidate is declared the winner. If no candidate receives a majority, then the candidate with the fewest first-choice votes is eliminated, and those votes are given to the next preferred candidate. If a candidate now has a majority of first-choice votes, that candidate is declared the winner. If no candidate receives a majority, this process continues until a candidate receives a majority.

FEATURED IN THE NEWS

VOTERS RANK ALL CHOICES SUCCESS FOR INSTANT RUNOFF VOTING IN SAN FRANCISCO

Last November, San Francisco proved to be a beacon in an otherwise tumultuous election season. In a time of polarized national politics and an alienated electorate, San Francisco embarked on an important innovation that points American democracy toward the future.

San Francisco elected seven seats on the city council (called the Board of Supervisors) using a method known as instant runoff voting (IRV). Several races were hotly contested, one race drawing a remarkable 22 candidates. Observers long used to the blood sport of San Francisco politics were amazed to see how candidates in several races engaged in more coalition building and less vicious negative attacks. Winners were all decided either on election night or within 72 hours after the polls had closed, and even skeptics were won

over. Two exit polls showed that city voters generally liked IRV and found it easy to use, including voters across racial and ethnic lines. National media including the New York Times, Washington Post, Associated Press and National Public Radio covered the successful election.

San Francisco will use IRV in future years for citywide offices like mayor and district attorney, joining the ranks of Ireland, Australia and London that use IRV to elect their highest offices. IRV simulates a series of runoff elections but finishes the job in a single election. Voters rank candidates for each race in order of choice: first, second, third. If your first choice gets eliminated from the “instant runoff,” your vote goes to your second-ranked candidate as your backup choice. The runoff rankings are used to determine which candidate has support from a popular majority, and accomplish this in a single election. Voters are liberated to vote for the

candidates they really like, no more spoiler candidates and “lesser of two evils” dilemmas.

Previously San Francisco decided majority winners in a December runoff election. Runoffs were expensive, costing the City more than \$3 million city-wide, and voter turnout often plummeted in the December election by as much as 50 percent. So San Francisco taxpayers will save millions of dollars by using IRV, and winners now are determined in the November election when voter turnout tends to be highest. Also, candidates didn’t need to raise more money for a second election and independent expenditures declined, significantly improving the campaign finance situation.

by Steven Hill and Rob Richie
Published on Wednesday, January
12, 2005 by CommonDreams.org

EXAMPLE 4

APPLYING AND COMPARING THE PLURALITY AND INSTANT RUNOFF METHODS OF VOTING Townsville is electing a new mayor. The candidates are Alturas (A), Bellum (B), Chan (C), and Dushay (D). The ranked ballots are tallied, and the results are summarized as shown in the voter preference table given in Figure 6.9.

		Number of Ballots Cast					
		15,527	15,287	10,023	9,105	7,978	5,660
Ranked Ballot	1st choice	C	B	C	B	A	D
	2nd choice	B	C	D	A	B	A
	3rd choice	A	D	A	D	C	C
	4th choice	D	A	B	C	D	B

FIGURE 6.9 Voter preference table for Mayor of Townsville.

- How many voters participated in the election?
- Use the plurality method to determine the winner.
- Use the instant runoff method to determine the winner.

SOLUTION

- a. To determine the number of voters, we find the sum of the row of numbers representing the number of ballots cast for each ranking:

$$15,527 + 15,287 + 10,023 + 9,105 + 7,978 + 5,660 = 63,580$$

Therefore, 63,580 voters participated in the election.

- b. Looking at the row of first-choice votes, we see that

A received 7,978 votes,

B received $15,287 + 9,105 = 24,392$ votes,

C received $15,527 + 10,023 = 25,550$ votes, and

D received 5,660 votes.

Because Chan (C) received the most votes (25,550), Mr. Chan becomes the new mayor of Townsville. However, it should be noted that Mr. Chan did not receive a majority of the votes; he received only $25,550/63,580 \approx 40.2\%$.

- c. First, we determine the number of votes needed to obtain a majority. Because 63,580 votes were cast, a candidate needs more than half of 63,580, that is, more than $63,580 \div 2 = 31,790$ votes; in other words, a candidate needs 31,791 votes to win. Because D received the fewest number of votes (5,660), Dushay is eliminated, and we obtain the modified voter preference table shown in Figure 6.10.

	Number of Ballots Cast					
	15,527	15,287	10,023	9,105	7,978	5,660
1st choice	C	B	C	B	A	A
2nd choice	B	C	A	A	B	C
3rd choice	A	A	B	C	C	B

FIGURE 6.10 Modified voter preference table for Mayor of Townsville.

Examining the new row of first-choice votes, we see that

A received $7,978 + 5,660 = 13,638$ votes,

B received $15,287 + 9,105 = 24,392$ votes, and

C received $15,527 + 10,023 = 25,550$ votes.

Because no candidate has a majority, we eliminate the candidate with the fewest votes, that is, we eliminate Alturas and obtain the modified voter preference table shown in Figure 6.11.

	Number of Ballots Cast					
	15,527	15,287	10,023	9,105	7,978	5,660
1st choice	C	B	C	B	B	C
2nd choice	B	C	B	C	C	B

FIGURE 6.11 Modified voter preference table for Mayor of Townsville.

Looking at the new row of first-choice votes, we see that

B received $15,287 + 9,105 + 7,978 = 32,370$ votes, and

C received $15,527 + 10,023 + 5,660 = 31,210$ votes.

Consequently, Ms. Bellum becomes the new mayor with $32,370/63,580 \approx 50.9\%$ of the vote.

Whereas the plurality method utilizes only a voter’s first choice, the instant runoff method is designed to accommodate a voter’s alternative choices. Another popular method that requires voters to rank their choices is the Borda count method.

The Borda Count Method

To win an election using the instant runoff method, a candidate must capture a majority of first-choice votes. Rather than tallying the first-choice votes only, we might want to tally the first-choice votes, the second-choice votes, the third-choice votes, and so on. This is the basis of the **Borda count method**, which is summarized in the following box.

BORDA COUNT METHOD OF VOTING

Each voter ranks all of the candidates; that is, each voter selects his or her first choice, second choice, third choice, and so on. If there are k candidates, each candidate receives k points for each first-choice vote, $(k - 1)$ points for each second-choice vote, $(k - 2)$ points for each third-choice vote, and so on. The candidate with the most total points is declared the winner.

EXAMPLE 5

APPLYING THE BORDA COUNT METHOD OF VOTING Use the Borda count method to determine the location of the party for the We Deliver employees referenced in Example 3.

SOLUTION

Recall that the employees were asked to rank the three locations (warehouse, park, and beach); the results are summarized in the voter preference table given in Figure 6.12.

		Number of Ballots Cast					
		9	14	15	4	2	16
Ranked Ballot	1st choice	W	W	P	P	B	B
	2nd choice	P	B	B	W	W	P
	3rd choice	B	P	W	B	P	W

FIGURE 6.12 Voter preference table.

First, we tally the votes for each location as shown in Figure 6.13. Because there were $k = 3$ candidates (warehouse, park, and beach), each first-choice vote is worth 3 points, each second-choice vote is worth 2 points, and each third-choice vote is worth 1 point.

	Warehouse	Park	Beach
1st choice	$9 + 14 = 23$ votes	$15 + 4 = 19$ votes	$2 + 16 = 18$ votes
2nd choice	$4 + 2 = 6$ votes	$9 + 16 = 25$ votes	$14 + 15 = 29$ votes
3rd choice	$15 + 16 = 31$ votes	$14 + 2 = 16$ votes	$9 + 4 = 13$ votes

FIGURE 6.13 Tally of votes.

Now determine the score for each candidate by multiplying the number of votes times the appropriate number of points as shown in Figure 6.14.

	Warehouse	Park	Beach
1st choice votes (3 points each)	23×3 points = 69 points	19×3 points = 57 points	18×3 points = 54 points
2nd choice votes (2 points each)	6×2 points = 12 points	25×2 points = 50 points	29×2 points = 58 points
3rd choice votes (1 point each)	31×1 points = 31 points	16×1 points = 16 points	13×1 points = 13 points

FIGURE 6.14 Points.

Summing each column, we obtain the results of the election:

$$\text{Warehouse} = 69 + 12 + 31 = 112 \text{ points}$$

$$\text{Park} = 57 + 50 + 16 = 123 \text{ points}$$

$$\text{Beach} = 54 + 58 + 13 = 125 \text{ points}$$

When the Borda count method is used, the beach is declared the winner because it has the most points (125).

As is illustrated by the examples referring to the We Deliver employees' party location, the outcome of the election depends on the voting system that is used. Specifically, the plurality method produced the warehouse as the winner, whereas the plurality with elimination and instant runoff methods declared the park to be the winner, and the Borda count method chose the beach. Obviously, it is very important to inform voters *before* the election as to which system will be used!

EXAMPLE 6

APPLYING THE BORDA COUNT METHOD OF VOTING Referring to Example 4, use the Borda count method to determine the next Mayor of Townsville.

SOLUTION

Referring to the voter preference table given in Figure 6.9, we tally the votes as shown in Figure 6.15.

	Alturas	Bellum	Chan	Dushay
1st choice	7,978 votes	$15,287 + 9,105$ = 24,392 votes	$15,527 + 10,023$ = 25,550 votes	5,660 votes
2nd choice	$9,105 + 5,660$ = 14,765 votes	$15,527 + 7,978$ = 23,505 votes	15,287 votes	10,023 votes
3rd choice	$15,527 + 10,023$ = 25,550 votes	0 votes	$7,978 + 5,660$ = 13,638 votes	$15,287 + 9,105$ = 24,392 votes
4th choice	15,287 votes	$10,023 + 5,660$ = 15,683 votes	9,105 votes	$15,527 + 7,978$ = 23,505 votes

FIGURE 6.15 Tally of votes.

Because there were $k = 4$ candidates (Alturas, Bellum, Chan, and Dushay), each first-choice vote is worth 4 points. Consequently, the second-, third-, and fourth-choice votes are worth 3, 2, and 1 points each, respectively. Now determine the score for each candidate by multiplying the number of votes times the appropriate number of points as shown in Figure 6.16.

	Alturas	Bellum	Chan	Dushay
1st choice (4 points each)	$7,978 \times 4$ = 31,912 points	$24,392 \times 4$ = 97,568 points	$25,550 \times 4$ = 102,200 points	$5,660 \times 4$ = 22,640 points
2nd choice (3 points each)	$14,765 \times 3$ = 44,295 points	$23,505 \times 3$ = 70,515 points	$15,287 \times 3$ = 45,861 points	$10,023 \times 3$ = 30,069 points
3rd choice (2 points each)	$25,550 \times 2$ = 51,100 votes	0×2 = 0 points	$13,638 \times 2$ = 27,276 points	$24,392 \times 2$ = 48,784 points
4th choice (1 point each)	$15,287 \times 1$ = 15,287 votes	$15,683 \times 1$ = 15,683 points	$9,105 \times 1$ = 9,105 points	$23,505 \times 1$ = 23,505 points

FIGURE 6.16 Points.

Summing each column, we obtain the results of the election:

$$\text{Alturas} = 31,912 + 44,295 + 51,100 + 15,287 = 142,594 \text{ points}$$

$$\text{Bellum} = 97,568 + 70,515 + 15,683 = 183,766 \text{ points}$$

$$\text{Chan} = 102,200 + 45,861 + 27,276 + 9,105 = 184,442 \text{ points}$$

$$\text{Dushay} = 22,640 + 30,069 + 48,784 + 23,505 = 124,998 \text{ points}$$

When the Borda count method is used, Mr. Chan becomes the new mayor because he has the most points (184,442).

The Pairwise Comparison Method

An election that has more than two candidates can be viewed as several mini-elections in which each possible pair of candidates compete against each other. That is, if there are three candidates, X, Y, and Z, we may consider the mini-elections or pairwise comparisons of X versus Y, X versus Z, and Y versus Z. We then determine the winner of each of these pairwise comparisons, and the candidate who wins the most of these is declared the winner of the overall election. This process is the basis of the **pairwise comparison method**, which is summarized in the following box.

PAIRWISE COMPARISON METHOD OF VOTING

Each voter ranks all of the candidates; that is, each voter selects his or her first choice, second choice, third choice, and so on. For each possible pairing of candidates, the candidate with the most votes receives 1 point; if there is a tie, each candidate receives $1/2$ point. The candidate who receives the most points is declared the winner.

How many possible pairwise comparisons are there? That is, if there are k candidates and we must select two for a mini-election, how many different

mini-elections are possible? The answer lies in using *combinations*, as illustrated in Section 2.4; that is, the order in which we select two candidates to compete does not matter (X versus Y is the same as Y versus X). Consequently, if there are k candidates, we must examine

$${}_k C_2 = \frac{k!}{(k-2)!2!} = \frac{k(k-1)}{2}$$

possible pairwise comparisons.

EXAMPLE 7

APPLYING THE PAIRWISE COMPARISON METHOD OF VOTING Use the pairwise comparison method to determine the location of the party for the We Deliver employees referenced in Example 3.

SOLUTION

Recall that the employees were asked to rank the three locations (warehouse, park, and beach) and the results are summarized in the voter preference table in Figure 6.17.

		Number of Ballots Cast					
		9	14	15	4	2	16
Ranked Ballot	1st choice	W	W	P	P	B	B
	2nd choice	P	B	B	W	W	P
	3rd choice	B	P	W	B	P	W

FIGURE 6.17 Voter preference table.

Because there are $k = 3$ “candidates,” we must examine

$${}_3 C_2 = \frac{3!}{(3-2)!2!} = \frac{3!}{1!2!} = \frac{3 \cdot 2 \cdot 1}{1 \cdot 2 \cdot 1} = 3$$

pairwise comparisons; specifically, we investigate W versus P, W versus B, and P versus B. For each comparison, we examine each column of the voter preference table to determine which candidate is preferred.

W versus P In the first column of Figure 6.17, we see that W is preferred over P; therefore, W receives the nine votes listed in column 1. In the second column, W is again ranked over P and receives fourteen votes. In the third column, P is preferred over W, so P receives fifteen votes. In the fourth column, P is again preferred over W and receives four votes. In the fifth column, W is preferred over P and receives two votes. Finally, P is preferred over W in the sixth column and receives sixteen votes. Tallying these results, we obtain the following totals:

$$\text{votes for W} = 9 + 14 + 2 = 25$$

$$\text{votes for P} = 15 + 4 + 16 = 35$$

For the comparison of W versus P, voters preferred P over W by a vote of 35 to 25. Consequently, P receives 1 point.

W versus B Examining Figure 6.17, we see that the voters in the first, second, and fourth columns all preferred W over B. Consequently, W receives the nine, fourteen, and four votes listed in those columns. We also see that the voters in the third, fifth, and sixth columns all preferred B over W, and so B receives the fifteen,

two, and sixteen votes listed in those columns. Tallying these results, we obtain the following totals:

$$\begin{aligned} \text{votes for W} &= 9 + 14 + 4 = 27 \\ \text{votes for B} &= 15 + 2 + 16 = 33 \end{aligned}$$

For the comparison of W versus B, voters preferred B over W by a vote of 33 to 27. Consequently, B receives 1 point.

P versus B Figure 6.17 indicates that the voters in the first, third, and fourth columns all ranked P over B. Therefore, P receives the nine, fifteen, and four votes listed in those columns. The voters in the remaining columns (second, fifth, and sixth) preferred B over P, and B receives fourteen, two, and sixteen votes accordingly. Tallying these results, we obtain the following totals:

$$\begin{aligned} \text{votes for P} &= 9 + 15 + 4 = 28 \\ \text{votes for B} &= 14 + 2 + 16 = 32 \end{aligned}$$

For the comparison of P versus B, voters preferred B over P by a vote of 32 to 28. Consequently, B receives 1 point.

Tallying the points, we see that B receives 2 points (B was preferred in two of the comparisons), P receives 1 point (P was preferred in one comparison), and W receives 0 points (W was not preferred in any of the comparisons). Consequently, B is declared the winner; the party will be held at the beach.

EXAMPLE 8

APPLYING THE PAIRWISE COMPARISON METHOD OF VOTING

Use the pairwise comparison method to determine the new mayor of Townsville referenced in Example 4.

SOLUTION

Recall that the candidates are Alturas (A), Bellum (B), Chan (C), and Dushay (D) and that 63,580 votes were cast. The ranked ballots are tallied, and the results are summarized as shown in the voter preference table in Figure 6.18.

		Number of Ballots Cast					
		15,527	15,287	10,023	9,105	7,978	5,660
Ranked Ballot	1st choice	C	B	C	B	A	D
	2nd choice	B	C	D	A	B	A
	3rd choice	A	D	A	D	C	C
	4th choice	D	A	B	C	D	B

FIGURE 6.18 Voter preference table for Mayor of Townsville.

Because there are $k = 4$ candidates, we must examine

$${}^4C_2 = \frac{4!}{(4-2)!2!} = \frac{4!}{2! \cdot 2!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1 \cdot 2 \cdot 1} = 6$$

pairwise comparisons; specifically, we investigate A versus B, A versus C, A versus D, B versus C, B versus D, and C versus D. For each comparison, we examine each column of the voter preference table to determine which candidate is preferred.

A versus B Figure 6.18 indicates that the voters in the third, fifth, and sixth columns all chose A over B. Therefore, A receives the $10,023 + 7,978 + 5,660 = 23,661$ votes listed in those columns. To determine the number of voters who prefer B over A, we could sum the votes listed in the remaining columns, or using a simpler approach, we subtract the 23,661 votes for A from the total number of votes to obtain the number of votes for B. Therefore, B receives $63,580 - 23,661 = 39,919$ votes. Consequently, B is preferred and receives 1 point.

A versus C Figure 6.18 indicates that the voters in the fourth, fifth, and sixth columns all preferred A over C. Therefore, the votes are tallied as follows:

$$A = 9,105 + 7,978 + 5,660 = 22,743 \text{ votes}$$

$$C = 63,580 - 22,743 = 40,837 \text{ votes}$$

Consequently, C is preferred and receives 1 point.

A versus D Figure 6.18 indicates that the voters in the first, fourth, and fifth columns all preferred A over D. Therefore, the votes are tallied as follows:

$$A = 15,527 + 9,105 + 7,978 = 32,610 \text{ votes}$$

$$D = 63,580 - 32,610 = 30,970 \text{ votes}$$

Consequently, A is preferred and receives 1 point.

B versus C Figure 6.18 indicates that the voters in the second, fourth, and fifth columns all preferred B over C. Therefore, the votes are tallied as follows:

$$B = 15,287 + 9,105 + 7,978 = 32,370 \text{ votes}$$

$$C = 63,580 - 32,370 = 31,210 \text{ votes}$$

Consequently, B is preferred and receives 1 point.

B versus D Figure 6.18 indicates that the voters in the first, second, fourth, and fifth columns all preferred B over D. Therefore, the votes are tallied as follows:

$$B = 15,527 + 15,287 + 9,105 + 7,978 = 47,897 \text{ votes}$$

$$D = 63,580 - 47,897 = 15,683 \text{ votes}$$

Consequently, B is preferred and receives 1 point.

C versus D Figure 6.18 indicates that the voters in the first, second, third, and fifth columns all preferred C over D. Therefore, the votes are tallied as follows:

$$C = 15,527 + 15,287 + 10,023 + 7,978 = 48,815 \text{ votes}$$

$$D = 63,580 - 48,815 = 14,765 \text{ votes}$$

Consequently, C is preferred and receives 1 point.

Tallying the points, we see that A receives 1 point, B receives 3 points, C receives 2 points, and D receives 0 points. B has the most points (3) and so is declared the winner; Ms. Bellum is the new mayor of Townsville.

Flaws of Voting Systems

As is illustrated by the examples involving the We Deliver employees' party location and the election of the mayor of Townsville, the outcome of an election may depend on the voting system that is used. Some people (especially the losing candidates) might claim that this is not fair. But what exactly is meant by the word

fair? Political scientists and mathematicians have created a list of criteria that any “fair” voting system should meet. Specifically, the four fairness criteria are known as the *majority criterion*, the *head-to-head criterion*, the *monotonicity criterion*, and the *irrelevant alternatives criterion*.

If you were a candidate in an election and you received a **majority** of the votes (that is, you received more than half of the votes), you would expect to be declared the winner. This seems to be a fair and logical conclusion; most people would agree that if a candidate received more than half of the votes, that candidate should be declared the winner. This conclusion is known as the **majority criterion of fairness**.

THE MAJORITY CRITERION

If candidate X receives a majority of the votes, then candidate X should be declared the winner.

Unfortunately, not all systems of voting satisfy this criterion; that is, in some circumstances, a candidate might receive a majority of the votes, yet another candidate might be declared the winner! Specifically, Exercise 23 will illustrate that the Borda count method can at times violate the majority criterion.

If you were a candidate in an election running against two other candidates, say, A and B, and you received more votes when compared to either A or B individually, you would expect to be declared the winner. In other words, suppose that you were running only against candidate A and you beat candidate A; then suppose that you were running only against candidate B and you beat candidate B. The fair and logical conclusion would be that you triumphed over each candidate, A and B, and therefore, you should be declared the winner. This conclusion is known as the **head-to-head criterion of fairness**.

THE HEAD-TO-HEAD CRITERION

If candidate X is favored when compared head-to-head (individually) with each of the other candidates, then candidate X should be declared the winner.

Unfortunately, not all systems of voting satisfy this criterion; that is, in some circumstances, a candidate might triumph in every possible head-to-head comparison, yet another candidate might be declared the overall winner! Specifically, Exercise 24 will illustrate that the plurality method can in fact violate the head-to-head criterion.

Many times, regular elections are preceded by preliminary, nonbinding elections called *straw votes*. Suppose you were the winning candidate in a straw vote. Then, the regular election was held, and the only changes in the votes were changes in your favor; that is, some people now voted for you instead of their original candidate. Under these conditions, a fair and logical conclusion would be that you should be declared the winner. This conclusion is known as the **monotonicity criterion of fairness**.

THE MONOTONICITY CRITERION

If candidate X wins an election and, in a subsequent election, the only changes are changes in favor of candidate X, then candidate X should be declared the winner.

Unfortunately, not all systems of voting satisfy this criterion; that is, in some circumstances, a candidate might win an election (or straw vote) and gather more votes in a second election but lose out to another candidate in the second election. Specifically, Exercise 25 will illustrate that the instant runoff method can in fact violate the monotonicity criterion.

Suppose you win an election. However, one or more of the (losing) candidates are subsequently determined to be ineligible to run in the election; consequently, they are removed from the ballot, and a recount is initiated. Under these conditions, a fair and logical conclusion would be that you should still be declared the winner. This conclusion is known as the **irrelevant alternatives criterion of fairness**.

THE IRRELEVANT ALTERNATIVES CRITERION

If candidate X wins an election and, in a recount, the only changes are that one or more of the losing candidates are removed from the ballot, then candidate X should still be declared the winner.

Unfortunately, not all systems of voting satisfy this criterion; that is, in some circumstances, a candidate might win an election, but on the removal of other candidates from the ballot, another candidate might be declared the winner! Specifically, Exercise 26 will illustrate that the pairwise comparison method can in fact violate the irrelevant alternatives criterion.

Each of the common systems of voting studied in this section can be shown to violate at least one of the four fairness criteria. Is it possible to create an ultimate system that satisfies all four criteria? The answer is “no”; it is mathematically impossible to create a system of voting that satisfies all four fairness criteria. This result was proven in 1951 by Kenneth Arrow and is known as **Arrow’s Impossibility Theorem**.

ARROW’S IMPOSSIBILITY THEOREM

It is mathematically impossible to create any system of voting (involving three or more candidates) that satisfies all four fairness criteria.

The systems of voting studied in this section are summarized in Figure 6.19.

System of Voting	Description
Plurality	Each person votes for his or her favorite candidate. The candidate who receives the most votes is declared the winner.
Instant Run-Off	Each voter ranks all of the candidates. If a candidate receives a majority (more than half) of first-choice votes, that candidate is declared the winner. If no candidate receives a majority, then the candidate with the fewest first-choice votes is eliminated and those votes are given to the next preferred candidate. If a candidate now has a majority of first-choice votes, that candidate is declared the winner. If no candidate receives a majority, this process continues until a candidate receives a majority.
Borda count	Each voter ranks all of the candidates. If there are k candidates, each candidate receives k points for each first-choice vote, $(k - 1)$ points for each second-choice vote, $(k - 2)$ points for each third-choice vote, and so on. The candidate with the most total points is declared the winner.
Pairwise comparison	Each voter ranks all of the candidates. For each possible pairing of candidates, the candidate with the most votes receives 1 point; if there is a tie, each candidate receives $\frac{1}{2}$ point. The candidate who receives the most points is declared the winner.

FIGURE 6.19 Systems of voting.

The systems of voting studied in this section each satisfy (or violate) the fairness criteria as summarized in Figure 6.20.

Fairness Criteria	System of Voting			
	Plurality Method	Instant Runoff Method	Borda Count Method	Pairwise Comparison Method
Majority Criterion	always satisfies	always satisfies	may not satisfy	always satisfies
Head-to-Head Criterion	may not satisfy	may not satisfy	may not satisfy	always satisfies
Monotonicity Criterion	always satisfies	may not satisfy	always satisfies	always satisfies
Irrelevant Alternatives Criterion	may not satisfy	may not satisfy	may not satisfy	may not satisfy

FIGURE 6.20 Fairness and flaws of the systems of voting.

6.1 EXERCISES

- Four candidates, Allioti, Baker, Cruz, and Daud, are running for president of the student government. After the polls close, votes are tallied, and the results in Figure 6.21 are obtained.

Candidate	Allioti	Baker	Cruz	Daud
Number of votes	314	155	1,052	479

FIGURE 6.21 Table for Exercise 1.

- How many votes were cast?
 - Using the plurality method of voting, which candidate wins?
 - Did the winner receive a majority of the votes?
- Five candidates, Edwards, Fischer, Gelinias, Horner, and Inclan, are running for president of the faculty senate. After the polls close, votes are tallied, and the results in Figure 6.22 are obtained.

Candidate	Edwards	Fischer	Gelinias	Horner	Inclan
Number of votes	32	25	17	102	24

FIGURE 6.22 Table for Exercise 2.

- How many votes were cast?
- Using the plurality method of voting, which candidate wins?
- Did the winner receive a majority of the votes?

- Three candidates, Arce, Edelstein, and Spence, are running for president of academic affairs at a local college. After the polls close, votes are tallied, and the results in Figure 6.23 are obtained.

Candidate	Arce	Edelstein	Spence
Number of votes	3,021	4,198	3,132

FIGURE 6.23 Table for Exercise 3.

- How many votes were cast?
 - Using the plurality method of voting, which candidate wins?
 - Did the winner receive a majority of the votes?
- Four candidates, Brecha, Parks, Wilcox, and Willett, are running for district supervisor. After the polls close, votes are tallied, and the results in Figure 6.24 are obtained.

Candidate	Brecha	Parks	Wilcox	Willett
Number of votes	3,007	2,957	10,541	2,851

FIGURE 6.24 Table for Exercise 4.

- How many votes were cast?
- Using the plurality method of voting, which candidate wins?
- Did the winner receive a majority of the votes?

- ▶ 5. The managers of Prints Alive (a local silk-screening business) are planning a party for their 30 employees. There are three possible locations for the party: the warehouse (W), the park (P), or the beach (B). The employees are asked to rank these choices in order of preference, and the results are summarized in Figure 6.25.

	Number of Ballots Cast			
	6	8	11	5
1st choice	P	P	B	W
2nd choice	B	W	W	B
3rd choice	W	B	P	P

FIGURE 6.25 Voter preference table for Exercise 5.

- How many votes were cast?
 - Use the plurality method of voting to determine the winner.
 - What percent of the votes did the winner in part (b) receive?
 - Use the instant runoff method to determine the winner.
 - What percent of the votes did the winner in part (d) receive?
 - Use the Borda count method to determine the winner.
 - How many points did the winner in part (f) receive?
 - Use the pairwise comparison method to determine the winner.
 - How many points did the winner in part (h) receive?
6. The members of a local service club are volunteering to clean up and modernize the playground at one of the elementary schools in town. There are three schools: Hidden Lakes (H), Strandwood (S), and Valhalla (V). The members are asked to rank these choices in order of preference, and the results are summarized in Figure 6.26.

	Number of Ballots Cast			
	13	8	10	19
1st choice	S	V	V	H
2nd choice	H	S	H	S
3rd choice	V	H	S	V

FIGURE 6.26 Voter preference table for Exercise 6.

- How many votes were cast?
- Use the plurality method of voting to determine the winner.

- What percent of the votes did the winner in part (b) receive?
- Use the instant runoff method to determine the winner.
- What percent of the votes did the winner in part (d) receive?
- Use the Borda count method to determine the winner.
- How many points did the winner in part (f) receive?
- Use the pairwise comparison method to determine the winner.
- How many points did the winner in part (h) receive?

7. The members of a youth club are raising money so that they can attend a summer camp. There are three camps in the area: Coastline (C), Pinewood (P), and The Ranch (R). The members are asked to rank these choices in order of preference, and the results are summarized in Figure 6.27.

	Number of Ballots Cast				
	19	12	10	11	13
1st choice	C	C	P	R	R
2nd choice	P	R	C	P	C
3rd choice	R	P	R	C	P

FIGURE 6.27 Voter preference table for Exercise 7.

- How many votes were cast?
 - Use the plurality method of voting to determine the winner.
 - What percent of the votes did the winner in part (b) receive?
 - Use the instant runoff method to determine the winner.
 - What percent of the votes did the winner in part (d) receive?
 - Use the Borda count method to determine the winner.
 - How many points did the winner in part (f) receive?
 - Use the pairwise comparison method to determine the winner.
 - How many points did the winner in part (h) receive?
8. The members of a local charitable group are raising money to send a group of neighborhood children to a special event. There are three events to choose from: the circus (C), the ice show (I), and the symphony (S). The members are asked to rank these choices in order of preference, and the results are summarized in Figure 6.28.
- How many votes were cast?
 - Use the plurality method of voting to determine the winner.

	Number of Ballots Cast				
	6	8	7	10	15
1st choice	C	C	I	I	S
2nd choice	I	S	C	S	C
3rd choice	S	I	S	C	I

FIGURE 6.28 Voter preference table for Exercise 8.

- What percent of the votes did the winner in part (b) receive?
 - Use the instant runoff method to determine the winner.
 - What percent of the votes did the winner in part (d) receive?
 - Use the Borda count method to determine the winner.
 - How many points did the winner in part (f) receive?
 - Use the pairwise comparison method to determine the winner.
 - How many points did the winner in part (h) receive?
9. Three candidates, Budd (B), Nirgiotis (N), and Shattuck (S), are running for union president. After the polls close, ranked ballots are tallied, and the results are summarized in Figure 6.29.

	Number of Ballots Cast					
	25	13	19	27	30	26
1st choice	B	B	N	N	S	S
2nd choice	N	S	S	B	B	N
3rd choice	S	N	B	S	N	B

FIGURE 6.29 Voter preference table for Exercise 9.

- How many votes were cast?
- Use the plurality method of voting to determine the winner.
- What percent of the votes did the winner in part (b) receive?
- Use the instant runoff method to determine the winner.
- What percent of the votes did the winner in part (d) receive?
- Use the Borda count method to determine the winner.
- How many points did the winner in part (f) receive?
- Use the pairwise comparison method to determine the winner.
- How many points did the winner in part (h) receive?

10. Three candidates, Maruyama (M), Peters (P), and Vilas (V), are running for district representative. After the polls close, ranked ballots are tallied, and the results are summarized in Figure 6.30.

	Number of Ballots Cast					
	675	354	451	387	601	297
1st choice	M	M	P	P	V	V
2nd choice	P	V	V	M	M	P
3rd choice	V	P	M	V	P	M

FIGURE 6.30 Voter preference table for Exercise 10.

- How many votes were cast?
 - Use the plurality method of voting to determine the winner.
 - What percent of the votes did the winner in part (b) receive?
 - Use the instant runoff method to determine the winner.
 - What percent of the votes did the winner in part (d) receive?
 - Use the Borda count method to determine the winner.
 - How many points did the winner in part (f) receive?
 - Use the pairwise comparison method to determine the winner.
 - How many points did the winner in part (h) receive?
11. Four candidates, Dolenz (D), Jones (J), Nesmith (N), and Tork (T), are running for director of public relations. After the polls close, ranked ballots are tallied, and the results are summarized in Figure 6.31.

	Number of Ballots Cast						
	225	134	382	214	81	197	109
1st choice	D	D	J	J	N	T	T
2nd choice	J	N	D	T	J	D	J
3rd choice	T	T	T	D	T	N	D
4th choice	N	J	N	N	D	J	N

FIGURE 6.31 Voter preference table for Exercise 11.

- How many votes were cast?
- Use the plurality method of voting to determine the winner.
- What percent of the votes did the winner in part (b) receive?
- Use the instant runoff method to determine the winner.

- e. What percent of the votes did the winner in part (d) receive?
- f. Use the Borda count method to determine the winner.
- g. How many points did the winner in part (f) receive?
- h. Use the pairwise comparison method to determine the winner.
- i. How many points did the winner in part (h) receive?
- ▶ 12. Four candidates, Harrison (H), Lennon (L), McCartney (M), and Starr (S), are running for regional manager. After the polls close, ranked ballots are tallied, and the results are summarized in Figure 6.32.

		Number of Ballots Cast						
		23	98	45	82	32	17	21
1st choice	H	L	L	M	M	S	S	
2nd choice	L	M	S	L	L	H	M	
3rd choice	S	H	M	H	S	L	H	
4th choice	M	S	H	S	H	M	L	

FIGURE 6.32 Voter preference table for Exercise 12.

- a. How many votes were cast?
- b. Use the plurality method of voting to determine the winner.
- c. What percent of the votes did the winner in part (b) receive?
- d. Use the instant runoff method to determine the winner.
- e. What percent of the votes did the winner in part (d) receive?
- f. Use the Borda count method to determine the winner.
- g. How many points did the winner in part (f) receive?
- h. Use the pairwise comparison method to determine the winner.
- i. How many points did the winner in part (h) receive?
13. Five candidates, Addley (A), Burke (B), Ciento (C), Darter (D), and Epp (E), are running for mayor. After the polls close, ranked ballots are tallied, and the results are summarized in Figure 6.33.
- a. How many votes were cast?
- b. Use the plurality method of voting to determine the winner.
- c. What percent of the votes did the winner in part (b) receive?
- d. Use the instant runoff method to determine the winner.

- e. What percent of the votes did the winner in part (d) receive?
- f. Use the Borda count method to determine the winner.
- g. How many points did the winner in part (f) receive?
- h. Use the pairwise comparison method to determine the winner.
- i. How many points did the winner in part (h) receive?

		Number of Ballots Cast						
		1,897	1,025	4,368	2,790	6,897	9,571	5,206
1st choice	A	A	B	C	D	D	E	
2nd choice	C	D	D	E	B	E	D	
3rd choice	D	C	C	D	C	B	C	
4th choice	B	E	E	B	E	A	B	
5th choice	E	B	A	A	A	C	A	

FIGURE 6.33 Voter preference table for Exercise 13.

- ▶ 14. Five candidates, Fino (F), Gempler (G), Holloway (H), Isho (I), and James (J), are running for president of the Polar Bear Swim Club. After the polls close, ranked ballots are tallied, and the results are summarized in Figure 6.34.

		Number of Ballots Cast						
		12	18	29	11	21	19	18
1st choice	F	F	G	H	I	J	J	
2nd choice	G	I	F	G	J	H	F	
3rd choice	I	G	H	F	G	I	G	
4th choice	J	H	J	I	F	G	I	
5th choice	H	J	I	J	H	F	H	

FIGURE 6.34 Voter preference table for Exercise 14.

- a. How many votes were cast?
- b. Use the plurality method of voting to determine the winner.
- c. What percent of the votes did the winner in part (b) receive?
- d. Use the instant runoff method to determine the winner.
- e. What percent of the votes did the winner in part (d) receive?
- f. Use the Borda count method to determine the winner.
- g. How many points did the winner in part (f) receive?

- h. Use the pairwise comparison method to determine the winner.
 - i. How many points did the winner in part (h) receive?
15. If there are six candidates in an election and voters are asked to rank all of the candidates, how many different rankings are possible?
 - ▶ 16. If there are eight candidates in an election and voters are asked to rank all of the candidates, how many different rankings are possible?
 17. If there are six candidates in an election and voters are asked to rank all of the candidates, how many different pairwise comparisons are there?
 - ▶ 18. If there are eight candidates in an election and voters are asked to rank all of the candidates, how many different pairwise comparisons are there?
 19. In an election, there are three candidates and 25 voters.
 - a. What is the maximum number of points that a candidate can receive using the Borda count method?
 - b. What is the minimum number of points that a candidate can receive using the Borda count method?
 - ▶ 20. In an election, there are four candidates and 75 voters.
 - a. What is the maximum number of points that a candidate can receive using the Borda count method?
 - b. What is the minimum number of points that a candidate can receive using the Borda count method?
 21. In an election, there are seven candidates.
 - a. What is the maximum number of points that a candidate can receive using the pairwise comparison method?
 - b. What is the minimum number of points that a candidate can receive using the pairwise comparison method?
 - ▶ 22. In an election, there are nine candidates.
 - a. What is the maximum number of points that a candidate can receive using the pairwise comparison method?
 - b. What is the minimum number of points that a candidate can receive using the pairwise comparison method?
 23. Candidates A, B, and C are being considered as supervisor of a local school district. There are thirteen directors on the school board, and they have ranked their choices as shown in Figure 6.35.
 - a. Does any candidate have a majority of first-choice votes? Who should win?
 - b. Use the Borda count method to determine the winner.
 - c. Does the Borda count method violate the majority criterion of fairness? Explain why or why not.

	Number of Ballots Cast		
	7	4	2
1st choice	A	B	C
2nd choice	B	C	B
3rd choice	C	A	A

FIGURE 6.35 Voter preference table for Exercise 23.

- ▶ 24. Candidates A, B, and C are being considered as chancellor of a local college district. There are twenty-one directors on the college board, and they have ranked their choices as shown in Figure 6.36.

	Number of Ballots Cast			
	9	2	4	6
1st choice	A	C	B	B
2nd choice	B	A	A	C
3rd choice	C	B	C	A

FIGURE 6.36 Voter preference table for Exercise 24.

- a. Use the plurality method to determine the winner.
 - b. Who wins when A is compared to B?
 - c. Who wins when A is compared to C?
 - d. Who wins when B is compared to C?
 - e. Does the plurality method violate the head-to-head criterion of fairness? Explain why or why not.
25. Candidates A, B, and C are being considered as director of a public service agency. There are twenty-seven trustees on the executive board, and after the initial discussion and casting of straw votes, the trustees have ranked their choices as shown in Figure 6.37.

	Number of Ballots Cast			
	6	8	4	9
1st choice	A	B	A	C
2nd choice	B	C	C	A
3rd choice	C	A	B	B

FIGURE 6.37 Original voter preference table for Exercise 25.

- a. Use the instant runoff method to determine the winner of the straw vote.
- b. After discussing the results of the straw vote, four trustees changed their votes to support candidate C as shown in Figure 6.38. Use the instant runoff method to determine the winner of the subsequent election.

	Number of Ballots Cast		
	6	8	13
1st choice	A	B	C
2nd choice	B	C	A
3rd choice	C	A	B

FIGURE 6.38 Revised voter preference table for Exercise 25.

- c. Does the instant runoff method violate the monotonicity criterion of fairness? Explain why or why not.
- ▶ 26. Candidates A, B, C, and D are being considered as director of an endowment for the arts trust fund. There are twenty-eight trustees on the executive board, and they have ranked their choices as shown in Figure 6.39.

	Number of Ballots Cast			
	10	8	6	4
1st choice	A	C	D	D
2nd choice	B	B	A	A
3rd choice	C	D	C	B
4th choice	D	A	B	C

FIGURE 6.39 Original voter preference table for Exercise 26.

- a. Use the pairwise comparison method to determine the winner.

- b. After the trustees had ranked the candidates, it was discovered that candidates B and C did not meet the minimum qualifications and hence were ineligible for the position. Consequently, candidates B and C were removed, and the voter preference table in Figure 6.40 was established. Use the pairwise comparison method to determine the new winner.

	10	8	6	4
1st	A	D	D	D
2nd	D	A	A	A

FIGURE 6.40 Revised voter preference table for Exercise 26.

- c. Does the pairwise comparison method violate the irrelevant alternatives criterion of fairness? Explain why or why not.



Answer the following questions using complete sentences and your own words.

• CONCEPT QUESTIONS

- What is a majority?
- What is a ranked ballot?
- What is a voter preference table?
- What are the four fairness criteria? Explain the meaning of each.
- What is Arrow's Impossibility Theorem?

6.2

Methods of Apportionment

OBJECTIVES

- Develop and apply Hamilton's Method of apportionment
- Develop and apply Jefferson's Method of apportionment
- Develop and apply Adams's Method of apportionment
- Develop and apply Webster's Method of apportionment
- Develop and apply the Hill-Huntington Method of apportionment

After the thirteen former British colonies won their independence, the “founding fathers” had to come up with a constitution to govern the newly formed nation. Needless to say, it was an enormous task to create the framework of a government



Christy, Howard Chandler (1873–1952)/Hall of Representatives, Washington, D.C., USA/The Bridgeman Art Library International



© Joseph Solmi, Visions of America/CORBIS

In the Constitution, the “founding fathers” created two legislative bodies: the Senate (two members per state) and the House of Representatives (apportioned to each state’s population).

rooted in revolution and driven by the underlying principle of “power to the people.” The framers of the Constitution believed that the former colonies, because of their diversity, should be relatively independent states; however, there should also be some sort of central or federal government to serve as a cohesive binding element to strengthen and secure the sovereignty of the nation. To that effect, the Preamble of the Constitution reads, “We, the People of the United States, in Order to form a more perfect Union, establish Justice, insure domestic Tranquility, provide for the common defense, promote the general Welfare, and secure the Blessings of Liberty to ourselves and our Posterity, do ordain and establish this Constitution for the United States of America.”

An important feature of any government is the method for creating the nation’s laws; this is commonly called the legislative branch of government. When delegates of the original thirteen states met in Philadelphia in 1787, a common debate was the manner in which the states would be represented in the federal legislature. Some states, in particular the smaller ones, advocated that each state receive the same number of representatives. On the other hand, many of the larger states wanted representation in proportion to the number of people residing in the individual states.

The result of this great debate was the creation of two distinct legislative bodies. Article I, Section 1 of the Constitution states, “All legislative Powers herein granted shall be vested in a Congress of the United States, which shall consist of a Senate and House of Representatives.” Article I, Section 3 created the Senate and states “The Senate of the United States shall be composed of two Senators from each state.” In contrast, Article I, Section 2 created the House of Representatives and states, “Representatives and direct taxes shall be apportioned among the several States which may be included within this Union, according to their respective Numbers, which shall be determined by adding to the whole Number of free Persons, including those bound to Service for a Term of Years, and excluding Indians not taxed, three-fifths of all other persons. The actual Enumeration shall be made within three Years after the first Meeting of the Congress of the United States, and within every subsequent Term of ten Years, in such Manner as they shall by Law direct. The Number of Representatives shall not exceed one for every thirty Thousand, but each State shall have at Least one Representative.”

It is very easy to interpret the phrase “The Senate of the United States shall be composed of two Senators from each state.” However, the phrase “Representatives shall be *apportioned* among the several states according to

their respective numbers” is nebulous at best. When we consult a standard dictionary, we find that the word **apportion** is defined as “to divide and distribute in shares according to a plan.” Although the Constitution specifies that Representatives be apportioned or distributed according to a plan, it does not specify the plan. Consequently, several methods of apportionment have been proposed, and several different methods have actually been used since the first apportionment in 1790.

In this section, we will study the following methods of apportionment:

- **Hamilton’s Method** was proposed by Alexander Hamilton (1757–1804) and was the first plan to be approved by Congress in 1790. However, President George Washington vetoed the plan (the first use of a presidential veto in the history of the United States). Eventually, Hamilton’s method was used following each federal government census (every ten years) from 1850 to 1900.
- **Jefferson’s Method** was proposed by Thomas Jefferson (1743–1826) and was the first method to actually be used; it was used following each census from 1790 to 1830.
- **Adams’s Method** was proposed by John Quincy Adams (1767–1848) and has never been used. The method merits investigation, as it is viewed as the exact opposite of Jefferson’s Method.
- **Webster’s Method** was proposed by Daniel Webster (1782–1852). It was used following the census of 1840 and then again in 1910 and 1930. (There was no apportionment in 1920.)
- **Hill-Huntington Method** was proposed by Joseph Hill (1860–1938) and Edward Huntington (1874–1952), and it has been used following every census from 1940 to the present. During the early twentieth century, Hill was the chief statistician at the U.S. Census Bureau, and Huntington was a professor of mathematics and mechanics at Harvard.

Basic Terminology

Before we discuss any specific method of apportionment, we must introduce some basic concepts and define some basic terms.

STANDARD DIVISOR

The **standard divisor**, denoted by d , is the ratio of the total population to the total number of seats to be allocated; d is found by dividing the total population by the total number of seats to be allocated.

$$d = \text{standard divisor} = \frac{\text{total population}}{\text{total number of seats}}$$

It is customary to round the standard divisor to two decimal places.

For example, if the total population is 201,000 people and ten seats are to be allocated, the standard divisor will be

$$\frac{201,000 \text{ people}}{10 \text{ seats}} = 20,100$$

Therefore, $d = 20,100$ people per seat, or alternatively, each seat will represent 20,100 people. Once the standard divisor has been calculated, each state's **standard quota** must be determined

STANDARD QUOTA

The **standard quota** (of a specific state), denoted by q , is the ratio of a state's population to the standard divisor; q is found by dividing the state's population by the standard divisor.

$$q = \text{standard quota} = \frac{\text{state's population}}{\text{standard divisor}} = \frac{\text{state's population}}{d}$$

It is customary to round the standard quota to two decimal places.

For example, if the population of state A was 94,700 and the standard divisor was $d = 20,100$, then state A's standard quota would be

$$\frac{94,700}{20,100} = 4.711442786 \dots$$

Therefore, $q = 4.71$ (rounded to two decimal places). Subscripts may be used to identify a specific state's standard quota, that is, we can use q_A to represent the standard quota of state A; hence, $q_A = 4.71$. Similarly, we use q_B to represent the standard quota of state B, q_C to represent the standard quota of state C, and so on. Rounding to two decimal places provides satisfactory results in most cases. However, if rounding to two decimal places creates an integer (when the standard quota is not an exact integer), then round to three (or more) decimal places so that an integer is not obtained. That is, if we round $q = \frac{94,700}{18,950} = 4.997361478 \dots$ to two decimal places, we obtain $q = 5.00$, an integer. Instead, we round to three decimal places and obtain $q = 4.997$.

A standard quota can be interpreted as the number of seats allocated to a specific state; therefore, if $q_A = 4.71$, state A should receive 4.71 seats. Because a state cannot receive part of a seat, we conclude that state A should receive at least four seats. The “whole number” part of a standard quota is called the **lower quota** of the state.

LOWER QUOTA

The **lower quota** (of a specific state) is the standard quota of a state truncated (rounded down) to a whole number. That is, the lower quota is the whole number part of a standard quota. (Use the complete, original standard quota, not the rounded version, when finding the lower quota.) If the standard quota is an exact integer, then the lower quota equals the standard quota.

Hamilton's Method

We begin our study of the methods of apportionment with Hamilton's Method. Hamilton's Method is the easiest and most direct method to implement; it was the first plan to be approved by Congress in 1790, although it was not officially used until 1850.

HAMILTON'S METHOD OF APPORTIONMENT

1. Using the standard divisor d , calculate the standard quotas and the lower quotas of each state. Initially, each state receives a number of seats equal to its lower quota.
2. If the sum of the lower quotas equals the total number of seats to be apportioned, the apportionment process is complete.
3. If the sum of the lower quotas is less than the total number of seats to be apportioned, then assign a seat to the state that has the highest decimal part in its standard quota.
4. Repeat step 3 (using the next highest decimal part) until the total number of seats has been apportioned.

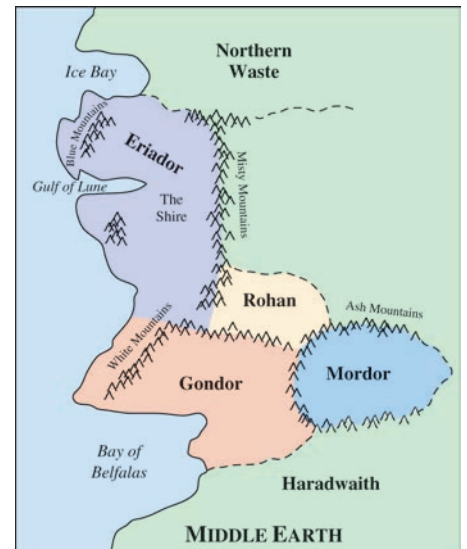


© Warner Brothers/PhotoFest

Inhabitants of Middle Earth.

EXAMPLE 1

APPLYING HAMILTON'S METHOD OF APPORTIONMENT Middle Earth is a fantastic world created by the literary genius of J. R. R. Tolkien; Middle Earth's principle inhabitants are hobbits, dwarves, elves, wizards, men, and many mutant forms of evil. Let us suppose that the good inhabitants of the realms of Gondor, Eriador, and Rohan have agreed to form a federation to foster mutual protection, trade, and cultural exchange. The (estimated) populations of these realms (or states) are given in Figure 6.41. The high council of the federation is to have ten seats. Use Hamilton's Method to apportion the seats among these three realms (states).



Realm (State)	Gondor	Eriador	Rohan	Total
Population (estimated)	94,700	72,600	33,700	201,000

FIGURE 6.41 Realms of Middle Earth.

SOLUTION

It is often advantageous to express large numbers that have several zeroes at the end in terms of simpler multiples. That is, rather than working with the numeral 94,700, we can express it as 94.7 thousand (or $94.7 \times 1,000$), and then use only the numeral 94.7. This will consistently simplify our calculations. Figure 6.42 shows the simplified population data in terms of thousands.

Realm (State)	Gondor	Eriador	Rohan	Total
Population (thousands)	94.7	72.6	33.7	201

FIGURE 6.42 Realms of Middle Earth.

First, we must find the standard divisor:

$$\begin{aligned} d &= \text{standard divisor} \\ &= \frac{\text{total population}}{\text{total number of seats}} \\ &= \frac{201}{10} = 20.1 \end{aligned}$$

Therefore, $d = 20.1$.

Now calculate each realm's standard quota, starting with Gondor (G):

$$\begin{aligned} q_G &= \text{Gondor's standard quota} \\ &= \frac{\text{Gondor's population}}{d} \\ &= \frac{94.7}{20.1} \\ &= 4.711442786 \dots \end{aligned}$$

Therefore $q_G = 4.71$.

In a similar fashion, we calculate the standard quotas for Eriador (E) and Rohan (R):

$$\begin{aligned} q_E &= \text{Eriador's standard quota} & q_R &= \text{state C's standard quota} \\ &= \frac{\text{Eriador's population}}{d} & &= \frac{\text{Rohan's population}}{d} \\ &= \frac{72.6}{20.1} & &= \frac{33.7}{20.1} \\ q_E &= 3.61 \quad \text{rounded to two} & q_R &= 1.68 \quad \text{rounded to two} \\ & \quad \text{decimal places} & & \quad \text{decimal places} \end{aligned}$$

The standard quotas and lower quotas for Gondor, Eriador, and Rohan are summarized in Figure 6.43.

Realm (State)	Gondor	Eriador	Rohan	Total
Population (thousands)	94.7	72.6	33.7	201
Standard Quota (using $d = 20.1$)	4.71	3.61	1.68	10
Lower Quota	4	3	1	8

FIGURE 6.43 Standard and lower quotas of the realms.

Because the sum of the lower quotas (8) is less than the total number of seats (10), we must decide who receives the two ($10 - 8 = 2$) “surplus” seats. According to Hamilton’s method, Gondor will receive the first additional seat (Figure 6.44) because Gondor has the highest decimal part in its standard quota (0.71 is larger than either 0.61 or 0.68), and Rohan will receive the next seat (because $0.68 > 0.61$).

Realm (State)	Gondor	Eriador	Rohan	Total
Population (thousands)	94.7	72.6	33.7	201
Standard Quota ($d = 20.1$)	4.71	3.61	1.68	10
Lower Quota	4	3	1	8
Additional Seats	1	0	1	2
Number of Seats	5	3	2	10

FIGURE 6.44 Hamilton’s Method applied to the realms of Middle Earth.

By using Hamilton’s Method, the ten seats will be apportioned as shown in Figure 6.45.

Realm (State)	Gondor	Eriador	Rohan	Total
Hamilton’s Apportionment	5	3	2	10

FIGURE 6.45 Final apportionment of the seats of the high council.

Jefferson’s Method

As we have seen, Hamilton’s Method utilizes the standard divisor ($d = \frac{\text{state's population}}{\text{total number of seats}}$), and this method may result in surplus seats that must be distributed accordingly. Consequently, some states receive preferential treatment by obtaining extra seats. Thomas Jefferson proposed a method in which no surplus seats are created. The essence of Jefferson’s Method is this: If the standard divisor creates surplus seats, then find a new divisor (called a modified divisor) that will result in all of the seats being allocated. As we shall see, Jefferson’s Method requires a bit of trial and error.

A **modified divisor** is a number that is close to the standard divisor, and it is denoted by d_m . A modified divisor may be less than or greater than the standard divisor. Jefferson’s method calls for a modified divisor that is *less than* the standard divisor, that is, $d_m < d$. Now, rather than calculating the standard quota, we calculate the **modified quota**.

MODIFIED QUOTA

The **modified quota** (of a specific state), denoted by q_m , is the ratio of a state’s population to the modified divisor; q_m is found by dividing the state’s population by the modified divisor.

$$q_m = \text{modified quota} = \frac{\text{state's population}}{\text{modified divisor}} = \frac{\text{state's population}}{d_m}$$

It is customary to round off the modified quota to two decimal places.

As a consequence, the modified quota (in Jefferson's Method) will always be larger than the standard quota (in Hamilton's Method) because we are dividing the state's population by a smaller number. (Dividing by a smaller number produces a larger result.) Because the modified quotas are larger than the standard quotas, our goal is to find a modified divisor that will automatically "use up" any surplus seats.

JEFFERSON'S METHOD OF APPORTIONMENT

1. Using the standard divisor d , calculate the standard quotas and the lower quotas of each state.
2. If the sum of the lower quotas equals the total number of seats to be apportioned, the apportionment process is complete, that is, each state receives a number of seats equal to its lower quota.
3. If the sum of the lower quotas does not equal the total number of seats to be apportioned, choose a modified divisor d_m less than the standard divisor, that is, $d_m < d$, and calculate the modified quotas and lower modified quotas.
4. Repeat step 3 until you find a modified divisor such that the sum of the lower modified quotas equals the total number of seats to be apportioned. Each state receives a number of seats equal to its lower modified quota, and the apportionment process is complete.

EXAMPLE 2

APPLYING JEFFERSON'S METHOD OF APPORTIONMENT As in Example 1, the populations of Gondor, Eriador, and Rohan are given in Figure 6.46. Use Jefferson's Method to apportion the ten seats of the high council (legislature) among these three realms.

Realm (State)	Gondor	Eriador	Rohan	Total
Population	94,700	72,600	33,700	201,000

FIGURE 6.46 Realms of Middle Earth.

SOLUTION

As we saw in Example 1, using Hamilton's Method and the standard divisor $d = 20.1$ results in two surplus seats, as shown in Figure 6.47.

Realm (State)	Gondor	Eriador	Rohan	Total
Population (thousands)	94.7	72.6	33.7	201
Standard Quota ($d = 20.1$)	4.71	3.61	1.68	10
Lower Quota	4	3	1	8

FIGURE 6.47 Standard and lower quotas of the realms.

Therefore, to apply Jefferson's Method, we must find a modified divisor, $d_m < d = 20.1$, such that no surplus is created. As a first guess, we try $d_m = 19$ and calculate each realm's modified quota and lower modified quota; that is, divide each population by $d_m = 19$, round to two decimal places, and then truncate (round down) to a whole number as shown in Figure 6.48.

Realm (State)	Gondor	Eriador	Rohan	Total
Population (thousands)	94.7	72.6	33.7	201
Standard Quota ($d = 20.1$)	4.71	3.61	1.68	10
Modified Quota ($d_m = 19$)	4.98	3.82	1.77	(not needed)
Lower Modified Quota	4	3	1	8

FIGURE 6.48 Modified and lower modified quotas using $d_m = 19$.

The modified divisor $d_m = 19$ still creates two surplus seats, so we must try an even smaller divisor, say, $d_m = 18$, and repeat the process; that is, divide each population by $d_m = 18$, round to two decimal places, and then truncate to a whole number as shown in Figure 6.49.

Realm (State)	Gondor	Eriador	Rohan	Total
Population (thousands)	94.7	72.6	33.7	201
Standard Quota ($d = 20.1$)	4.71	3.61	1.68	10
Modified Quota ($d_m = 19$)	4.98	3.82	1.77	(not needed)
Modified Quota ($d_m = 18$)	5.26	4.03	1.87	(not needed)
Lower Modified Quota	5	4	1	10

FIGURE 6.49 Jefferson's Method using $d_m = 18$.

The modified divisor $d_m = 18$ has properly apportioned the ten seats, and the final results are shown in Figure 6.50.

Realm (State)	Gondor	Eriador	Rohan	Total
Jefferson's Apportionment	5	4	1	10
Hamilton's Apportionment	5	3	2	10

FIGURE 6.50 Comparison of final apportionments.

Figure 6.50 shows that different methods of apportionment can lead to different allocations of legislative seats. Observe that Jefferson's Method favors larger states while Hamilton's Method favors smaller states. That is, under Jefferson's Method, Eriador received an extra seat (and the population of Eriador is greater than that of Rohan), while under Hamilton's Method, Rohan received an extra seat (and the population of Rohan is less than that of Eriador). We will explore these "flaws" in more detail in Section 6.3.

Adams's Method

As we have seen, Jefferson's Method utilizes a modified divisor that is less than the standard divisor, and the modified quotas are truncated (rounded down) to a whole

number (lower modified quota). Like Jefferson, John Quincy Adams proposed using a modified divisor; however, his approach was the opposite of Jefferson's: Adams proposed using a modified divisor *greater than* the standard divisor, and he rounded his modified quotas *up* to the **upper quota**.

UPPER QUOTA

The **upper quota** (of a specific state) is the standard (or modified) quota of a state rounded up to the next whole number. (Use the complete, original standard quota, not the rounded version, when finding the upper quota.) If the standard quota is an exact integer, then the upper quota equals the standard quota plus 1.

ADAMS'S METHOD OF APPORTIONMENT

1. Using the standard divisor d , calculate the standard quotas and the upper quotas of each state.
2. If the sum of the upper quotas equals the total number of seats to be apportioned, the apportionment process is complete, that is, each state receives a number of seats equal to its upper quota.
3. If the sum of the upper quotas does not equal the total number of seats to be apportioned, choose a modified divisor d_m greater than the standard divisor, that is, $d_m > d$ and calculate the modified quotas and upper modified quotas.
4. Repeat step 3 until you find a modified divisor such that the sum of the upper modified quotas equals the total number of seats to be apportioned. Each state receives a number of seats equal to its upper modified quota, and the apportionment process is complete.

In Adams's Method, the modified quota will always be smaller than the standard quota (in Hamilton's Method) because we are dividing the state's population by a larger number. (Dividing by a larger number produces a smaller result.) Because the modified quotas are smaller than the standard quotas (and we will use upper modified quotas), our goal is to find a modified divisor that will automatically allocate the proper number of seats.

EXAMPLE 3

APPLYING ADAMS'S METHOD OF APPORTIONMENT As in Example 1, the populations of Gondor, Eriador, and Rohan are given in Figure 6.51. Use Adams's Method to apportion the ten seats of the high council among these three realms.

Realm (State)	Gondor	Eriador	Rohan	Total
Population	94,700	72,600	33,700	201,000

FIGURE 6.51 Realms of Middle Earth.

SOLUTION

As we saw in Example 1, Hamilton's Method and the standard divisor $d = 20.1$ results in two unused seats as shown in Figure 6.52. Figure 6.52 also shows the upper quotas (standard quotas rounded up) for each realm.

Because the sum of the upper quotas (11) does not equal the total number of seats to be allocated (10), the apportionment is not complete (we have allocated

Realm (State)	Gondor	Eriador	Rohan	Total
Population (thousands)	94.7	72.6	33.7	201
Standard Quota ($d = 20.1$)	4.71	3.61	1.68	10
Lower Quota	4	3	1	8
Upper Quota	5	4	2	11

FIGURE 6.52 Standard, lower, and upper quotas of the realms.

one seat too many). Therefore, to apply Adams's Method, we must find a modified divisor, $d_m > d = 20.1$, such that the seats will be properly allocated. As a first guess, we try $d_m = 23$ and calculate each realm's modified quota and upper modified quota, that is, divide each population by $d_m = 23$, round to two decimal places, and then round up to a whole number as shown in Figure 6.53.

Realm (State)	Gondor	Eriador	Rohan	Total
Population (thousands)	94.7	72.6	33.7	201
Standard Quota ($d = 20.1$)	4.71	3.61	1.68	10
Modified Quota ($d_m = 23$)	4.12	3.16	1.47	(not needed)
Upper Modified Quota	5	4	2	11

FIGURE 6.53 Modified and upper modified quotas using $d_m = 23$.

The modified divisor $d_m = 23$ does not properly allocate the ten seats, so we must try an even larger divisor. Divisors need not equal whole numbers. For example, we can choose $d_m = 23.8$ and repeat the process; that is, we divide each population by $d_m = 23.8$, round to two decimal places, and find the upper quotas as shown in Figure 6.54.

Realm (State)	Gondor	Eriador	Rohan	Total
Population (thousands)	94.7	72.6	33.7	201
Standard Quota ($d = 20.1$)	4.71	3.61	1.68	10
Modified Quota ($d_m = 23$)	4.12	3.16	1.47	(not needed)
Modified Quota ($d_m = 23.8$)	3.98	3.05	1.42	(not needed)
Upper Modified Quota	4	4	2	10

FIGURE 6.54 Adams's Method using $d_m = 23.8$.

Because the sum of the upper quotas (10) equals the total number of seats to be allocated (10), the apportionment process is complete. By using Adams's Method, the ten seats will be apportioned as shown in Figure 6.55.

Realm (State)	Gondor	Eriador	Rohan	Total
Adams's Apportionment	4	4	2	10
Jefferson's Apportionment	5	4	1	10
Hamilton's Apportionment	5	3	2	10

FIGURE 6.55 Comparison of final apportionments.

Once again, we see (Figure 6.55) that different methods of apportionment can lead to significantly different allocations of seats in a legislature (high council). However, different methods of apportionment do not always lead to different allocations; different methods may lead to the same allocation of seats. Note that Adams's Method favors smaller states; that is, under Adams's Method, Eriador "took away" a seat from Gondor (and the population of Eriador is less than that of Gondor). Different methods of apportionment have "flaws," which will be examined in Section 6.3.

Webster's Method

As we have seen, both Hamilton's Method and Jefferson's Method consistently round *down* the quotas and thereby utilize *lower* quotas, whereas Adams's Method consistently rounds *up* the quotas and thereby utilizes *upper* quotas. Webster's Method differs from the previous methods in that this method utilizes *regular* rounding rules, that is, round down if a number is less than 5 and round up if it is 5 or more.

WEBSTER'S METHOD OF APPORTIONMENT

1. Using the standard divisor d , calculate the standard quotas of each state and use the regular rules of rounding to round each standard quota to a whole number.
2. If the sum of the rounded standard quotas equals the total number of seats to be apportioned, the apportionment process is complete, that is, each state receives a number of seats equal to its rounded standard quota.
3. If the sum of the rounded standard quotas does not equal the total number of seats to be apportioned, choose a modified divisor d_m that is different from the standard divisor (either less than or greater than d), and calculate the modified quotas and rounded modified quotas.
4. Repeat step 3 until you find a modified divisor such that the sum of the rounded modified quotas equals the total number of seats to be apportioned. Each state receives a number of seats equal to its rounded modified quota, and the apportionment process is complete.

EXAMPLE 4

APPLYING WEBSTER'S METHOD OF APPORTIONMENT As in Example 1, the populations of Gondor, Eriador, and Rohan are given in Figure 6.56. Use Webster's Method to apportion the ten seats of the high council among these three realms.

Realm (State)	Gondor	Eriador	Rohan	Total
Population	94,700	72,600	33,700	201,000

FIGURE 6.56 Realms of Middle Earth.

SOLUTION

As we saw in Example 1, the standard divisor $d = 20.1$ produces the standard quotas shown in Figure 6.57. Figure 6.57 also shows the rounded quotas (using the rules of regular rounding) for each realm.

Realm (State)	Gondor	Eriador	Rohan	Total
Population (thousands)	94.7	72.6	33.7	201
Standard Quota ($d = 20.1$)	4.71	3.61	1.68	10
Rounded Quota	5	4	2	11

FIGURE 6.57 Standard and (regular) rounded quotas of the realms.

The sum of the rounded quotas (11) is too large, so we must choose a modified divisor, $d_m > d = 20.1$, because dividing by a larger number creates a smaller result. As a first guess, we try $d_m = 21$ and calculate each realm's modified quota, that is, divide each population by $d_m = 21$, round to two decimal places, and then round to a whole number as shown in Figure 6.58.

Realm (State)	Gondor	Eriador	Rohan	Total
Population (thousands)	94.7	72.6	33.7	201
Standard Quota ($d = 20.1$)	4.71	3.61	1.68	10
Modified Quota ($d_m = 21$)	4.51	3.46	1.60	(not needed)
Rounded Quota	5	3	2	10

FIGURE 6.58 Webster's Method using $d_m = 21$.

Because the sum of the rounded quotas (10) equals the total number of seats to be allocated (10), the apportionment process is complete. By using Webster's Method, the ten seats will be apportioned as shown in Figure 6.59.

Realm (State)	Gondor	Eriador	Rohan	Total
Webster's Apportionment	5	3	2	10
Adams's Apportionment	4	4	2	10
Jefferson's Apportionment	5	4	1	10
Hamilton's Apportionment	5	3	2	10

FIGURE 6.59 Comparison of final apportionments.

In Example 4, the apportionment due to Webster's Method is the same as that due to Hamilton's Method. This does not always occur. Sometimes different methods lead to the same allocation of seats; other times, different methods may lead to different allocation of seats. It should also be noted that Webster's Method tends to favor smaller states.

Hill-Huntington Method

As we have seen, the methods of Hamilton and Jefferson consistently round *down* the quotas, the method advocated by Adams consistently rounds *up* the quotas, and quotas are subjected to the rules of *regular* rounding when Webster's Method is applied. The Hill-Huntington Method is similar to Webster's Method in that sometimes the quotas are rounded up and sometimes they are rounded down. Before this method is discussed, we must first explore the geometric mean.

When you study algebra, you undoubtedly encounter the calculation of an “average,” that is, many people recall that the “average” of two numbers is the sum of the numbers divided by 2. (See Section 4.2.) Technically, this type of average is called the **arithmetic mean**, that is, the arithmetic mean of a group of numbers equals the sum of the numbers divided by the number of numbers. There are other types of means. In particular, the Hill-Huntington Method utilizes the geometric mean.

GEOMETRIC MEAN

Given two numbers x and y , the **geometric mean** of x and y , denoted by gm , is the square root of the product of x and y . That is,

$$gm = \text{geometric mean} = \sqrt{xy}$$

For example, the geometric mean of 2 and 18 is $gm = \sqrt{2 \times 18} = \sqrt{36} = 6$, whereas the arithmetic mean is $\frac{2+18}{2} = \frac{20}{2} = 10$. We now describe the Hill-Huntington Method of apportionment.

HILL-HUNTINGTON METHOD OF APPORTIONMENT

1. Using the standard divisor d , calculate the standard quotas, lower quotas, and upper quotas of each state.
2. For each state, calculate the geometric mean (rounded to two decimal places) of its lower quota and upper quota. If the standard quota is less than the geometric mean, round the quota down; if the standard quota is greater than or equal to the geometric mean, round the quota up.
3. If the sum of the rounded standard quotas equals the total number of seats to be apportioned, the apportionment process is complete, that is, each state receives a number of seats equal to its rounded standard quota.
4. If the sum of the rounded standard quotas does not equal the total number of seats to be apportioned, choose a modified divisor d_m different from the standard divisor (either less than or greater than d), and calculate the modified quotas and rounded modified quotas.
5. Repeat step 4 until you find a modified divisor such that the sum of the rounded modified quotas equals the total number of seats to be apportioned. Each state receives a number of seats equal to its rounded modified quota, and the apportionment process is complete.

EXAMPLE 5

APPLYING THE HILL-HUNTINGTON METHOD OF APPORTIONMENT

As in Example 1, the populations of Gondor, Eriador, and Rohan are given in Figure 6.60. Use the Hill-Huntington Method to apportion the ten seats of the high council among these three realms.

Realm (State)	Gondor	Eriador	Rohan	Total
Population	94,700	72,600	33,700	201,000

FIGURE 6.60 Realms of Middle Earth.

SOLUTION

As we saw in Example 3, the standard divisor $d = 20.1$ produces the standard, lower, and upper quotas shown in Figure 6.61.

Beginning with Gondor, we now calculate the geometric mean of the lower and upper quotas for each state:

$$gm_G = \sqrt{4 \times 5} = \sqrt{20} = 4.472135955 \dots$$

Therefore, $gm_G = 4.47$.

Realm (State)	Gondor	Eriador	Rohan	Total
Population (thousands)	94.7	72.6	33.7	201
Standard Quota ($d = 20.1$)	4.71	3.61	1.68	10
Lower Quota	4	3	1	8
Upper Quota	5	4	2	11

FIGURE 6.61 Standard, lower, and upper quotas of the realms.

In a similar fashion, we calculate the geometric means for Eriador (E) and Rohan (R):

$gm_E = \text{Eriador's geometric mean}$ $= \sqrt{3 \times 4}$ $= \sqrt{12}$ $gm_E = 3.46 \quad \text{rounded to two decimal places}$	$gm_R = \text{Rohan's geometric mean}$ $= \sqrt{1 \times 2}$ $= \sqrt{2}$ $gm_R = 1.41 \quad \text{rounded to two decimal places}$
---	--

The standard quotas and geometric means for Gondor, Eriador, and Rohan are summarized in Figure 6.62.

Realm (State)	Gondor	Eriador	Rohan	Total
Population (thousands)	94.7	72.6	33.7	201
Standard Quota ($d = 20.1$)	4.71	3.61	1.68	10
Geometric Mean	4.47	3.46	1.41	(not needed)

FIGURE 6.62 Geometric means of the realms.

For each state, the standard quota is greater than the geometric mean; consequently, each standard quota is rounded up as shown in Figure 6.63.

Realm (State)	Gondor	Eriador	Rohan	Total
Population (thousands)	94.7	72.6	33.7	201
Standard Quota ($d = 20.1$)	4.71	3.61	1.68	10
Geometric Mean	4.47	3.46	1.41	(not needed)
Rounded Quota	5	4	2	11

FIGURE 6.63 Rounded quotas of the realms.

The sum of the rounded quotas (11) is too large, so we choose a modified divisor, $d_m > d = 20.1$, because dividing by a larger number creates a smaller result. As a first guess, we try $d_m = 21$ and calculate each realm’s modified quota; that is, we divide each population by $d_m = 21$, round to two decimal places, and then round to a whole number (using the geometric mean) as shown in Figure 6.64.

Realm (State)	Gondor	Eriador	Rohan	Total
Population (thousands)	94.7	72.6	33.7	201
Standard Quota ($d = 20.1$)	4.71	3.61	1.68	10
Geometric Mean	4.47	3.46	1.41	(not needed)
Modified Quota ($d_m = 21$)	4.51	3.46	1.60	(not needed)
Rounded Quota	5	4	2	11

FIGURE 6.64 The Hill-Huntington Method using $d_m = 21$.

The sum of the rounded quotas (11) is still too large, so we must choose an even larger modified divisor, say, $d_m = 21.2$, and repeat the process. That is, we divide each population by $d_m = 21.2$, round to two decimal places, and then round to a whole number (using the geometric mean) as shown in Figure 6.65.

Realm (State)	Gondor	Eriador	Rohan	Total
Population (thousands)	94.7	72.6	33.7	201
Standard Quota ($d = 20.1$)	4.71	3.61	1.68	10
Geometric Mean	4.47	3.46	1.41	(not needed)
Modified Quota ($d = 21.2$)	4.47	3.42	1.59	(not needed)
Rounded Quota	5	3	2	10

FIGURE 6.65 The Hill-Huntington Method using $d_m = 21.2$.

Because the sum of the rounded quotas (10) equals the total number of seats to be allocated (10), the apportionment process is complete. By using the Hill-Huntington Method, the ten seats will be apportioned as shown in Figure 6.66.

Realm (State)	Gondor	Eriador	Rohan	Total
Hill-Huntington's Apportionment	5	3	2	10
Webster's Apportionment	5	3	2	10
Adams's Apportionment	4	4	2	10
Jefferson's Apportionment	5	4	1	10
Hamilton's Apportionment	5	3	2	10

FIGURE 6.66 Comparison of final apportionments.

Additional Seats

Once the seats of a legislature have been allocated, it might be decided that the size of the legislature should be increased; that is, new seats might be added to an existing apportionment. Who gets the new seats? The answer lies in the calculation of **Hill-Huntington numbers**.

HILL-HUNTINGTON NUMBER

The **Hill-Huntington number** for a state, denoted HHN , is the square of the state's population divided by the product of its current number of seats n and $n + 1$. That is,

$$HHN = \frac{(\text{state's population})^2}{n(n + 1)}$$

where n = state's current number of seats.

In allocating a new seat to an existing legislature, the state with the highest HHN should receive the seat.

EXAMPLE 6

CALCULATING HILL-HUNTINGTON NUMBERS As we saw in Example 5, the populations of Gondor, Eriador, and Rohan and the apportionment of the ten seats via the Hill-Huntington Method are given in Figure 6.67. Suppose the high council decides to add an additional seat; that is, it is decided that the council should now consist of eleven seats. Use Hill-Huntington numbers to determine which realm should receive the new seat.

Realm (State)	Gondor	Eriador	Rohan	Total
Population (thousands)	94.7	72.6	33.7	201
Number of Seats	5	3	2	10

FIGURE 6.67 Apportionment of the high council.

SOLUTION

Beginning with Gondor, we calculate the Hill-Huntington number for each realm:

$$HHN_G = \frac{94.7^2}{5(5 + 1)} = \frac{94.7^2}{30} = 298.936333 \dots$$

Therefore, $HHN_G = 298.94$.

In a similar fashion, we calculate the Hill-Huntington numbers for Eriador (E) and Rohan (R):

$$\begin{aligned}
 HHN_E &= \text{Eriador's Hill-Huntington number} & HHN_R &= \text{Rohan's Hill-Huntington number} \\
 &= \frac{72.6^2}{3(3 + 1)} & &= \frac{33.7^2}{2(2 + 1)} \\
 &= \frac{72.6^2}{12} & &= \frac{33.7^2}{6} \\
 HHN_E &= 439.23 \text{ rounded to two decimal places} & HHN_R &= 189.28 \text{ rounded to two decimal places}
 \end{aligned}$$

The Hill-Huntington numbers for the three realms are summarized in Figure 6.68.

Realm (State)	Gondor	Eriador	Rohan
Hill-Huntington Number	298.94	439.23	189.28

FIGURE 6.68 Hill-Huntington numbers for each realm.

Because Eriador has the highest *HHN* (439.23 is greater than 298.23 or 189.28), Eriador should receive the new seat.

Why does this method work? Why does the state with the highest Hill-Huntington number deserve the additional seat? The answer lies in what is called *relative unfairness* and in the application of algebra. Consider the following scenario: A community college has a main campus and a satellite campus that is located several miles from the main campus. The main campus has an enrollment of 23,000 students with a faculty of 460; the satellite has an enrollment of 2,000 with a faculty of 40. If the college hires one new instructor, which campus is more deserving of receiving the new hire?

Using intuition, it might seem appropriate to calculate student-to-teacher ratios; that is, the campus that has more students per teacher would appear to be more deserving. However, in our scenario, both campuses have the same ratio, as shown in the following calculations:

<p>Main Campus</p> $ \begin{aligned} &\frac{\text{number of students}}{\text{number of teachers}} \\ &= \frac{23,000}{460} \\ &= 50 \text{ students per teacher} \end{aligned} $	<p>Satellite Campus</p> $ \begin{aligned} &\frac{\text{number of students}}{\text{number of teachers}} \\ &= \frac{2,000}{40} \\ &= 50 \text{ students per teacher} \end{aligned} $
---	--

To continue, we must investigate the meaning of *relative unfairness*.

Suppose two groups, *A* and *B*, are vying for an additional representative. Let p_A and p_B denote the populations of groups *A* and *B*, respectively, and let a and b denote the current number of representatives of groups *A* and *B*, respectively. The mean representations of *A* and *B* are given by $\frac{p_A}{a}$ and $\frac{p_B}{b}$, respectively (the student-to-teacher ratios shown above). Now, if *A* receives one new representative, its mean representation becomes $\frac{p_A}{a + 1}$, whereas if *B* receives the new representative, its mean representation becomes $\frac{p_B}{b + 1}$.

Our goal is to allocate the new representative in such a way as to minimize the relative unfairness of the allocation. Relative unfairness can be interpreted as the

difference (subtraction) in mean representations compared to (divided by) the new mean representation. So if A receives one new representative, the difference in mean representations is $\frac{p_B}{b} - \frac{p_A}{a+1}$, and the relative unfairness to A may be expressed as

$$\frac{\frac{p_B}{b} - \frac{p_A}{a+1}}{\frac{p_A}{a+1}}$$

In a similar fashion, if B receives the new representative, the relative unfairness to B may be expressed as

$$\frac{\frac{p_A}{a} - \frac{p_B}{b+1}}{\frac{p_B}{b+1}}$$

For the sake of argument, let us now suppose that the relative unfairness to A is smaller than the relative unfairness to B ; consequently, A deserves the new representative. We can express this analytically as follows:

Relative unfairness to $A <$ relative unfairness to B

$$\left[\frac{\frac{p_B}{b} - \frac{p_A}{a+1}}{\frac{p_A}{a+1}} \right] < \left[\frac{\frac{p_A}{a} - \frac{p_B}{b+1}}{\frac{p_B}{b+1}} \right]$$

Because $\frac{p_A}{a+1}$ and $\frac{p_B}{b+1}$ are both positive, we multiply each side of the inequality by the expression $\left(\frac{p_A}{a+1}\right)\left(\frac{p_B}{b+1}\right)$ and obtain the following:

$$\left(\frac{p_A}{a+1}\right)\left(\frac{p_B}{b+1}\right)\left[\frac{\frac{p_B}{b} - \frac{p_A}{a+1}}{\frac{p_A}{a+1}}\right] < \left(\frac{p_A}{a+1}\right)\left(\frac{p_B}{b+1}\right)\left[\frac{\frac{p_A}{a} - \frac{p_B}{b+1}}{\frac{p_B}{b+1}}\right]$$

multiplying each side by the expression $\left(\frac{p_A}{a+1}\right)\left(\frac{p_B}{b+1}\right)$

$$\left(\frac{p_B}{b+1}\right)\left[\frac{p_B}{b} - \frac{p_A}{a+1}\right] < \left(\frac{p_A}{a+1}\right)\left[\frac{p_A}{a} - \frac{p_B}{b+1}\right]$$

canceling $\frac{p_A}{a+1}$ on the left side and $\frac{p_B}{b+1}$ on the right side

$$\frac{p_B^2}{b(b+1)} - \frac{p_A p_B}{(a+1)(b+1)} < \frac{p_A^2}{a(a+1)} - \frac{p_A p_B}{(a+1)(b+1)} \quad \text{distributing}$$

$$\frac{p_B^2}{b(b+1)} < \frac{p_A^2}{a(a+1)} \quad \text{adding } \frac{p_A p_B}{(a+1)(b+1)} \text{ to each side}$$

$HHN_B < HHN_A$ definition of Hill-Huntington numbers

Therefore, if the relative unfairness to A is less than the relative unfairness to B , we see that $HHN_B < HHN_A$, and consequently, A deserves the new seat; that is, the group with the highest HHN should receive an additional seat. Regarding our scenario involving a community college and its satellite, we conclude that the main campus should receive the new instructor because $HHN_{main} > HHN_{satellite}$. (Verify this!)

We conclude this section with an example that recaps the five methods of apportionment.

EXAMPLE 7

SUMMARIZING ALL FIVE METHODS OF APPORTIONMENT Suppose that the governors of six New England region states have agreed to form an interstate bureau to foster historical awareness, tourism, and commerce. The bureau will have sixteen seats, and the populations of the states are given in Figure 6.69.

State	Connecticut (CT)	Maine (ME)	Massachusetts (MA)	New Hampshire (NH)	Rhode Island (RI)	Vermont (VT)
Population	3,405,584	1,274,923	6,349,097	1,235,786	1,048,319	608,827

FIGURE 6.69 New England region states. *Source:* Bureau of the Census, 2000.

- Use Hamilton’s Method to apportion the seats.
- Use Jefferson’s Method to apportion the seats.
- Use Adams’s Method to apportion the seats.
- Use Webster’s Method to apportion the seats.
- Use the Hill-Huntington Method to apportion the seats.

SOLUTION

To simplify calculations, we express the populations in millions, rounded to three decimal places, as shown in Figure 6.70.

State	CT	ME	MA	NH	RI	VT	Total
Population (millions)	3.406	1.275	6.349	1.236	1.048	0.609	13.923

FIGURE 6.70 New England region states.

- The standard divisor is $d = \text{total population} / \text{total number of seats} = 13.923 / 16 = 0.87$, rounded to two decimal places. Now divide each state’s population by 0.87 and round to two decimal places to obtain the standard quotas and lower quotas shown in Figure 6.71.

State	CT	ME	MA	NH	RI	VT	Total
Population (millions)	3.406	1.275	6.349	1.236	1.048	0.609	13.923
Standard Quota ($d = 0.87$)	3.91	1.47	7.30	1.42	1.20	0.70	16
Lower Quota	3	1	7	1	1	0	13

FIGURE 6.71 Standard and lower quotas of New England region states.

Finally, we allocate three additional seats to the three states (CT, VT, ME) that have the highest decimal parts (0.91, 0.70, 0.47) in their standard quotas and obtain the final apportionment shown in Figure 6.72.

State	CT	ME	MA	NH	RI	VT	Total
Lower Quota	3	1	7	1	1	0	13
Additional Seats	1	1	0	0	0	1	3
Hamilton’s Apportionment	4	2	7	1	1	1	16

FIGURE 6.72 Final apportionment using Hamilton’s Method.

- Using the modified divisor $d_m = 0.7$ (less than d), we obtain the modified quotas, lower modified quotas, and final apportionment shown in Figure 6.73.

State	CT	ME	MA	NH	RI	VT	Total
Population (millions)	3.406	1.275	6.349	1.236	1.048	0.609	13.923
Modified Quota ($d_m = 0.7$)	4.87	1.82	9.07	1.77	1.50	0.87	(not needed)
Lower Modified Quota	4	1	9	1	1	0	16
Jefferson's Apportionment	4	1	9	1	1	0	16

FIGURE 6.73 Final apportionment using Jefferson's Method.

c. Using the modified divisor $d_m = 1.1$ (greater than d), we obtain the modified quotas, upper modified quotas, and final apportionment shown in Figure 6.74.

State	CT	ME	MA	NH	RI	VT	Total
Population (millions)	3.406	1.275	6.349	1.236	1.048	0.609	13.923
Modified Quota ($d_m = 1.1$)	3.10	1.16	5.77	1.12	0.95	0.55	(not needed)
Upper Modified Quota	4	2	6	2	1	1	16
Adams' Apportionment	4	2	6	2	1	1	16

FIGURE 6.74 Final apportionment using Adams's Method.

d. Using the modified divisor $d_m = 0.85$ (different from d), we obtain the modified quotas, rounded modified quotas, and final apportionment shown in Figure 6.75.

State	CT	ME	MA	NH	RI	VT	Total
Population (millions)	3.406	1.275	6.349	1.236	1.048	0.609	13.923
Modified Quota ($d_m = 0.85$)	4.01	1.50	7.47	1.45	1.23	0.72	(not needed)
Rounded Modified Quota	4	2	7	1	1	1	16
Webster's Apportionment	4	2	7	1	1	1	16

FIGURE 6.75 Final apportionment using Webster's Method.

e. Using the modified divisor $d_m = 0.9$ (different from d), we obtain the modified quotas, lower modified quotas, geometric means, rounded quotas, and final apportionment shown in Figure 6.76.

State	CT	ME	MA	NH	RI	VT	Total
Population (millions)	3.406	1.275	6.349	1.236	1.048	0.609	13.923
Modified Quota ($d_m = 0.9$)	3.78	1.42	7.05	1.37	1.16	0.68	16
Lower Modified Quota	3	1	7	1	1	0	(not needed)
Geometric Mean	$\sqrt{3 \times 4}$ = 3.46	$\sqrt{1 \times 2}$ = 1.41	$\sqrt{7 \times 8}$ = 7.48	$\sqrt{1 \times 2}$ = 1.41	$\sqrt{1 \times 2}$ = 1.41	$\sqrt{0 \times 1}$ = 0	(not needed)
Rounded Quota	4	2	7	1	1	1	16
Hill-Huntington Apportionment	4	2	7	1	1	1	16

FIGURE 6.76 Final apportionment using the Hill-Huntington Method.



TOPIC X THE U.S. HOUSE OF REPRESENTATIVES APPORTIONMENT IN THE REAL WORLD

After the adoption of the Constitution, the United States of America took its first census in the year 1790. This census, along with Jefferson's method of apportionment, was used to distribute the 105 seats of the first House of Representatives. The population (in descending magnitude) and apportionment of each of the states are listed in Figure 6.77.

It could be argued that in 1790, Virginia was the state with the most power or influence because it had the greatest number of representatives (nineteen). Notice that on the basis of the first apportionment, New York and North Carolina had equal power (each had ten representatives), while North Carolina had twice as many representatives as New Jersey (five).

Over time, the relative rankings of the "power" of these states has dramatically changed. For instance, on the basis of the 2000 census, New York now has more than twice as many representatives as North Carolina has (twenty-nine compared to thirteen), while North Carolina and New Jersey are now equal (thirteen each). Go to www.census.gov/population/www/censusdata/apportionment/index.html to see the current apportionment of the House of Representatives.

As we have seen, different methods of apportionment can lead to different allocations of seats. In the following example and in the exercises, we will explore the distribution of seats of the first House of Representatives, using several different methods of apportionment.

State	Population	Representatives
Virginia	630,560	19
Massachusetts	475,327	14
Pennsylvania	432,879	13
North Carolina	353,523	10
New York	331,589	10
Maryland	278,514	8
Connecticut	236,841	7
South Carolina	206,236	6
New Jersey	179,570	5
New Hampshire	141,822	4
Vermont	85,533	2
Georgia	70,835	2
Kentucky	68,705	2
Rhode Island	68,446	2
Delaware	55,540	1
Total	3,615,920	105

FIGURE 6.77 The first census and apportionment of the United States (1790).

For comparison, we have included all five apportionments in Figure 6.78.

State	CT	ME	MA	NH	RI	VT	Total
Hamilton's Apportionment	4	2	7	1	1	1	16
Jefferson's Apportionment	4	1	9	1	1	0	16
Adams's Apportionment	4	2	6	2	1	1	16
Webster's Apportionment	4	2	6	2	1	1	16
Hill-Huntington Apportionment	4	2	7	1	1	1	16

FIGURE 6.78 Comparison of final apportionments.

EXAMPLE 8

VERIFYING THE APPORTIONMENT OF THE FIRST HOUSE OF REPRESENTATIVES OF THE UNITED STATES Use Jefferson's Method of apportionment and the population data in Figure 6.77 to verify the apportionment of the 105 seats in the 1790 House of Representatives (as shown in Figure 6.77).

SOLUTION

To simplify calculations, we express the populations in hundred thousands, rounded to two decimal places, as shown in Figure 6.79.

State	Population (hundred thousands)
Virginia	6.31
Massachusetts	4.75
Pennsylvania	4.33
North Carolina	3.54
New York	3.32
Maryland	2.79
Connecticut	2.37
South Carolina	2.06
New Jersey	1.80
New Hampshire	1.42
Vermont	0.86
Georgia	0.71
Kentucky	0.69
Rhode Island	0.68
Delaware	0.56
Total	36.19

FIGURE 6.79 The census of the first United States.

First, we must find the standard divisor:

$$d = \frac{\text{total population}}{\text{total number of seats}} = \frac{36.19}{105} = 0.3446666667$$

Therefore, $d = 0.34$.

Now, calculate each state's standard quota, that is, divide each population by $d = 0.34$, round to two decimal places, and then truncate to a whole number to obtain the lower quotas, all as shown in Figure 6.80.

Because the sum of the lower quotas (100) is less than the number of seats to be allocated (105), the standard quota creates five surplus seats. Therefore, to apply Jefferson's Method, we must find a modified divisor, $d_m < d = 0.34$ such that no surplus is created. After trial and error, we find that the modified divisor $d_m = 0.33$ will properly apportion the 105 seats, and the final results are shown in Figure 6.81.

State	Population (hundred thousands)	Standard Quota ($d = 0.34$)	Lower Quota
Virginia	6.31	$6.31/0.34 = 18.56$	18
Massachusetts	4.75	$4.75/0.34 = 13.97$	13
Pennsylvania	4.33	$4.33/0.34 = 12.74$	12
North Carolina	3.54	$3.54/0.34 = 10.41$	10
New York	3.32	$3.32/0.34 = 9.76$	9
Maryland	2.79	$2.79/0.34 = 8.21$	8
Connecticut	2.37	$2.37/0.34 = 6.97$	6
South Carolina	2.06	$2.06/0.34 = 6.06$	6
New Jersey	1.80	$1.80/0.34 = 5.29$	5
New Hampshire	1.42	$1.42/0.34 = 4.18$	4
Vermont	0.86	$0.86/0.34 = 2.53$	2
Georgia	0.71	$0.71/0.34 = 2.09$	2
Kentucky	0.69	$0.69/0.34 = 2.03$	2
Rhode Island	0.68	$0.68/0.34 = 2.00$	2
Delaware	0.56	$0.56/.34 = 1.65$	1
Total	36.19		100

FIGURE 6.80 Standard and lower quotas of the first states.

State	Population (hundred thousands)	Modified Quota ($d_m = 0.33$)	Lower Modified Quota
Virginia	6.31	$6.31/0.33 = 19.12$	19
Massachusetts	4.75	$4.75/0.33 = 14.39$	14
Pennsylvania	4.33	$4.33/0.33 = 13.12$	13
North Carolina	3.54	$3.54/0.33 = 10.73$	10
New York	3.32	$3.32/0.33 = 10.06$	10
Maryland	2.79	$2.79/0.33 = 8.45$	8
Connecticut	2.37	$2.37/0.33 = 7.18$	7
South Carolina	2.06	$2.06/0.33 = 6.24$	6
New Jersey	1.80	$1.80/0.33 = 5.45$	5
New Hampshire	1.42	$1.42/0.33 = 4.30$	4
Vermont	0.86	$0.86/0.33 = 2.61$	2
Georgia	0.71	$0.71/0.33 = 2.15$	2
Kentucky	0.69	$0.69/0.33 = 2.09$	2
Rhode Island	0.68	$0.68/0.33 = 2.06$	2
Delaware	0.56	$0.56/.33 = 1.70$	1
Total	36.19		105

FIGURE 6.81 Jefferson's Method applied to the first states.

6.2 EXERCISES

1. A small country consists of three states, *A*, *B*, and *C*; the population of each state is given in Figure 6.82. The country's legislature is to have forty seats.

State	A	B	C
Population	900,000	700,000	400,000

FIGURE 6.82 Table for Exercise 1.

- Express each state's population (and the total population) in terms of thousands.
 - Find the standard divisor.
 - Find each state's standard, lower, and upper quotas.
2. A small country consists of three states, *A*, *B*, and *C*; the population of each state is given in Figure 6.83. The country's legislature is to have fifty seats.

State	A	B	C
Population	810,000	720,000	510,000

FIGURE 6.83 Table for Exercise 2.

- Express each state's population (and the total population) in terms of thousands.
 - Find the standard divisor.
 - Find each state's standard, lower and upper quotas.
3. Suppose that the governors of three Middle Atlantic region states have agreed to form an interstate bureau to foster historical awareness, tourism, and commerce. The bureau will have fifteen seats, and the populations of the states are given in Figure 6.84.

State	New York	Pennsylvania	New Jersey
Population	18,976,821	12,281,054	8,414,347

FIGURE 6.84 Table for Exercise 3. *Source:* Bureau of the Census, 2000.

- Express each state's population (and the total population) in terms of millions, rounded to three decimal places.
 - Find the standard divisor.
 - Find each state's standard, lower, and upper quotas.
- ▶ 4. Suppose that the governors of three midwestern region states have agreed to form an interstate bureau to foster historical awareness, tourism, and commerce. The bureau will have ten seats, and the populations of the states are given in Figure 6.85.

State	Michigan	Wisconsin	Minnesota
Population	9,938,480	5,363,704	4,919,485

FIGURE 6.85 Table for Exercise 4. *Source:* Bureau of the Census, 2000.

- Express each state's population (and the total population) in terms of millions, rounded to three decimal places.
 - Find the standard divisor.
 - Find each state's standard, lower, and upper quotas.
5. Use Hamilton's Method to apportion the legislative seats in Exercise 1.
6. Use Hamilton's Method to apportion the legislative seats in Exercise 2.
7. Use Hamilton's Method to apportion the bureau seats in Exercise 3.
- ▶ 8. Use Hamilton's Method to apportion the bureau seats in Exercise 4.
- ▶ 9. A local school district contains four middle schools: Applewood, Boatwright, Castlerock, and Dunsmuir. The number of students attending each school is given in Figure 6.86. The school district has received a generous donation of 200 graphing calculators for student use. Use Hamilton's Method to determine how the calculators should be divided among the schools.

School	Applewood	Boatwright	Castlerock	Dunsmuir
Students	1,768	1,357	1,091	893

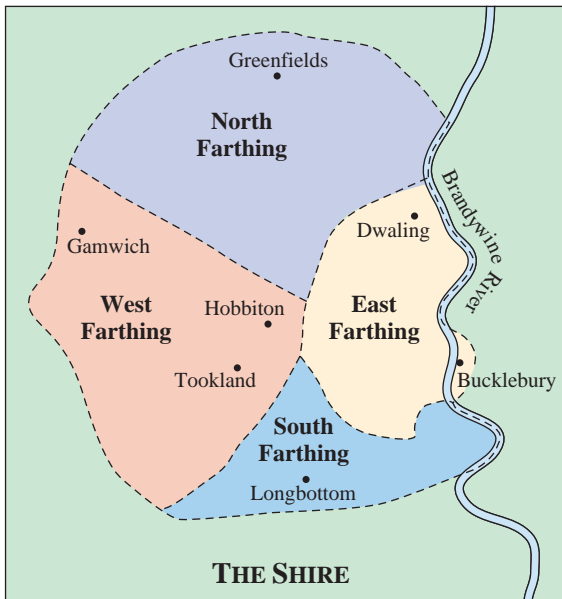
FIGURE 6.86 Table for Exercise 9.

10. A local school district contains four elementary schools: Elmhurst, Fernwood, Greenbriar, and Hawthorne. The number of students attending each school is given in Figure 6.87. The school district has received a generous donation of sixty digital cameras for student use. Use Hamilton's Method to determine how the cameras should be divided among the schools.

School	Elmhurst	Fernwood	Greenbriar	Hawthorne
Students	1,214	1,008	901	766

FIGURE 6.87 Table for Exercise 10.

11. In J.R.R. Tolkien's Middle Earth, a region known as the Shire is homeland to numerous clans of hobbits. The



Shire is divided into four regions: North Farthing, South Farthing, East Farthing, and West Farthing. Suppose the hobbits decide to form an association to foster the preservation of their cultural and culinary traditions. If the association is to have twenty-four seats, use Hamilton's Method and the (estimated) population data given in Figure 6.88 to apportion the seats.

Region	North Farthing	South Farthing	East Farthing	West Farthing
Population (estimated)	2,680	6,550	2,995	8,475

FIGURE 6.88 Table for Exercise 11.

12. In J.R.R. Tolkien's Middle Earth, the regions in and around the Misty Mountains are inhabited by elves, dwarves, and Ents (tree shepherds). Elves reside in Rivendell and Lothlórien, dwarves dwell in Moria, and Ents roam the forests of Fangorn. Suppose that these inhabitants decide to form a federation to foster mutual protection, historical preservation, and cultural exchange. If the federation is to have twenty-one seats, use Hamilton's Method and the (estimated) population data given in Figure 6.89 to apportion the seats.

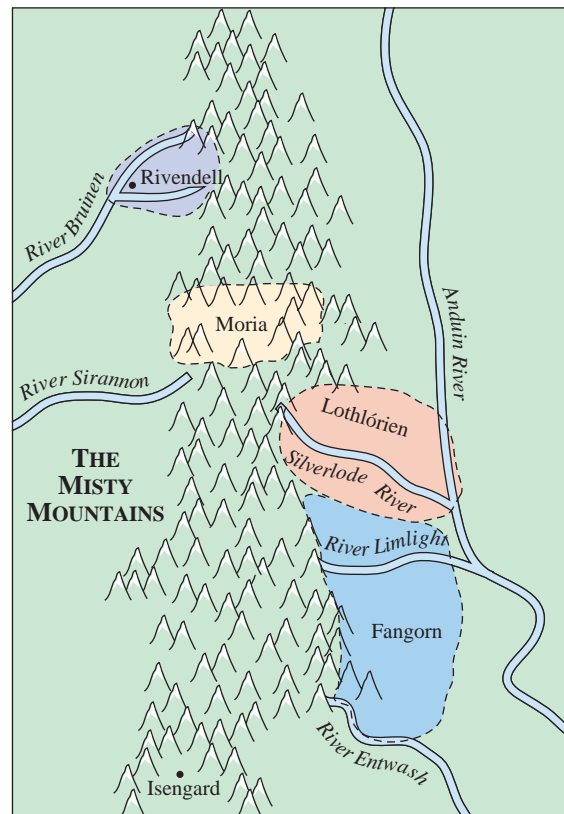
Region	Rivendell	Lothlórien	Moria	Fangorn
Population (estimated)	5,424	4,967	6,821	587

FIGURE 6.89 Table for Exercise 12.

13. Suppose that the governments of several Scandinavian countries have agreed to form an international bureau to

foster tourism, commerce, and cultural awareness. The bureau will have twenty seats, and the populations of the countries are given in Figure 6.90. Use Hamilton's Method to apportion the bureaucratic seats.

- ▶ 14. Suppose that the governments of several North African countries have agreed to form an international bureau to foster tourism, commerce, and education. The bureau will have twenty-five seats, and the populations of the countries are given in Figure 6.91. Use Hamilton's Method to apportion the bureau seats.
- 15. Suppose that the governments of several Central American countries have agreed to form an international bureau to foster tourism, commerce, and education. The bureau will have twenty-five seats, and the populations of the countries are given in Figure 6.92. Use Hamilton's Method to apportion the bureau seats.
- ▶ 16. Suppose that the governments of several Southeast Asian countries have agreed to form an international bureau to foster tourism, commerce, and education. The bureau will have thirty seats, and the populations of the countries are given in Figure 6.93. Use Hamilton's Method to apportion the bureau seats.
- ▶ 17.
 - a. Use Jefferson's Method to apportion the legislative seats in Exercise 5.
 - b. Use Adams's Method to apportion the legislative seats in Exercise 5.
 - c. Use Webster's Method to apportion the legislative seats in Exercise 5.



Country	Denmark	Finland	Iceland	Norway	Sweden
Population (thousands)	5,413	5,215	294	4,575	8,986

FIGURE 6.90 National populations of Scandinavia, 2004. *Source:* U.S. Department of Commerce.

Country	Algeria	Egypt	Libya	Morocco	Tunisia
Population (thousands)	32,129	76,117	5,632	32,209	9,975

FIGURE 6.91 National populations of North African countries, 2004. *Source:* U.S. Department of Commerce.

Country	Costa Rica	El Salvador	Guatemala	Honduras	Nicaragua	Panama
Population (thousands)	3,957	6,588	14,281	6,824	5,360	3,000

FIGURE 6.92 National populations of Central American countries, 2004. *Source:* U.S. Department of Commerce.

Country	Indonesia	Malaysia	Philippines	Taiwan	Thailand	Vietnam
Population (thousands)	238,453	23,522	86,242	22,750	64,866	82,690

FIGURE 6.93 National populations of Southeast Asian countries, 2004. *Source:* U.S. Department of Commerce.

18.
 - a. Use Jefferson's Method to apportion the legislative seats in Exercise 6.
 - b. Use Adams's Method to apportion the legislative seats in Exercise 6.
 - c. Use Webster's Method to apportion the legislative seats in Exercise 6.
19.
 - a. Use Jefferson's Method to apportion the bureau seats in Exercise 7.
 - b. Use Adams's Method to apportion the bureau seats in Exercise 7.
 - c. Use Webster's Method to apportion the bureau seats in Exercise 7.
- ▶ 20.
 - a. Use Jefferson's Method to apportion the bureau seats in Exercise 8.
 - b. Use Adams's Method to apportion the bureau seats in Exercise 8.
 - c. Use Webster's Method to apportion the bureau seats in Exercise 8.
21.
 - a. Use Jefferson's Method to determine how the calculators should be divided among the school in Exercise 9.
 - b. Use Adams's Method to determine how the calculators should be divided among the school in Exercise 9.
 - c. Use Webster's Method to determine how the calculators should be divided among the school in Exercise 9.
22.
 - a. Use Jefferson's Method to determine how the cameras should be divided among the school in Exercise 10.
 - b. Use Adams's Method to determine how the cameras should be divided among the schools in Exercise 10.
 - c. Use Webster's Method to determine how the cameras should be divided among the schools in Exercise 10.
23.
 - a. Use Jefferson's Method to apportion the associations seats in Exercise 11.
 - b. Use Adams's Method to apportion the association seats in Exercise 11.
 - c. Use Webster's Method to apportion the association seats in Exercise 11.
24.
 - a. Use Jefferson's Method to apportion the federation seats in Exercise 12.
 - b. Use Adams's Method to apportion the federation seats in Exercise 12.
 - c. Use Webster's Method to apportion the federation seats in Exercise 12.
25.
 - a. Use Jefferson's Method to apportion the bureau seats in Exercise 13.
 - b. Use Adams's Method to apportion the bureau seats in Exercise 13.
 - c. Use Webster's Method to apportion the bureau seats in Exercise 13.

- ▶ 26. a. Use Jefferson's Method to apportion the bureau seats in Exercise 14.
 b. Use Adams's Method to apportion the bureau seats in Exercise 14.
 c. Use Webster's Method to apportion the bureau seats in Exercise 14.
27. a. Use Jefferson's Method to apportion the bureau seats in Exercise 15.
 b. Use Adams's Method to apportion the bureau seats in Exercise 15.
 c. Use Webster's Method to apportion the bureau seats in Exercise 15.
- ▶ 28. a. Use Jefferson's Method to apportion the bureau seats in Exercise 16.
 b. Use Adams's Method to apportion the bureau seats in Exercise 16.
 c. Use Webster's Method to apportion the bureau seats in Exercise 16.
29. Use the Hill-Huntington Method to apportion the legislative seats in Exercise 5.
30. Use the Hill-Huntington Method to apportion the legislative seats in Exercise 6.
31. Use the Hill-Huntington Method to apportion the bureau seats in Exercise 7.
32. Use the Hill-Huntington Method to apportion the bureau seats in Exercise 8.
- ▶ 33. Use the Hill-Huntington Method to determine how the calculators should be divided among the school in Exercise 9.
34. Use the Hill-Huntington Method to determine how the cameras should be divided among the schools in Exercise 10.
35. Use the Hill-Huntington Method to apportion the association seats in Exercise 11.
36. Use the Hill-Huntington Method to apportion the federation seats in Exercise 12.
37. Use the Hill-Huntington Method to apportion the bureau seats in Exercise 13.
- ▶ 38. Use the Hill-Huntington Method to apportion the bureau seats in Exercise 14.
39. Use the Hill-Huntington Method to apportion the bureau seats in Exercise 15.
- ▶ 40. Use the Hill-Huntington Method to apportion the bureau seats in Exercise 16.
- ▶ 41. A community college has a main campus and a satellite campus that is located several miles away from the main campus. The main campus has an enrollment of 25,000 students with a faculty of 465; the satellite has an enrollment of 2,600 with a faculty of 47. The college decides to hire one new instructor. Use Hill-Huntington numbers to determine which location should receive the new instructor.

42. A waste management district operates two recycling locations: Alpha and Beta. Alpha services 20,000 households and has a staff of 23 workers; Beta's staff of 34 services 30,000 households. The district decides to hire one new worker. Use Hill-Huntington numbers to determine which location should receive the new staff member.
43. A school district has three elementary schools: Agnesi, Banach, and Cantor. If the district decides to hire one new teacher, use Hill-Huntington numbers to determine which school should receive the new instructor. The current enrollments and faculty of the schools are given in Figure 6.94.

	Agnesi	Banach	Cantor
Students	567	871	666
Teachers	21	32	25

FIGURE 6.94 Table for Exercise 43.

- ▶ 44. A school district has three middle schools: Descartes, Euclid, and Fermat. If the district decides to hire one new teacher, use Hill-Huntington numbers to determine which school should receive the new instructor. The current enrollments and faculty of the schools are given in Figure 6.95.

	Descartes	Euclid	Fermat
Students	752	984	883
Teachers	25	33	29

FIGURE 6.95 Table for Exercise 44.

- ▶ 45. Use Hamilton's Method of apportionment and the population data in Figure 6.77 to apportion the 105 seats in the 1790 House of Representatives. Does this hypothetical apportionment differ from the actual apportionment (shown in Figure 6.77)? If so, how?
46. Use Adams's Method of apportionment and the population data in Figure 6.77 to apportion the 105 seats in the 1790 House of Representatives. Does this hypothetical apportionment differ from the actual apportionment (shown in Figure 6.77)? If so, how?
47. Use Webster's Method of apportionment and the population data in Figure 6.77 to apportion the 105 seats in the 1790 House of Representatives. Does this hypothetical apportionment differ from the actual apportionment (shown in Figure 6.77)? If so, how?
48. Use the Hill-Huntington Method of apportionment and the population data in Figure 6.77 to apportion

the 105 seats in the 1790 House of Representatives. Does this hypothetical apportionment differ from the actual apportionment (shown in Figure 6.77)? If so, how?



Answer the following questions using complete sentences and your own words.

• CONCEPT QUESTIONS

49. What is apportionment?
50. What is a standard divisor?
51. What is a modified divisor?
52. What is a standard quota?
53. What is a modified quota?
54. What are lower and upper quotas?

• HISTORY QUESTIONS

55. What was the first method of apportionment of the House of Representatives to be approved by Congress?
56. What was the first method of apportionment of the House of Representatives to actually be used?
57. Why are the answers to Exercises 55 and 56 different?
58. Since the founding of the United States, what methods of apportionment of the House of Representatives have been used? What is their chronology?

59. What method of apportionment of the House of Representatives is currently being used?



60. In 1790, the United States had 3,615,920 people, and the House of Representatives had 105 seats. Therefore, on average, each seat represented approximately 34,437 people.
 - a. What was the population of the United States in 1850? How many seats did the House have? On average, how many people did each seat represent?
 - b. What was the population of the United States in 1900? How many seats did the House have? On average, how many people did each seat represent?
 - c. What was the population of the United States in 1950? How many seats did the House have? On average, how many people did each seat represent?
 - d. What was the population of the United States in 2000? How many seats did the House have? On average, how many people did each seat represent?
 - e. When you compare the answers to parts (a) through (d), what conclusion(s) can you make about the average representation in the House of Representatives?

Some useful links for this web project are listed on the text website: www.cengage.com/math/johnson

6.3 Flaws of Apportionment

OBJECTIVES

- Define and investigate the Quota Rule
- Define and investigate the Alabama Paradox
- Define and investigate the New States Paradox
- Define and investigate the Population Paradox

In the previous section, we studied several different methods of apportionment. Quite often, different methods of apportionment produce different allotments of the items being apportioned. Are the methods “fair”? Is one method “better” than the rest? We will attempt to answer these questions in this section.

The Quota Rule

The first step in any method of apportionment is the calculation of the standard divisor and the standard quota. That is,

$$d = \text{standard divisor} = \frac{\text{total population}}{\text{total number of seats}}$$

$$q = \text{standard quota} = \frac{\text{state's population}}{\text{standard divisor}} = \frac{\text{state's population}}{d}$$

Most people would agree that in any fair method of apportionment, each state should receive either its lower quota (truncated standard quota) of seats or its upper quota (rounded up standard quota) of seats. After all, if a state received too few seats, the state would appear to have been penalized, whereas if it received too many, favoritism could be alleged. Consequently, this fundamental criterion of fairness is referred to as the **Quota Rule**.

THE QUOTA RULE

The apportionment of a group should equal either the lower quota of the group or the upper quota of the group.

EXAMPLE 1

APPLYING JEFFERSON'S METHOD AND INVESTIGATING THE QUOTA RULE A small nation of 18 million people is composed of four states as shown in Figure 6.96. The national legislature has 120 seats.

State	A	B	C	D	Total
Population	1,548,000	2,776,000	3,929,000	9,747,000	18,000,000

FIGURE 6.96 National population figures.

- Use Jefferson's Method to apportion the 120 legislative seats.
- Compare the final apportionment with the lower and upper quotas of each state.

SOLUTION

- After each state's population is expressed in terms of millions, we find the standard divisor:

$$d = \text{standard divisor} = \frac{\text{total population}}{\text{total number of seats}} = \frac{18.0}{120} = 0.15$$

However, the standard divisor $d = 0.15$ results in two surplus seats, as shown in Figure 6.97.

State	A	B	C	D	Total
Population (millions)	1.548	2.776	3.929	9.747	18.000
Standard Quota ($d = 0.15$)	10.32	18.51	26.19	64.98	120
Lower Quota	10	18	26	64	118

FIGURE 6.97 Standard and lower quotas of the states.

Therefore, to apply Jefferson's Method, we must find a modified divisor, $d_m < d = 0.15$, such that no surplus is created. As a first guess, we try $d_m = 0.147$ and calculate each state's modified quota and lower modified quota, that is, we divide each state's population by $d_m = 0.147$, round to two decimal places, and then truncate (round down) to a whole number, as shown in Figure 6.98.

State	A	B	C	D	Total
Population (millions)	1.548	2.776	3.929	9.747	18.000
Standard Quota ($d = 0.15$)	10.32	18.51	26.19	64.98	120
Modified Quota ($d_m = 0.147$)	10.53	18.88	26.73	66.31	122.45
Lower Modified Quota	10	18	26	66	120

FIGURE 6.98 Modified and lower modified quotas using $d_m = 0.147$.

The modified divisor $d_m = 0.147$ has properly apportioned the 120 seats, and the final results are shown in Figure 6.99.

State	A	B	C	D	Total
Jefferson's Apportionment	10	18	26	66	120

FIGURE 6.99 Final apportionment using Jefferson's Method.

- b. The standard quota, lower quota, upper quota, and final apportionment for each state are shown in Figure 6.100.

State	A	B	C	D	Total
Standard Quota ($d = 0.15$)	10.32	18.51	26.19	64.98	120
Lower Quota	10	18	26	64	118
Upper Quota	11	19	27	65	122
Jefferson's Apportionment	10	18	26	66	120

FIGURE 6.100 Comparison of apportionment with lower and upper quotas.

Notice that the apportionments for A, B, and C are "fair" in that they equal either the lower quota or the upper quota of the state. However, the apportionment for D (66 seats) is "unfair" because it is greater than D's upper quota of 65.

Example 1 illustrates the fact that Jefferson's Method can violate the Quota Rule. In a similar fashion, it can be shown that Adams's Method, Webster's Method, and the Hill-Huntington Method can all produce apportionments that violate the Quota Rule. In contrast, Hamilton's Method always produces a fair apportionment with respect to the Quota Rule; it always allocates either the lower

quota or the upper quota. However, even though it satisfies the Quota Rule, Hamilton's Method gives rise to several other problems. In particular, we shall investigate the Alabama Paradox, the New States Paradox, and the Population Paradox.

The Alabama Paradox

In 1870, the population of the United States was 38,558,371, and the House of Representatives had 292 seats. Ten years later, the 1880 census recorded a population of 50,189,209—a 30% increase. Subsequently, the number of seats in the House of Representatives was to be increased. A discussion on whether to increase to 299 or 300 seats ensued. At the time, there were thirty-eight states in the Union, and Hamilton's Method was the official means of apportionment. The U.S. Census Office calculated the apportionment for each state using both 299 and 300 seats. It was then discovered that with 299 seats, Alabama would receive eight seats, whereas with 300, Alabama's share would be only seven. That is, an increase of one seat overall would result in a reduction in a state's apportionment. Consequently, this scenario has become known as the **Alabama Paradox**.

THE ALABAMA PARADOX

Adding one new seat to the total number of seats being allocated causes one of the states to lose one of its seats (even though the population has not changed).

EXAMPLE 2

APPLYING HAMILTON'S METHOD AND INVESTIGATING THE ALABAMA PARADOX A small nation of 12 million people is composed of three states as shown in Figure 6.101. The national legislature has 150 seats.

State	A	B	C	Total
Population (thousands)	595	5,615	5,790	12,000

FIGURE 6.101 National population figures.

- Use Hamilton's Method to apportion the 150 legislative seats.
- Suppose the total number of seats increases by one. Reapportion the 151 legislative seats.
- Compare the apportionments in parts (a) and (b).

SOLUTION

- First, we must find the standard divisor:

$$d = \text{standard divisor} = \frac{\text{total population}}{\text{total number of seats}} = \frac{12,000}{150} = 80$$

Therefore, $d = 80$.

Now calculate each state's standard quota (divide each state's population by $d = 80$) and lower quota (truncate the standard quotas to whole numbers), as shown in

Figure 6.102. By using Hamilton's Method, state A receives the one additional seat because the decimal part of state A's standard quota is highest (0.44); the final apportionment is shown in the last row of Figure 6.102.

State	A	B	C	Total
Population (thousands)	595	5,615	5,790	12,000
Standard Quota ($d = 80$)	7.44	70.19	72.38	150.01
Lower Quota	7	70	72	149
Additional Seats	1	0	0	1
Number of Seats	8	70	72	150

FIGURE 6.102 Hamilton's apportionment of 150 seats.

b. The new standard divisor is

$$d = \frac{\text{total population}}{\text{total number of seats}} = \frac{12,000}{151} = 79.47019868 \dots$$

Therefore, $d = 79.47$ (rounded to two decimal places).

Now calculate each state's standard quota (divide each state's population by $d = 79.47$) and lower quota (truncate the standard quotas to whole numbers), as shown in Figure 6.103. By using Hamilton's Method, state C and state B each receive an additional seat because the decimal parts of their standard quotas are highest (0.86 and 0.66); the final apportionment is shown in the last row of Figure 6.103.

State	A	B	C	Total
Population (thousands)	595	5,615	5,790	12,000
Standard Quota ($d = 79.47$)	7.49	70.66	72.86	151.01
Lower Quota	7	70	72	149
Additional Seats	0	1	1	2
Number of Seats	7	71	73	151

FIGURE 6.103 Hamilton's apportionment of 151 seats.

c. Figure 6.104 summarizes the two apportionments found in parts (a) and (b). Although the states' populations did not change, state A lost one of its seats when one new seat was added to the legislature; this is an illustration of the Alabama Paradox. Notice that the state that lost a seat is also the smallest state. This is typical of the so-called Alabama Paradox; large states benefit at the expense of small states.

State	Apportionment of 150 seats	Apportionment of 151 seats
A	8	7
B	70	71
C	72	73

FIGURE 6.104 Comparison of apportionments.

The New States Paradox

Oklahoma was admitted to the United States in 1907 as the forty-sixth state. On the basis of its populations, it was determined that the new state should have five seats in the House of Representatives. Therefore, the House was increased from 386 seats to 391 with the intention of not altering the apportionment of the other forty-five states. However, when Hamilton’s Method was applied to the new House of 391 seats, it was discovered that the apportionments for two other states would be affected; Maine would gain one seat at the expense of New York losing a seat! Therefore, after the 1910 census, Congress jettisoned Hamilton’s Method, and Webster’s Method was reinstated. Consequently, this scenario has become known as the **New States Paradox**.

THE NEW STATES PARADOX
 Adding a new state, and the corresponding number of seats based on its population, alters the apportionments for some of the other states.

EXAMPLE 3

APPLYING HAMILTON’S METHOD AND INVESTIGATING THE NEW STATES PARADOX A small nation of 10 million people is composed of two states, as shown in Figure 6.105. The national legislature has 100 seats.

State	A	B	Total
Population	3,848,000	6,152,000	10,000,000

FIGURE 6.105 National population figures.

- a. Use Hamilton’s Method to apportion the 100 legislative seats.
- b. Suppose state C has a population of 4,332,000 and joins the union. How many new seats should be added to the legislature?
- c. Apportion the new total number of seats.
- d. Compare the apportionments in parts (a) and (c).

SOLUTION

- a. After expressing each state’s population in terms of thousands (an arbitrary decision), we must find the standard divisor:

$$d = \text{standard divisor} = \frac{\text{total population}}{\text{total number of seats}} = \frac{10,000}{100} = 100$$

Therefore, $d = 100$.

Now calculate each state’s standard quota (divide each state’s population by $d = 100$) and lower quota (truncate the standard quotas to whole numbers), as shown in Figure 6.106. By using Hamilton’s Method, state B receives the one additional seat

State	A	B	Total
Population (thousands)	3,848	6,152	10,000
Standard Quota ($d = 100$)	38.48	61.52	100
Lower Quota	38	61	99
Additional Seats	0	1	1
Number of Seats	38	62	100

FIGURE 6.106 Hamilton’s apportionment of the original 100 seats.

because the decimal part of state B's standard quota is highest (0.52); the final apportionment is shown in the last row of Figure 6.106.

- b. In terms of thousands, state C's population of 4,332,000 would be expressed as 4,332 thousand. Therefore, state C's quota of (new) seats is $\frac{4,332}{100} = 43.32$, or 43 seats. That is, when state C is admitted to the union, forty-three seats should be added to the legislature; the total number of seats to be allocated is now $100 + 43 = 143$ seats.
- c. We must now find the new standard divisor:

$$\begin{aligned} d = \text{standard divisor} &= \frac{\text{total population}}{\text{total number of seats}} = \frac{10,000 + 4,332}{100 + 43} \\ &= \frac{14,332}{143} = 100.2237762 \dots \end{aligned}$$

Therefore, $d = 100.22$.

Now calculate each state's standard quota (divide each population by $d = 100.22$) and lower quota (truncate the standard quotas to whole numbers), as shown in Figure 6.107. By using Hamilton's Method, state A receives the one additional seat because the decimal part of state A's standard quota is highest (0.40); the final apportionment is shown in the last row of Figure 6.107.

State	A	B	C	Total
Population (thousands)	3,848	6,152	4,332	14,332
Standard Quota ($d = 100.22$)	38.40	61.38	43.22	143
Lower Quota	38	61	43	142
Additional Seat	1	0	0	1
Number of Seats	39	61	43	143

FIGURE 6.107 Hamilton's apportionment of the new 143 seats.

- d. Figure 6.108 summarizes the apportionments from parts (a) and (c). On comparison, we see that when state C joined the union, state A gained 1 seat and state B lost one seat. That is, the addition of state C altered the apportionments for the other states; this is an illustration of the New States Paradox.

State	A	B	C	Total
(Original) Number of Seats	38	62	0	100
(New) Number of Seats	39	61	43	143

FIGURE 6.108 Comparison of apportionments.

The Population Paradox

Throughout history, the populations of the various states have been in flux; changes in population subsequently affect each state's apportionment of seats in the House of Representatives. Hypothetically, if one state grew at a fast rate while another grew at a slow rate, it might seem more appropriate that the state with the larger growth rate be entitled to an "extra" seat, if any became available. However, Hamilton's Method does not adhere to this premise. During the late 1800s and

early 1900s, Virginia was growing at rates much faster than Maine; for example, from 1890 to 1910, Virginia's population increased by 24.5%, while Maine's growth was only 12.3%. To everyone's surprise, Virginia lost one of its seats to Maine in the early 1900s. Consequently, this scenario has become known as the **Population Paradox**.

THE POPULATION PARADOX

One state loses a seat to another state even though the first state is growing at a faster rate; state A loses a seat to state B even though state A is growing at a faster rate than state B.

EXAMPLE 4

APPLYING HAMILTON'S METHOD AND INVESTIGATING THE POPULATION PARADOX A school district of slightly more than 6,000 students is composed of three campuses, as shown in Figure 6.109. The district has twelve educational specialists to assign to three locations.

Campus	A	B	C	Total
Enrollment (2009)	777	1,792	3,542	6,111
Enrollment (2010)	820	1,880	3,645	6,345

FIGURE 6.109 School district enrollments.

- Use Hamilton's Method and the 2009 enrollments to apportion the twelve specialists.
- Use Hamilton's Method and the 2010 enrollments to reapportion the twelve specialists.
- Find the differences, if any, in the number of specialists apportioned to each campus in 2009 versus 2010.
- Which campus had the highest growth rate (percentagewise) from 2009 to 2010? Do the growth rates of the campuses conform with the differences found in part (c)?

SOLUTION

- First, we must find the standard divisor for 2009:

$$d = \text{standard divisor} = \frac{\text{total enrollment}}{\text{total number of specialists}} = \frac{6,111}{12} = 509.25$$

Therefore, $d = 509.25$.

Now calculate each campus's standard quota (divide each enrollment by $d = 509.25$) and lower quota (truncate the standard quotas to whole numbers), as shown in Figure 6.110. By using Hamilton's Method, campus C and campus A each receive one additional specialist because the decimal parts of their standard quotas are highest (0.96 and 0.53); the 2009 apportionment is shown in the last row of Figure 6.110.

Campus	A	B	C	Total
Enrollment	777	1,792	3,542	6,111
Standard Quota ($d = 509.25$)	1.53	3.52	6.96	12.01
Lower Quota	1	3	6	10
Additional Specialists	1	0	1	2
Number of Specialists (in 2009)	2	3	7	12

FIGURE 6.110 Hamilton's apportionment in 2009.

- b. Now we find the standard divisor for 2010:

$$d = \text{standard divisor} = \frac{\text{total enrollment}}{\text{total number of specialists}} = \frac{6,345}{12} = 528.75$$

Therefore, $d = 528.75$.

Now calculate each campus's standard quota (divide each enrollment by $d = 528.75$) and lower quota (truncate the standard quotas to whole numbers), as shown in Figure 6.111. By using Hamilton's Method, campus C and campus B each receive one additional specialist because the decimal parts of their standard quotas are highest (0.89 and 0.56); the 2010 apportionment is shown in the last row of Figure 6.111.

Campus	A	B	C	Total
Enrollment	820	1,880	3,645	6,345
Standard Quota ($d = 509.25$)	1.55	3.56	6.89	12
Lower Quota	1	3	6	10
Additional Specialists	0	1	1	2
Number of Specialists (in 2010)	1	4	7	12

FIGURE 6.111 Hamilton's apportionment in 2010.

- c. Comparing the number of specialists apportioned in 2009 and 2010 (Figure 6.112), we see that campus A lost one specialist, campus B gained one specialist, and the apportionment of campus C was unchanged.

Campus	A	B	C
Number of Specialists (in 2009)	2	3	7
Number of Specialists (in 2010)	1	4	7
Difference	-1	+1	0

FIGURE 6.112 Comparison of apportionment in 2009 and 2010.

- d. To calculate the growth rate of each campus (expressed as a percentage), we must find the difference in the enrollment (2010 minus 2009), divide the difference by the original (2009) enrollment, and then multiply by 100. These calculations are summarized in Figure 6.113.

Campus	A	B	C
2009 Enrollment	777	1,792	3,542
2010 Enrollment	820	1,880	3,645
Difference (increase)	$820 - 777 = 43$	$1,880 - 1,792 = 88$	$3,645 - 3,542 = 103$
Proportion	$\frac{43}{777}$ $= 0.05534 \dots$	$\frac{88}{1,792}$ $= 0.04910 \dots$	$\frac{103}{3,542}$ $= 0.02907 \dots$
Percentage growth	5.5%	4.9%	2.9%

FIGURE 6.113 Growth rates from 2009 and 2010.

Consequently, campus A had the highest growth rate (5.5%) from 2009 to 2010. However, the growth rates do not conform with the differences in apportionment of specialists (Figure 6.114).

Campus	A	B	C
Enrollment growth rate	5.5%	4.9%	2.9%
Change in apportionment	lost 1 specialist	gained 1 specialist	no change

FIGURE 6.114 Comparison of growth rates and changes in apportionment.

Notice that campus A grew at a faster rate than campus B ($5.5\% > 4.9\%$), but campus A lost one of its specialists to campus B! This is an illustration of the Population Paradox; one group loses an apportioned position to another group even though the first group is growing at a faster rate.

The Balinski-Young Impossibility Theorem

We have studied several methods of apportionment, specifically the methods of Hamilton, Jefferson, Adams, Webster, and Hill-Huntington. We have also seen various problems associated with these methods of apportionment. Although Hamilton's Method does not violate the Quota Rule, it has several flaws in that it gives rise to the Alabama, New States, and Population Paradoxes. In contrast, it can be shown that the methods of Jefferson, Adams, Webster, and Hill-Huntington are all immune to these paradoxes; however, they may all violate the Quota Rule. In addition, some methods favor large states, whereas other methods favor small states. In other words, if a method of apportionment does not have one type of problem, it has another type of problem. In the quest for the ideal method of apportionment, can we ever hope to find a method that will satisfy the Quota Rule, be immune from paradoxes, and favor neither large nor small states? Unfortunately, the answer is "no." In 1980, mathematicians Michel Balinski and H. Peyton Young proved that it is mathematically impossible for any method of apportionment to satisfy the Quota Rule and not produce any paradoxes. This is known as the **Balinski-Young Impossibility Theorem**.

THE BALINSKI-YOUNG IMPOSSIBILITY THEOREM

There is no perfect method of apportionment: If a method satisfies the Quota Rule, it must give rise to paradoxes; and if a method does not give rise to paradoxes, it must violate the Quota Rule.

Just as Arrow's Impossibility Theorem (Section 6.1) ruled out the possibility of the existence of a perfect system of voting, the Balinski-Young Impossibility Theorem rules out the possibility of creating a perfect method of apportionment.

Figure 6.115 lists several methods of apportionment and summarizes the flaws discussed in this section.

Flaw	Method of Apportionment				
	Hamilton	Jefferson	Adams	Webster	Hill-Huntington
May violate the Quota Rule.	No	Yes	Yes	Yes	Yes
May produce the Alabama Paradox.	Yes	No	No	No	No
May produce the New States Paradox.	Yes	No	No	No	No
May produce the Population Paradox.	Yes	No	No	No	No

FIGURE 6.115 Flaws of the methods of apportionment.

6.3 EXERCISES

1. A small country consists of three states: A, B, and C. The population of each state is given in Figure 6.116. The country's legislature is to have thirty-two seats.

State	A	B	C
Population	3,500,000	4,200,000	16,800,000

FIGURE 6.116 Table for Exercise 1.

- Express each state's population (and the total population) in terms of millions.
 - Find the standard divisor, and round it off to two decimal places.
 - Find each state's standard, lower, and upper quotas.
 - Use Jefferson's Method to apportion the thirty-two seats.
 - Is the Quota Rule violated? Explain.
2. A small country consists of three states, A, B, C. The population of each state is given in Figure 6.117. The country's legislature is to have ninety-six seats.

State	A	B	C
Population	1,200,000	3,400,000	15,400,000

FIGURE 6.117 Table for Exercise 2.

- Express each state's population (and the total population) in terms of millions.
- Find the standard divisor, and round it off to two decimal places.

- Find each state's standard, lower, and upper quotas.
- Use Jefferson's Method to apportion the ninety-six seats.
- Is the Quota Rule violated? Explain.

3. A small country consists of three states: A, B, and C. The population of each state is given in Figure 6.118. The country's legislature is to have seventy-nine seats.

State	A	B	C
Population	3,500,000	4,200,000	16,800,000

FIGURE 6.118 Table for Exercise 3.

- Express each state's population (and the total population) in terms of millions.
 - Find the standard divisor, and round it off to two decimal places.
 - Find each state's standard, lower, and upper quotas.
 - Use Adams's Method to apportion the seventy-nine seats.
 - Is the Quota Rule violated? Explain.
4. A small country consists of three states: A, B, and C. The population of each state is given in Figure 6.119. The country's legislature is to have seventy-three seats.

State	A	B	C
Population	1,200,000	3,400,000	15,400,000

FIGURE 6.119 Table for Exercise 4.

- a. Express each state's population (and the total population) in terms of millions.
 - b. Find the standard divisor, and round it off to two decimal places.
 - c. Find each state's standard, lower, and upper quotas.
 - d. Use Adams's Method to apportion the seventy-three seats.
 - e. Is the Quota Rule violated? Explain.
5. A small country consists of four states: A, B, C, and D. The population of each state is given in Figure 6.120. The country's legislature is to have 200 seats.

State	A	B	C	D
Population	1,200,000	3,400,000	17,500,000	19,400,000

FIGURE 6.120 Table for Exercise 5.

- a. Express each state's population (and the total population) in terms of millions.
 - b. Find the standard divisor, and round it off to two decimal places.
 - c. Find each state's standard, lower, and upper quotas.
 - d. Use Webster's Method to apportion the 200 seats.
 - e. Is the Quota Rule violated? Explain.
6. A small country consists of four states: A, B, C, and D. The population of each state is given in Figure 6.121. The country's legislature is to have 177 seats.

State	A	B	C	D
Population	1,100,000	3,200,000	17,300,000	19,400,000

FIGURE 6.121 Table for Exercise 6.

- a. Express each state's population (and the total population) in terms of millions.
 - b. Find the standard divisor, and round it off to two decimal places.
 - c. Find each state's standard, lower, and upper quotas.
 - d. Use Webster's Method to apportion the 177 seats.
 - e. Is the Quota Rule violated? Explain.
- ▶ 7. A small country consists of four states: A, B, C, and D. The population of each state is given in Figure 6.122. The country's legislature is to have 200 seats.

State	A	B	C	D
Population	1,200,000	3,400,000	17,500,000	19,400,000

FIGURE 6.122 Table for Exercise 7.

- a. Express each state's population (and the total population) in terms of millions.
 - b. Find the standard divisor, and round it off to two decimal places.
 - c. Find each state's standard, lower, and upper quotas.
 - d. Use the Hill-Huntington Method to apportion the 200 seats.
 - e. Is the Quota Rule violated? Explain.
8. A small country consists of four states: A, B, C, and D. The population of each state is given in Figure 6.123. The country's legislature is to have 177 seats.

State	A	B	C	D
Population	1,100,000	3,200,000	17,300,000	19,400,000

FIGURE 6.123 Table for Exercise 8.

- a. Express each state's population (and the total population) in terms of millions.
 - b. Find the standard divisor and round it off to two decimal places.
 - c. Find each state's standard, lower, and upper quotas.
 - d. Use the Hill-Huntington Method to apportion the 177 seats.
 - e. Is the Quota Rule violated? Explain.
9. A small country consists of three states: A, B, and C. The population of each state is given in Figure 6.124. The country's legislature is to have 120 seats.

State	A	B	C
Population	690,000	5,700,000	6,410,000

FIGURE 6.124 Table for Exercise 9.

- a. Express each state's population (and the total population) in terms of thousands.
 - b. Find the standard divisor and round it off to two decimal places.
 - c. Use Hamilton's Method to apportion the 120 legislative seats.
 - d. Suppose the total number of seats increases by one. Reapportion the 121 legislative seats.
 - e. Is the Alabama Paradox exhibited? Explain.
- ▶ 10. A small country consists of three states: A, B, C. The population of each state is given in Figure 6.125. The country's legislature is to have 140 seats.

State	A	B	C
Population	700,000	5,790,000	6,510,000

FIGURE 6.125 Table for Exercise 10.

- a. Express each state's population (and the total population) in terms of thousands.
- b. Find the standard divisor and round it off to two decimal places.
- c. Use Hamilton's Method to apportion the 140 legislative seats.
- d. Suppose the total number of seats increases by one. Reapportion the 141 legislative seats.
- e. Is the Alabama Paradox exhibited? Explain.
- ▶ 11. A small country consists of three states: A, B, and C. The population of each state is given in Figure 6.126. The country's legislature is to have 110 seats.

State	A	B	C
Population	1,056,000	1,844,000	2,100,000

FIGURE 6.126 Table for Exercise 11.

- a. Express each state's population (and the total population) in terms of thousands.
- b. Find the standard divisor and round it off to two decimal places.
- c. Use Hamilton's Method to apportion the 110 legislative seats.
- d. Suppose state D has a population of 2,440,000 and joins the union. How many new seats should be added to the legislature?
- e. Apportion the new total number of seats.
- f. Is the New States Paradox exhibited? Explain
12. A small country consists of three states: A, B, and C. The population of each state is given in Figure 6.127. The country's legislature is to have 150 seats.

State	A	B	C
Population	2,056,000	1,844,000	3,100,000

FIGURE 6.127 Table for Exercise 12.

- a. Express each state's population (and the total population) in terms of thousands.
- b. Find the standard divisor and round it off to two decimal places.
- c. Use Hamilton's Method to apportion the 150 legislative seats.
- d. Suppose state D has a population of 2,500,000 and joins the union. How many new seats should be added to the legislature?
- e. Apportion the new total number of seats.
- f. Is the New States Paradox exhibited? Explain.

13. A school district has three campuses with enrollments given in Figure 6.128. The district has eleven educational specialists to assign to the locations.

Campus	A	B	C
Enrollment (2005)	780	1,700	3,520
Enrollment (2006)	820	1,880	3,740

FIGURE 6.128 Table for Exercise 13.

- a. Use Hamilton's Method and the 2005 enrollments to apportion the eleven specialists.
- b. Use Hamilton's Method and the 2006 enrollments to reapportion the eleven specialists.
- c. Find the growth rate (from 2005 to 2006) for each campus.
- d. Is the Population Paradox exhibited? Explain.
- ▶ 14. A school district has three campuses with enrollments given in Figure 6.129. The district has twelve educational specialists to assign to the locations.

Campus	A	B	C
Enrollment (2005)	750	1,680	3,980
Enrollment (2006)	788	1,858	4,210

FIGURE 6.129 Table for Exercise 14.

- a. Use Hamilton's Method and the 2005 enrollments to apportion the twelve specialists.
- b. Use Hamilton's Method and the 2006 enrollments to reapportion the twelve specialists.
- c. Find the growth rate (from 2005 to 2006) for each campus.
- d. Is the Population Paradox exhibited? Explain.



Answer the following questions using complete sentences and your own words.

• CONCEPT QUESTIONS

15. What is the Quota Rule? Explain its meaning.
16. What is the Alabama Paradox? Explain its meaning.
17. What is the New States Paradox? Explain its meaning.
18. What is the Population Paradox? Explain its meaning.
19. What is the Balinski-Young Impossibility Theorem? Explain its meaning.

6

CHAPTER REVIEW

TERMS

Adams's Method
Alabama Paradox
apportion
arithmetic mean
Arrow's Impossibility Theorem
Balinski-Young Impossibility Theorem
Borda count method

geometric mean
Hamilton's Method
head-to-head criterion of fairness
Hill-Huntington Method
Hill-Huntington number
irrelevant alternatives criterion of fairness
Jefferson's Method
lower quota
majority

majority criterion of fairness
modified divisor
modified quota
monotonicity criterion of fairness
New States Paradox
pairwise comparison method
plurality method of voting

plurality with elimination method
Population Paradox
ranked-choice or instant runoff method
Quota Rule
standard divisor
standard quota
upper quota
voter preference table
Webster's Method

REVIEW EXERCISES

1. The Metropolitan Symphonic Orchestra has volunteered to play a benefit concert, with proceeds to support music programs at local high schools. The organizers need to know what music will be featured at the concert. The conductor has asked each member of the orchestra to rank the following composers: Beethoven (B), Mozart (M), Tchaikovsky (T), and Vivaldi (V). The ranked ballots have been tallied, and the results are summarized in Figure 6.130.

- How many votes were cast?
- Use the plurality method of voting to determine the winner.
- What percent of the votes did the winner in part (b) receive?
- Use the instant runoff method to determine the winner.
- What percent of the votes did the winner in part (d) receive?

	Number of Ballots Cast						
	5	7	6	12	13	17	14
1st choice	M	M	M	B	B	V	T
2nd choice	B	V	T	T	M	M	V
3rd choice	V	B	V	V	T	B	B
4th choice	T	T	B	M	V	T	M

FIGURE 6.130 Voter preference table for Exercise 1.

- Use the Borda count method to determine the winner.
 - How many points did the winner in part (f) receive?
 - Use the pairwise comparison method to determine the winner.
 - How many points did the winner in part (h) receive?
2. A small country consists of three states: A, B, and C. The population of each state is given in Figure 6.131. The country's legislature is to have seventy-six seats.

State	A	B	C
Population	1,200,000	2,300,000	2,500,000

FIGURE 6.131 Table for Exercise 2.

- Express each state's population (and the total population) in terms of thousands.

- b. Find the standard divisor and round it off to two decimal places.
 - c. Find each state's standard, lower, and upper quotas.
3. Use Hamilton's Method to apportion the legislative seats in Exercise 2.
 4. Use Jefferson's Method to apportion the legislative seats in Exercise 2.
 5. Use Adams's Method to apportion the legislative seats in Exercise 2.
 6. Use Webster's Method to apportion the legislative seats in Exercise 2.
 7. Use the Hill-Huntington Method to apportion the legislative seats in Exercise 2.
 8. Suppose that the governments of several South American countries have agreed to form an international bureau to foster tourism, commerce, and education. The bureau will have twenty-five seats, and the populations of the countries are given in Figure 6.132. Use Hamilton's Method to apportion the bureau seats.

Country	Argentina	Bolivia	Chile	Paraguay	Uruguay
Population (thousands)	39,145	8,724	15,824	6,191	3,399

FIGURE 6.132 National population figures, 2004. *Source:* U.S. Department of Commerce.

9. Use Jefferson's Method to apportion the bureau seats in Exercise 8.
10. Use Adams's Method to apportion the bureau seats in Exercise 8.
11. Use Webster's Method to apportion the bureau seats in Exercise 8.
12. Use the Hill-Huntington Method to apportion the bureau seats in Exercise 8.
13. A school district has three high schools: Leibniz, Maclaurin, and Napier. If the district decides to hire one new teacher, use Hill-Huntington numbers to determine which school gets the new instructor. The current enrollments and faculty of the schools are given in Figure 6.133.

	Leibniz	Maclaurin	Napier
Students	987	1,242	1,763
Teachers	25	31	44

FIGURE 6.133 Table for Exercise 13.

14. A small country consists of three states: A, B, and C. The population of each state is given in Figure 6.134. The country's legislature is to have thirty-one seats.

State	A	B	C
Population	3,500,000	4,100,000	16,800,000

FIGURE 6.134 Table for Exercise 14.

- a. Express each state's population (and the total population) in terms of millions.
 - b. Find the standard divisor, and round it off to two decimal places.
 - c. Find each state's standard, lower, and upper quotas.
 - d. Use Jefferson's Method to apportion the thirty-one seats.
 - e. Is the Quota Rule violated? Explain.
15. A small country consists of three states: A, B, and C. The population of each state is given in Figure 6.135. The country's legislature is to have seventy-two seats.

State	A	B	C
Population	1,600,000	3,500,000	15,300,000

FIGURE 6.135 Table for Exercise 15.

- a. Express each state's population (and the total population) in terms of millions.
 - b. Find the standard divisor, and round it off to two decimal places.
 - c. Find each state's standard, lower, and upper quotas.
 - d. Use Adams's Method to apportion the seventy-two seats.
 - e. Is the Quota Rule violated? Explain.
16. A small country consists of four states: A, B, C, and D. The population of each state is given in Figure 6.136. The country's legislature is to have 201 seats.

State	A	B	C	D
Population	1,350,000	3,430,000	17,600,000	19,650,000

FIGURE 6.136 Figure for Exercise 16.

- a. Express each state's population (and the total population) in terms of millions.
 - b. Find the standard divisor, and round it off to two decimal places.
 - c. Find each state's standard, lower, and upper quotas.
 - d. Use Webster's Method to apportion the 201 seats.
 - e. Is the Quota Rule violated? Explain.
17. A small country consists of four states: A, B, C, and D. The population of each state is given in Figure 6.137. The country's legislature is to have 201 seats.

State	A	B	C	D
Population	1,100,000	3,500,000	17,600,000	19,400,000

FIGURE 6.137 Figure for Exercise 17.

- Express each state's population (and the total population) in terms of millions.
 - Find the standard divisor, and round it off to two decimal places.
 - Find each state's standard, lower, and upper quotas.
 - Use the Hill-Huntington Method to apportion the 201 seats.
 - Is the Quota Rule violated? Explain.
18. A small country consists of three states: A, B, and C. The population of each state is given in Figure 6.138. The country's legislature is to have 140 seats.

State	A	B	C
Population	700,000	5,790,000	6,520,000

FIGURE 6.138 Figure for Exercise 18.

- Express each state's population (and the total population) in terms of thousands.
 - Find the standard divisor and round it off to two decimal places.
 - Use Hamilton's Method to apportion the 140 legislative seats.
 - Suppose the total number of seats increases by one. Reapportion the 141 legislative seats.
 - Is the Alabama Paradox exhibited? Explain.
19. A small country consists of three states: A, B, and C. The population of each state is given in Figure 6.139. The country's legislature is to have 110 seats.

State	A	B	C
Population	1,057,000	1,942,000	2,001,000

FIGURE 6.139 Figure for Exercise 19.

- Express each state's population (and the total population) in terms of thousands.
- Find the standard divisor and round it off to two decimal places.
- Use Hamilton's Method to apportion the 110 legislative seats.
- Suppose state D has a population of 2,450,000 and joins the union. How many new seats should be added to the legislature?

- Apportion the new total number of seats.
 - Is the New States Paradox exhibited? Explain.
20. A school district has three campuses with enrollments given in Figure 6.140. The district has twelve educational specialists to assign to the locations.

Campus	A	B	C
Enrollment (2005)	760	1,670	3,980
Enrollment (2006)	788	1,858	4,210

FIGURE 6.140 Figure for Exercise 20.

- Use Hamilton's Method and the 2005 enrollments to apportion the twelve specialists.
- Use Hamilton's Method and the 2006 enrollments to reapportion the twelve specialists.
- Find the growth rate (from 2005 to 2006) for each campus.
- Is the Population Paradox exhibited? Explain.



Answer the following questions using complete sentences and your own words.

• CONCEPT QUESTIONS

- What is a majority?
- What is a ranked ballot?
- What is a voter preference table?
- What is a standard divisor?
- What is a standard quota?
- What is apportionment?
- What is the Quota Rule? Explain its meaning.
- What is the Alabama Paradox? Explain its meaning.
- What is the New States Paradox? Explain its meaning.
- What is the Population Paradox? Explain its meaning.
- What is the Balinski-Young Impossibility Theorem? Explain its meaning.
- What are the four fairness criteria? Explain their meanings.
- What is Arrow's Impossibility Theorem?

• HISTORY QUESTIONS

- What methods of apportionment for the House of Representatives have actually been used? What is their chronology?
- What method of apportionment for the House of Representatives is currently being used?