McGraw Hill ALEKS®

Essential Statistics, 3e

Navidi | Monk

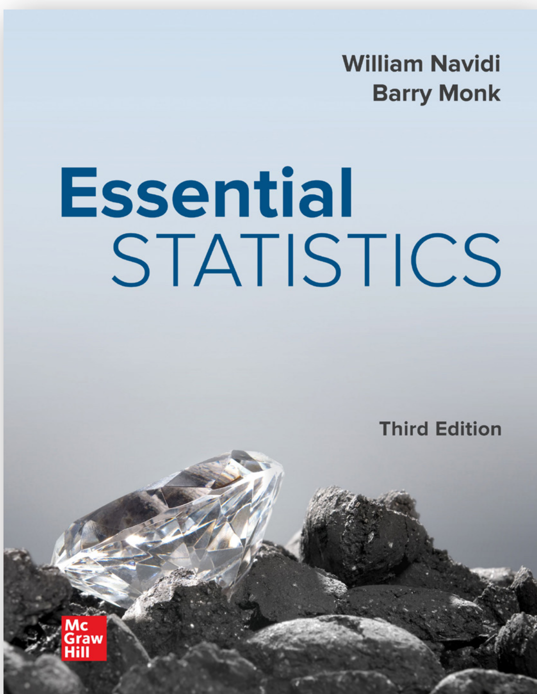


Table of Contents

Essential STATISTICS

Third Edition

William Navidi
Colorado School of Mines

Barry Monk
Middle Georgia State University

McGraw Hill

ESSENTIAL STATISTICS, THIRD EDITION

This book is printed on acid-free paper.

1 2 3 4 5 6 7 8 9 LWI 26 25 24 23 22 21

ISBN 978-1-260-35979-4 (bound edition)
MHID 1-260-35979-4 (bound edition)

ISBN 978-1-260-49216-3 (loose-leaf edition)
MHID 1-260-49216-8 (loose-leaf edition)

ISBN 978-1-264-57183-3 (instructor's edition)
MHID 1-264-57183-6 (instructor's edition)

Portfolio Manager: *Chrissy Skogen*
Product Developer: *Megan Platt*
Marketing Manager: *Emily Digiovanna*
Content Project Managers: *Jeni McAtee, Rachael Hillebrand*
Buyer: *Laura Fuller*
Designer: *Beth Blech*
Content Licensing Specialist: *Jacob Sullivan*
Cover Image: *Joe Belanger/Alamy*
Compositor: *Aptara®, Inc.*

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

Library of Congress Cataloging-in-Publication Data

Names: Navidi, William Cyrus, author. | Monk, Barry (Barry J.), author.
Title: Essential statistics / William Navidi, Colorado School of Mines, Barry Monk, Middle Georgia State University.
Description: Third edition. | New York, NY : McGraw-Hill Education, [2022] | Includes index.
Identifiers: LCCN 2020039533 (print) | LCCN 2020039534 (ebook) | ISBN 9781260359794 (hardcover) | ISBN 9781260492163 (spiral bound) | ISBN 9781264571833 (hardcover) | ISBN 9781260492194 (ebook) | ISBN 9781260492293 (ebook other)
Subjects: LCSH: Mathematical statistics—Textbooks.
Classification: LCC QA276.12 .N386 2022 (print) | LCC QA276.12 (ebook) | DDC 519.5—dc23
LC record available at https://lccn.loc.gov/2020039533
LC ebook record available at https://lccn.loc.gov/2020039534

To Catherine, Sarah, and Thomas

—William Navidi

To Shaun, Dawn, and Ben

—Barry Monk

About the Authors

©William Navidi

William Navidi is a professor of Applied Mathematics and Statistics at the Colorado School of Mines in Golden, Colorado. He received a Bachelor's degree in Mathematics from New College, a Master's degree in Mathematics from Michigan State University, and a Ph.D. in Statistics from the University of California at Berkeley. Bill began his teaching career at the County College of Morris, a two-year college in Dover, New Jersey. He has taught mathematics and statistics at all levels, from developmental through the graduate level. Bill has written two Engineering Statistics textbooks for McGraw Hill. In his spare time, he likes to play racquetball.

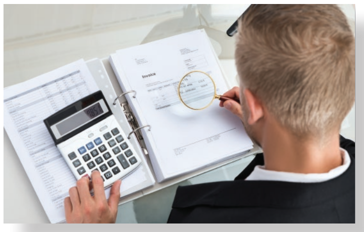
©Dawn Sherry

Barry Monk is a Professor of Mathematics with Middle Georgia State University in Macon, Georgia. Barry received a Bachelor of Science in Mathematical Statistics, a Master of Arts in Mathematics specializing in Optimization and Statistics, and a Ph.D. in Applied Mathematics, all from the University of Alabama. Barry has been teaching Introductory Statistics since 1992 in the classroom and online environments. Barry has a minor in Creative Writing and is a skilled jazz pianist.

Brief Contents

Contents

CHAPTER 5 Discrete Probability Distributions 205



Andriy Popov/123RF

CHAPTER 6 The Normal Distribution 237

Steve Allen/Getty Images

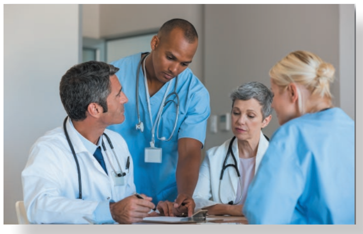
CHAPTER 7 Confidence Intervals 297

Schroptschop/E+/Getty Images

CHAPTER 8 Hypothesis Testing 351

Murat Taner/Getty Images

CHAPTER 9 Inferences on Two Samples 421


Ridofranz/Getty Images

CHAPTER 10 Tests with Qualitative Data 479


Mark Scott/Getty Images

CHAPTER 11 Correlation and Regression 505


Stockbyte/PunchStock

Preface

This book is designed for an introductory course in statistics. In addition to presenting the mechanics of the subject, we have endeavored to explain the concepts behind them in a writing style as straightforward, clear, and engaging as we could make it. As practicing statisticians, we have done everything possible to ensure that the material is accurate and correct. We believe that this book will enable instructors to explore statistical concepts in depth yet remain easy for students to read and understand.

To achieve this goal, we have incorporated a number of useful pedagogical features:

Features

- **Check Your Understanding Exercises:** After each concept is explained, one or more exercises are immediately provided for students to be sure they are following the material. These exercises provide students with confidence that they are ready to go on, or alert them to the need to review the material just covered.
- **Explain It Again:** Many important concepts are reinforced with additional explanation in these marginal notes.
- **Real Data:** Statistics instructors universally agree that the use of real data engages students and convinces them of the usefulness of the subject. A great many of the examples and exercises use real data. Some data sets explore topics in health or social sciences, while others are based in popular culture such as movies, contemporary music, or video games.
- **Integration of Technology:** Many examples contain screenshots from the TI-84 Plus calculator, MINITAB, and Excel. Each section contains detailed, step-by-step instructions, where applicable, explaining how to use these forms of technology to carry out the procedures explained in the text.
- **Interpreting Technology:** Many exercises present output from technology and require the student to interpret the results.
- **Write About It:** These exercises, found at the end of each chapter, require students to explain statistical concepts in their own words.
- **Case Studies:** Each chapter begins with a discussion of a real problem. At the end of the chapter, a case study demonstrates applications of chapter concepts to the problem.
- **In-Class Activities:** At the end of each chapter, activities are suggested that reinforce some concepts presented in the chapter.

Flexibility

We have endeavored to make our book flexible enough to work effectively with a wide variety of instructor styles and preferences. We cover both the *P*-value and critical value approaches to hypothesis testing, so instructors can choose to cover either or both of these methods.

Instructors differ in their preferences regarding the depth of coverage of probability. A light treatment of the subject may be obtained by covering Section 4.1 and skipping the rest of the chapter. More depth can be obtained by covering Section 4.2.

Supplements

Supplements, including a Corequisite Workbook, online homework, videos, guided student notes, and PowerPoint presentations, play an increasingly important role in the educational process. As authors, we have adopted a hands-on approach to the development of our supplements, to make sure that they are consistent with the style of the text and that they work effectively with a variety of instructor preferences. In particular, our online homework package offers instructors the flexibility to choose whether the solutions that students view are based on tables or technology, where applicable.

New in This Edition

The third edition of the book is intended to extend the strengths of the second. Some of the changes are:

- Discussions of the investigative process of statistics have been added, in accordance with recommendations of the GAISE report.
- In-class activities have been added to each chapter
- Material on the ratio and interval levels of measurement has been added.
- Material on bell-shaped histograms has been added.

- A discussion of the use of sample means to estimate population means has been added.
- Material on the uniform distribution has been added.
- A new section on multiple testing has been added.
- A new objective on the reasoning used in hypothesis testing has been added.
- New conceptual exercises regarding assumptions in constructing confidence intervals and performing hypothesis tests have been added.
- Additional material on Type I and Type II errors has been added.
- Objectives on the relationship between confidence intervals and the margin of error, calculating the sample size needed for a confidence interval of a given width, and the difference between confidence and probability are now presented in a context where the population standard deviation is unknown.
- Objectives on the relationship between confidence intervals and hypothesis tests, the relationship between the level of a test and the probability of error, the importance of reporting *P*-values, and the difference between statistical and practical significance are now presented in a context where the population standard deviation is unknown.
- Material on confidence intervals and hypothesis tests for paired samples now immediately follows the corresponding material for independent samples.
- A large number of new exercises have been included, many of which involve real data from recent sources.
- A large number of new exercises have been added to the online homework system, ALEKS. These include new conceptual questions and stepped-out solutions for the TI-84 Plus calculator and Excel.
- Several of the case studies have been updated.
- The exposition has been improved in a number of places.

William Navidi
Barry Monk

Acknowledgments

We are indebted to many people for contributions at every stage of development. Colleagues and students who reviewed the evolving manuscript provided many valuable suggestions. In particular, Charla Baker, John Trimboli, Don Brown, and Duane Day contributed to the supplements, and Mary Wolfe helped create the video presentations. Ashlyn Munson contributed a number of exercises, Abby Noble, Lindsay Lewis, and Ryan Melendez contributed many valuable ideas for revisions, and Tim Chappell and Lisa Rombes played an important role in the development of our digital content.

The staff at McGraw Hill has been extremely capable and supportive. Project Manager Jeni McAtee was always patient and helpful. We owe a debt of thanks to Emily DiGiovanna, Debbie McFarland, and Mary Ellen Rahn for their creative marketing and diligence in spreading the word about our book. We appreciate the guidance of our editors, Chrissy Skogen and Megan Platt, whose input has considerably improved the final product.

William Navidi
Barry Monk

Feedback from Statistics Instructors

Paramount to the development of ***Essential Statistics*** was the invaluable feedback provided by the instructors from around the country who reviewed the manuscript while it was in development.

A Special Thanks to All of the Symposia and Focus Group Attendees

James Adair, *Dyersburg State Community College*
Andrea Adlman, *Ventura College*
Leandro Alvarez, *Miami Dade College*
Simon Aman, *City Colleges of Chicago*
Diane Benner, *Harrisburg Area Community College*
Karen Brady, *Columbus State Community College*
Liliana Brand, *Northern Essex Community College*
Denise Brown, *Collin College–Spring Creek*
Don Brown, *Middle Georgia State University*
Mary Brown, *Harrisburg Area Community College*
Gerald Busald, *San Antonio College*
Anna Butler, *Polk State College*
Robert Cappetta, *College of DuPage*
Joe Castillo, *Broward College*
Michele Catterton, *Harford Community College*
Tim Chappell, *Metropolitan Communiity College - Penn Valley*
Ivette Chuca, *El Paso Community College*
James Coker, *University of Maryland Global Campus*
James Condor, *State College of Florida*
Milena Cuellar, *LaGuardia Community College*
Phyllis Curtiss, *Grand Valley State University*
Hema Deshmukh, *Mercyhurst University*
Mitra Devkota, *Shawnee State University*
Sue Jones Dobbyn, *Pellissippi State Community College*
Nataliya Doroshenko, *University of Memphis*
Rob Eby, *Blinn College–Bryan Campus*
Charles Wayne Ehler, *Anne Arundel Community College*
Franco Fedele, *University of West Florida*
Robert Fusco, *Broward College*
Wojciech Golik, *Lindenwood University*
Tim Grant, *Southwestern Illinois College*
Nina Greenberg, *University of New Mexico*
Todd Hendricks, *Georgia State University, Perimeter College*
Mary Hill, *College of DuPage*
Steward Huang, *University of Arkansas–Fort Smith*
Vera Hu-Hyneman, *Suffolk County Community College*
Laura Iossi, *Broward College*
Vera Ioudina, *Texas State University*
Brittany Juraszek, *Santa Fe College*
Maryann Justinger, *Erie Community College–South Campus*
Joseph Karnowski, *Norwalk Community College*
Esmarie Kennedy, *San Antonio College*
Lynette Kenyon, *Collin College–Plano*
Raja Khoury, *Collin College–Plano*
Alexander Kolesnik, *Ventura College*
Holly Kresch, *Diablo Valley College*
JoAnn Kump, *West Chester University*
Dan Kumpf, *Ventura College*
Erica Kwiatkowski-Egizio, *Joliet Junior College*
Hui Liu, *Georgia State University*
Pam Lowry, *Bellevue College*
Corey Manchester, *Grossmont College*
Scott McDaniel, *Middle Tennessee State University*
Mikal McDowell, *Cedar Valley College*
Ryan Melendez, *Arizona State University*
Lynette Meslinsky, *Erie Community College*

Penny Morris, *Polk State College*
Brittany Mosby, *Pellissippi State Community College*
Cindy Moss, *Skyline College*
Kris Mudunuri, *Long Beach City College*
Linda Myers, *Harrisburg Area Community College*
Sean Nguyen, *San Francisco State University*
Kim Page, *Middle Tennessee State University*
Ronald Palcic, *Johnson County Community College*
Matthew Pragel, *Harrisburg Area Community College*
Blanche Presley, *Middle Georgia State University*
Ahmed Rashed, *Richland College*
Cyndi Roemer, *Union County College*
Ginger Rowell, *Middle Tennessee State University*
Sudipta Roy, *Kanakee Community College*
Douglas Ruvolo, *University of Akron*
Ligo Samuel, *Austin Peay State University*
Jamal Salahat, *Owens State Community College*
Kathy Shay, *Middlesex County College*
Laura Shick, *Clemson University*
Larry Shrewsbury, *Southern Oregon University*
Shannon Solis, *San Jacinto College-North*
Tommy Thompson, *Cedar Valley College*
John Trimboli, *Middle Georgia State University*
Rita Sowell, *Volunteer State Community College*
Chris Turner, *Pensacola State College*
Jo Tucker, *Tarrant County College*
Dave Vinson, *Pellissippi State Community College*
Laurie Waites, *University of South Alabama*
Henry Wakhungu, *Indiana University*
Bin Wang, *University of South Alabama*
Daniel Wang, *Central Michigan University*
Rachel Webb, *Portland State University*
Jennifer Zeigenfuse, *Anne Arundel Community College*

Manuscript Review Panels

Alisher Abdullayev, *American River College*
Andrea Adlman, *Ventura College*
Olcay Akman, *Illinois State University*
Raid Amin, *University of West Florida*
Wesley Anderson, *Northwest Vista College*
Peter Arvanites, *Rockland Community College*
Diana Asmus, *Greenville Technical College*
John Avioli, *Christopher Newport University*
Robert Bass, *Gardner-Webb University*
Robbin Bates-Yelverton, *Park University*
Lynn Beckett-Lemus, *El Camino College*
Diane Benner, *Harrisburg Area Community College*
Abraham Biggs, *Broward College*
Wes Black, *Illinois Valley Community College*
Gregory Bloxom, *Pensacola State College*
Dale Bowman, *University of Memphis*
Brian Bradie, *Christopher Newport University*
Tonia Broome, *Gaston College*
Donna Brouilette, *Georgia State University, Perimeter College*
Allen Brown, *Wabash Valley College*
Denise Brown, *Collin Community College*
Don Brown, *Middle Georgia State University*
Mary Brown, *Harrisburg Area Community College*
Jennifer Bryan, *Oklahoma Christian University*
William Burgin, *Gaston College*
Gerald Busald, *San Antonio College*
David Busekist, *Southeastern Louisiana University*
Lynn Cade, *Pensacola State College*
Elizabeth Carrico, *Illinois Central College*
Connie Carroll, *Guilford Technical Community College*
Joseph Castillo, *Broward College*
Linda Chan, *Mount San Antonio College & Pasadena City College*
Ayona Chatterjee, *University of West Georgia*
Chand Chauhan, *Indiana University Purdue University Fort Wayne*
Pinyuen Chen, *Syracuse University*
Askar Choudhury, *Illinois State University*
Lee Clendenning, *University of North Georgia*
James Condor, *State College of Florida–Manatee*
Natalie Creed, *Gaston College*
John Curran, *Eastern Michigan University*
John Daniels, *Central Michigan University*
Shibasish Dasgupta, *University of South Alabama*
Nataliya Doroshenko, *University of Memphis*
Brandon Doughery, *Montgomery County Community College*
Larry Dumais, *American River College*
Christina Dwyer, *State College of Florida–Manatee*
Wayne Ehler, *Anne Arundel Community College*
Mark Ellis, *Central Piedmont Community College*
Angela Everett, *Chattanooga State Technical Community College*
Franco Fedele, *University of West Florida*
Harshini Fernando, *Purdue University—North Central*
Art Fortgang, *Southern Oregon University*
Thomas Fox, *Cleveland State Community College*
Robert Fusco, *Broward College*
Linda Galloway, *Kennesaw State University*
David Garth, *Truman State University*
Sharon Giles, *Grossmont Community College*
Mary Elizabeth Gore, *Community College of Baltimore County*
Carrie Grant, *Flagler College*

Delbert Greear, *University of North Georgia*
Jason Greshman, *Nova Southeastern University*
David Gurney, *Southeastern Louisiana University*
Chris Hail, *Union University–Jackson*
Ryan Harper, *Spartanburg Community College*
Phillip Harris, *Illinois Central College*
James Harrington, *Adirondack Community College*
Matthew He, *Nova Southeastern University*
Mary Beth Headlee, *State College of Florida–Manatee*
James Helmreich, *Marist College*
Todd Hendricks, *Georgia State University, Perimeter College*
Jada Hill, *Richland College*
Mary Hill, *College of DuPage*
William Huepenbecker, *Bowling Green State University–Firelands*
Patricia Humphrey, *Georgia Southern University*
Nancy Johnson, *State College of Florida–Manatee*
Maryann Justinger, *Erie Community College–South Campus*
Joseph Karnowski, *Norwalk Community College*
Susitha Karunaratne, *Purdue University—North Central*
Ryan Kasha, *Valencia College—West Campus*
Joseph Kazimir, *East Los Angeles College*
Esmarie Kennedy, *San Antonio College*
Lynette Kenyon, *Collin College*
Gary Kersting, *North Central Michigan College*
Raja Khoury, *Collin College*
Heidi Kiley, *Suffolk County Community College–Selden*
Daniel Kim, *Southern Oregon University*
Ann Kirkpatrick, *Southeastern Louisiana University*
John Klages, *County College of Morris*
Karon Klipple, *San Diego City College*
Matthew Knowlen, *Horry Georgetown Tech College*
JoAnn Kump, *West Chester University*
Alex Kolesnik, *Ventura College*
Bohdan Kunciw, *Salisbury University*
Erica Kwiatkowski-Egizio, *Joliet Junior College*
William Langston, *Finger Lakes Community College*
Tracy Leshan, *Baltimore City Community College*
Nicole Lewis, *East Tennessee State University*
Jiawei Liu, *Georgia State University*
Fujia Lu, *Endicott College*
Timothy Maharry, *Northwestern Oklahoma State University*
Aldo Maldonado, *Park University*
Kenneth Mann, *Catawba Valley Community College*
James Martin, *Christopher Newport University*
Erin Martin-Wilding, *Parkland College*
Amina Mathias, *Cecil College*
Catherine Matos, *Clayton State University*
Angie Matthews, *Broward College*
Mark McFadden, *Montgomery County Community College*
Karen McKarnin, *Allen Community College*
Penny Morris, *Polk Community College*
B. K. Mudunuri, *Long Beach City College–CalPoly Pomona*
Linda Myers, *Harrisburg Area Community College*
Miroslaw Mystkowski, *Gardner-Webb University*
Shai Neumann, *Brevard Community College*
Francis Nkansah, *Bunker Hill Community College*
Karen Orr, *Roane State Community College*
Richard Owens, *Park University*
Irene Palacios, *Grossmont College*
Luca Petrelli, *Mount Saint Mary's University*
Blanche Presley, *Middle Georgia State University*
Robert Prince, *Berry College*
Richard Puscas, *Georgia State University, Perimeter College*
Ramaswamy Radhakrishnan, *Illinois State University*
Leela Rakesh, *Central Michigan University*
Gina Reed, *University of North Georgia*
Andrea Reese, *Daytona State College–Daytona Beach*
Jim Robison-Cox, *Montana State University*
Alex Rolon, *Northampton Community College*
Jason Rosenberry, *Harrisburg Area Community College*
Yolanda Rush, *Illinois Central College*
Loula Rytikova, *George Mason University*
Fary Sami, *Harford Community College*
Vicki Schell, *Pensacola State College*
Carol Schoen, *University of Wisconsin–Eau Claire*
Pali Sen, *University of North Florida*
Rosa Seyfried, *Harrisburg Area Community College*
Larry Shrewsbury, *Southern Oregon University*
Abdallah Shuaibi, *Truman College*
Rick Silvey, *University of Saint Mary*
Russell Simmons, *Brookhaven College*
Peggy Slavik, *University of Saint Mary*
Karen Smith, *University of West Georgia*
Pam Stogsdill, *Bossier Parish Community College*
Susan Surina, *George Mason University*
Victor Swaim, *Southeastern Louisiana University*
Scott Sykes, *University of West Georgia*

Van Tran, *San Francisco State University*
John Trimboli, *Middle Georgia State University*
Barbara Tucker, *Tarrant County College South East*
Steven Forbes Tuckey, *Jackson Community College*
Christopher Turner, *Pensacola State College*
Anke Van Zuylen, *College of William and Mary*
Dave Vinson, *Pellissippi State Community College*
Joseph Walker, *Georgia State University*
James Wan, *Long Beach City College*
Xiaohong Wang, *Central Michigan University*
Jason Willis, *Gardner-Webb University*
Fuzhen Zhang, *Nova Southeastern University*
Yichuan Zhao, *Georgia State University*
Deborah Ziegler, *Hannibal LaGrange University*
Bashar Zogheib, *Nova Southeastern University*
Stephanie Zwyghuizen, *Jamestown Community College*

McGraw Hill ALEKS®

Personalized for your students, flexible for your approach.

Every student has different needs and enters your course with varied levels of preparation. ALEKS® pinpoints what students already know, what they don't and, most importantly, what they're ready to learn next. ALEKS® provides flexibility and control when shaping your course so you can reach every student.

ALEKS® Keeps Students Focused with a Personalized Learning Path

ALEKS® creates an optimized path with an ongoing cycle of learning and assessment, celebrating students' small wins along the way with positive real-time feedback. Rooted in research and analytics, ALEKS® improves student outcomes by fostering better preparation, increased motivation and knowledge retention.

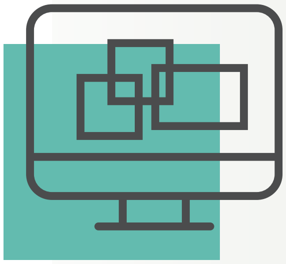
Your Course, Your Way

In a single platform, ALEKS® provides a balance of personalized, adaptive practice and non-adaptive assignments for application and assessment. This supports students individually while allowing you to create assignments that align to your course goals.

By differentiating instruction for students, ALEKS® unlocks the power of instructional time, group work, and applications and prepares students to make real-world connections.

//CODiE//

2019 SIIA CODiE WINNER



Flexible Implementation

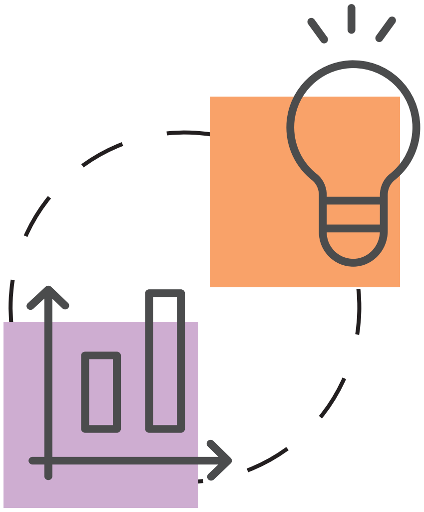
ALEKS® enables you to structure your course regardless of your instruction style and format. From a traditional classroom, to various corequisite models, to an online prep course before the start of the term, ALEKS® can supplement your instruction or play a lead role in delivering the content.

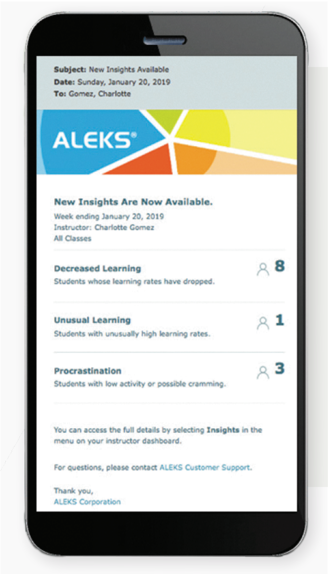
Teaching online? Respondus integration with ALEKS® supports secure testing with LockDown Browser and Respondus Monitor for Tests, Quizzes, and Scheduled Knowledge Checks.

Outcomes & Efficacy

Our commitment to improve student outcomes services a wide variety of implementation models and best practices, from lecture-based to labs and corequisites to summer prep courses. Our case studies illustrate our commitment to help you reach your course goals and our research demonstrates our drive to support all students, regardless of their math background and preparation level.

*visit **bit.ly/outcomesandefficacy** to review empirical data from ALEKS® users around the country





Turn Data Into Actionable Insights

ALEKS® Reports are designed to inform your instruction and create more meaningful interactions with your students when they need it the most. ALEKS® Insights alert you when students might be at risk of falling behind so that you can take immediate action. Insights summarize students exhibiting at least one of four negative behaviors that may require intervention including Failed Topics, Decreased Learning, Unusual Learning and Procrastination & Cramming.

DIGITAL EDGE50 AWARDS 2019

Winner of 2019 Digital Edge 50 Award for Data Analytics!

bit.ly/ALEKS_MHE

Supplements

Multimedia Supplements

This text is available with the adaptive online learning platform ALEKS. Rooted in 20 years of research and analytics, ALEKS identifies what students know, what they don't know, and what they are ready to learn, then personalizes a learning path tailored to each student. This personalization keeps students focused on what they need to learn and prepares them to make real-world connections. ALEKS also delivers textbook-aligned exercises, videos, and the many author-created digital supplements that accompany this text.

With **McGraw Hill Create**™, you can easily rearrange chapters, combine material from other content sources, and quickly upload content you have written like your course syllabus or teaching notes. Find the content you need in Create by searching through thousands of leading McGraw Hill textbooks. Arrange your book to fit your teaching style. Create even allows you to personalize your book's appearance by selecting the cover and adding your name, school, and course information. Assemble a Create book, and you'll receive a complimentary print review copy in 3–5 business days or a complimentary electronic review copy (eComp) via email in minutes. Go to www.mcgrawhillcreate.com today and experience how McGraw Hill Create™ empowers you to teach *your* students *your* way.

Computerized Test Bank Online (instructors only)

This computerized test bank, available online to adopting instructors, utilizes TestGen® cross-platform test generation software to quickly and easily create customized exams. Using hundreds of test items taken directly from the text, TestGen allows rapid test creation and flexibility for instructors to create their own questions from scratch with the ability to randomize number values. Powerful search and sort functions help quickly locate questions and arrange them in any order, and built-in mathematical templates let instructors insert stylized text, symbols, graphics, and equations directly into questions without need for a separate equation editor.

Videos

Author-produced lecture videos introduce concepts, definitions, formulas, and problem-solving procedures to help students better comprehend the topic at hand. Exercise videos illustrate the authors working through selected exercises, following the solution methodology employed in the text. These videos are closed-captioned for the hearing-impaired and meet the Americans with Disabilities Act Standards for Accessible Design.

SPSS Student Version for Windows

A student version of SPSS statistical software is available with copies of this text. Consult your McGraw Hill representative for details.

Guided Student Notes

Guided notes provide instructors with the framework of day-by-day class activities for each section in the book. Each lecture guide can help instructors make more efficient use of class time and can help keep students focused on active learning. Students who use the lecture guides have the framework of well-organized notes that can be completed with the instructor in class.

Data Sets

Data sets from selected exercises have been pre-populated into MINITAB, TI-Graph Link, Excel, SPSS, and comma-delimited ASCII formats for student and instructor use. These files are available on the text's website.

MINITAB 17 Manual

With guidance from the authors, this manual includes material from the book to provide seamless use from one to the other, providing additional practice in applying the chapter concepts while using the MINITAB program.

TI-84 Plus Graphing Calculator Manual

This friendly, author-influenced manual teaches students to learn about statistics and solve problems by using this calculator while following each text chapter.

Excel Manual

This workbook, specially designed to accompany the text by the authors, provides additional practice in applying the chapter concepts while using Excel.

Print Supplements

Annotated Instructor's Edition (instructors only)

The Annotated Instructor's Edition contains answers to all exercises. The answers to most questions are printed in blue next to each problem. Answers not appearing on the page can be found in the Answer Appendix at the end of the book.

Statistics Corequisite Workbook

This workbook, written by co-author Barry Monk, is designed to provide corequisite remediation of the necessary skills for an introductory statistics course. The included topics are largely independent of one another and may be used in any order that works best for the instructor. The workbook is available online or can be ordered in print format through Create.

Index of Applications

Library/book applications

Life sciences applications. *See* Biology/life sciences applications

Manufacturing applications

Medical/medical research/alternative medicine applications. *See also* Health/healthcare applications

Miscellaneous applications

Mike Flippo/123RF

chapter 2

Graphical Summaries of Data

Introduction

Are hybrid cars more fuel efficient than non-hybrid cars? Increasing prices of gasoline, along with concerns about the environment, have made fuel efficiency an important concern. To determine whether hybrid cars get better mileage than non-hybrid cars, we will compare the U.S. EPA mileage ratings for model year 2019 hybrid cars with the mileages for model year 2019 small non-hybrid cars. The following tables present the results, in miles per gallon.

Mileage Ratings for 2019 Hybrid Cars							
40	28	21	43	23	56	22	28
28	48	31	41	19	26	39	48
26	46	19	30	55	23	29	41
33	21	42	50	22	42	50	29
26	29	34	46	52	42	30	41
43	58	25	25	44	44	22	19
27	49	21	23	52	41	24	52
24	23	46					

Source: www.fueleconomy.gov

Mileage Ratings for 2019 Small Non-hybrid Cars							
33	32	28	32	34	33	27	25
31	26	30	33	32	32	29	27
33	30	29	31	30	37	34	30
35	34	36	33	35	33	39	36
32	36	22	35	36	33	28	31
27	28	35	27	31	32	29	34
30	33	29	32	32	30	32	29
26	30	32	29	31	31	32	33

Source: www.fueleconomy.gov

It is hard to tell from the lists of numbers whether the mileages differ substantially between hybrid and non-hybrid cars. What is needed are methods to summarize the data, so that their most important features stand out. One way to do this is by constructing graphs that allow us to visualize the important features of the data. In this chapter, we will learn how to construct many of the most commonly used graphical summaries. In the case study at the end of the chapter, you will be asked to use graphical methods to compare the mileages between hybrid and non-hybrid cars.

SECTION 2.1 Graphical Summaries for Qualitative Data

Objectives

1. Construct frequency distributions for qualitative data
2. Construct bar graphs
3. Construct pie charts

Objective 1 Construct frequency distributions for qualitative data

Frequency Distributions for Qualitative Data

How do retailers analyze their sales data to determine which methods of payment are most popular? Table 2.1 presents a list compiled by a retailer. The retailer accepts four types of credit cards: Visa, MasterCard, American Express, and Discover. The list contains the types of credit cards used by the last 50 customers.

Table 2.1 Types of Credit Cards Used

Discover	Visa	Visa	Am. Express	Visa
Visa	Visa	Am. Express	MasterCard	Visa
Am. Express	MasterCard	Visa	Visa	Visa
Visa	Am. Express	Am. Express	MasterCard	Visa
MasterCard	Visa	Discover	Am. Express	Discover
Visa	Am. Express	Discover	Visa	MasterCard
Visa	Visa	Visa	Visa	MasterCard
MasterCard	Am. Express	Visa	MasterCard	Visa
MasterCard	Discover	MasterCard	Visa	Visa
MasterCard	Discover	Am. Express	Discover	Visa

Table 2.1 is typical of data in raw form. It is a big list, and it's hard to gather much information simply by looking at it. To make the important features of the data stand out, we construct summaries. The starting point for many summaries is a frequency distribution.

DEFINITION

- The **frequency** of a category is the number of times it occurs in the data set.
- A **frequency distribution** is a table that presents the frequency for each category.

EXAMPLE 2.1

Construct a frequency distribution

Construct a frequency distribution for the data in Table 2.1.

Solution

To construct a frequency distribution, we begin by tallying the number of observations in each category and recording the totals in a table. Table 2.2 (page 37) presents a frequency distribution for the credit card data. We have included the tally marks in this table, but in practice it is permissible to omit them.

Table 2.2 Frequency Distribution for Credit Cards

Credit Card	Tally	Frequency
MasterCard	~~IIII~~ ~~IIII~~ I	11
Visa	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ III	23
Am. Express	~~IIII~~ IIII	9
Discover	~~IIII~~ II	7

CAUTION

When constructing a frequency distribution, be sure to check that the sum of all the frequencies is equal to the total number of observations.

It's a good idea to perform a check by adding the frequencies, to be sure that they add up to the total number of observations. In Table 2.2, the frequencies add up to 50, as they should.

Relative frequency distributions

A frequency distribution tells us exactly how many observations are in each category. Sometimes we are interested in the proportion of observations in each category. The proportion of observations in a category is called the *relative frequency* of the category.

Explain It Again

Difference between frequency and relative frequency: The frequency of a category is the number of items in the category. The relative frequency is the proportion of items in the category.

DEFINITION

The **relative frequency** of a category is the frequency of the category divided by the sum of all the frequencies.

$$\text{Relative frequency} = \frac{\text{Frequency}}{\text{Sum of all frequencies}}$$

We can add a column of relative frequencies to the frequency distribution. The resulting table is called a *relative frequency distribution*.

DEFINITION

A **relative frequency distribution** is a table that presents the relative frequency of each category. Often the frequency is presented as well.

EXAMPLE 2.2

Constructing a relative frequency distribution

Construct a relative frequency distribution for the data in Table 2.2.

Solution

We compute the relative frequencies for each type of credit card in Table 2.2 by using the following steps.

Step 1: Find the total number of observations by summing the frequencies.

$$\text{Sum of frequencies} = 11 + 23 + 9 + 7 = 50$$

Step 2: Find the relative frequency for the first category, MasterCard.

$$\text{Relative frequency for MasterCard} = \frac{11}{50} = 0.22$$

Step 3: Find the relative frequencies for the remaining categories.

$$\text{Relative frequency for Visa} = \frac{23}{50} = 0.46$$

$$\text{Relative frequency for Am. Express} = \frac{9}{50} = 0.18$$

$$\text{Relative frequency for Discover} = \frac{7}{50} = 0.14$$

Table 2.3 on page 38 presents a relative frequency distribution for the data in Table 2.2.

Table 2.3 Relative Frequency Distribution for Credit Cards

Credit Card	Frequency	Relative Frequency
MasterCard	11	0.22
Visa	23	0.46
Am. Express	9	0.18
Discover	7	0.14

Check Your Understanding

1. The following table lists the types of aircraft for the landings that occurred during a day at a small airport. ("Single" refers to single-engine and "Twin" refers to twin-engine.)

Types of Aircraft Landing at an Airport

Twin	Single	Helicopter	Turboprop	Twin	Single
Turboprop	Jet	Jet	Turboprop	Turboprop	Single
Jet	Single	Single	Twin	Twin	Turboprop
Helicopter	Single	Single	Single	Twin	Single
Jet	Twin	Twin	Single	Twin	Twin

a. Construct a frequency distribution for these data.
b. Construct a relative frequency distribution for these data.

Answers are on page 48.

Objective 2 Construct bar graphs

Bar Graphs

A **bar graph** is a graphical representation of a frequency distribution. A bar graph consists of rectangles of equal width, with one rectangle for each category. The heights of the rectangles represent the frequencies or relative frequencies of the categories. Example 2.3 shows how to construct a bar graph for the credit card data in Table 2.3.

EXAMPLE 2.3

Constructing bar graphs

Construct a frequency bar graph and a relative frequency bar graph for the credit card data in Table 2.3.

Solution

Step 1: Construct a horizontal axis. Place the category names along this axis, evenly spaced.
Step 2: Construct a vertical axis to represent the frequency or the relative frequency.
Step 3: Construct a bar for each category, with the heights of the bars equal to the frequencies or relative frequencies of their categories. The bars should not touch and should all be of the same width.

Figure 2.1(a) presents a frequency bar graph, and Figure 2.1(b) presents a relative frequency bar graph. The graphs are identical except for the scale on the vertical axis.

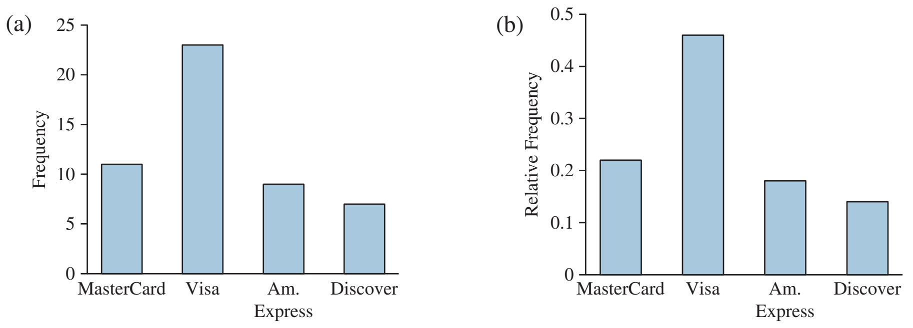


Figure 2.1 (a) Frequency bar graph. (b) Relative frequency bar graph.

Pareto charts

Sometimes it is desirable to construct a bar graph in which the categories are presented in order of frequency or relative frequency, with the largest frequency or relative frequency on the left and the smallest one on the right. Such a graph is called a **Pareto chart**. Pareto charts are useful when it is important to see clearly which are the most frequently occurring categories.

EXAMPLE 2.4

Constructing a Pareto chart

Construct a relative frequency Pareto chart for the data in Table 2.3.

Solution

Figure 2.2 presents the result. It is just like Figure 2.1(b) except that the bars are ordered from tallest to shortest.

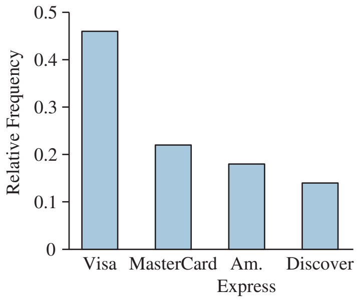


Figure 2.2 Pareto chart for the credit card data

Horizontal bars

The bars in a bar graph can be either horizontal or vertical. Horizontal bars are sometimes more convenient when the categories have long names.

EXAMPLE 2.5

Constructing bar graphs with horizontal bars

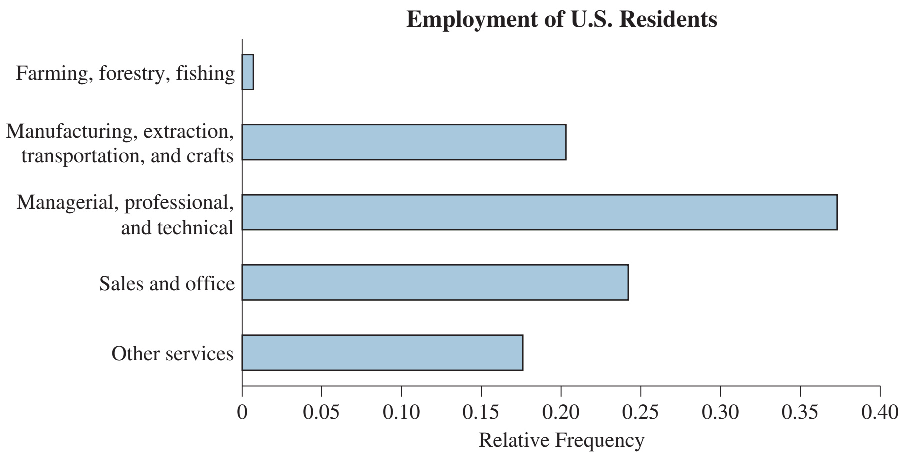
The following relative frequency distribution categorizes employed U.S. residents by type of employment in a recent year. Construct a relative frequency bar graph.

Type of Employment	Relative Frequency
Farming, forestry, fishing	0.007
Manufacturing, extraction, transportation, and crafts	0.203
Managerial, professional, and technical	0.373
Sales and office	0.242
Other services	0.176

Source: CIA—*The World Factbook*

Solution

The bar graph follows. We use horizontal bars, because the category names are long.



Side-by-side bar graphs

Sometimes we want to compare two bar graphs that have the same categories. The best way to do this is to construct both bar graphs on the same axes, putting bars that correspond to the same category next to each other. The result is called a **side-by-side bar graph**. As an illustration, Table 2.4 presents the number of active users, in millions, of several popular websites in 2017 and 2018.

Table 2.4 Monthly Active Users, in Millions

Website	2017	2018
Facebook	2129	2320
Gmail	1200	1500
Instagram	800	1000
LinkedIn	260	260
Twitter	330	321

Sources: Statista, Wikipedia, omnireagency.com

We would like to visualize the changes in the number of users between 2017 and 2018. Figure 2.3 presents a side-by-side bar graph. The bar graph clearly shows that the number of users of Facebook, Gmail, and Instagram grew noticeably, while the number of LinkedIn users held steady and the number of Twitter users decreased slightly.

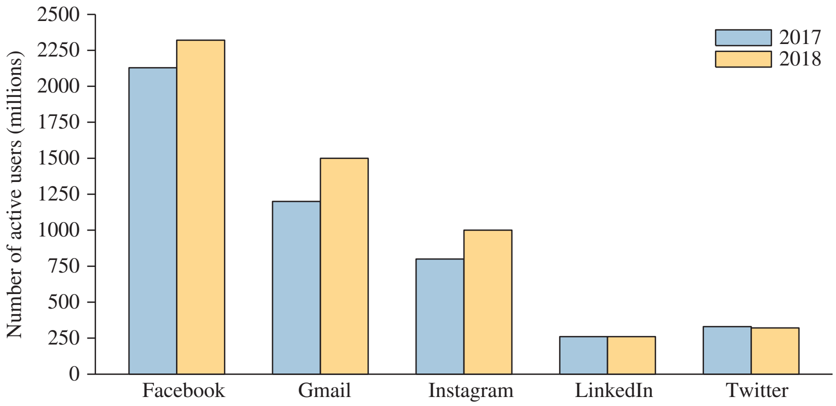


Figure 2.3 Side-by-side bar graph for the number of monthly active users of various websites

Objective 3 Construct pie charts

Pie Charts

A **pie chart** is an alternative to the bar graph for displaying relative frequency information. A pie chart is a circle. The circle is divided into sectors, one for each category. The relative sizes of the sectors match the relative frequencies of the categories. For example, if a category has a relative frequency of 0.25, then its sector takes up 25% of the circle. It is customary to label each sector with its relative frequency, expressed as a percentage. Example 2.6 illustrates the method for constructing a pie chart.

EXAMPLE 2.6

Constructing a pie chart

Construct a pie chart for the credit card data in Table 2.3.

Solution

For each category, we must determine how large the sector for that category must be. Since there are 360 degrees in a circle, we multiply the relative frequency of the category by 360 to determine the number of degrees in the sector for that category. For example, the relative frequency for the MasterCard category is 0.22. Therefore, the size of the sector for this category is $0.22 \cdot 360° = 79°$. Table 2.5 (page 41) presents the results for all the categories.

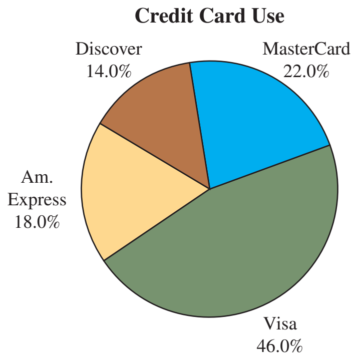


Figure 2.4 Pie chart for the credit card data in Table 2.5

Table 2.5 Sizes of Sectors for Pie Chart of Credit Card Data

Credit Card	Frequency	Relative Frequency	Size of Sector
MasterCard	11	0.22	79°
Visa	23	0.46	166°
Am. Express	9	0.18	65°
Discover	7	0.14	50°

Figure 2.4 presents the pie chart.

Constructing pie charts by hand is tedious. However, many software packages, such as MINITAB and Excel, can draw them. Step-by-step instructions for constructing a pie chart in MINITAB and Excel are presented in the Using Technology section on page 43.

Check Your Understanding

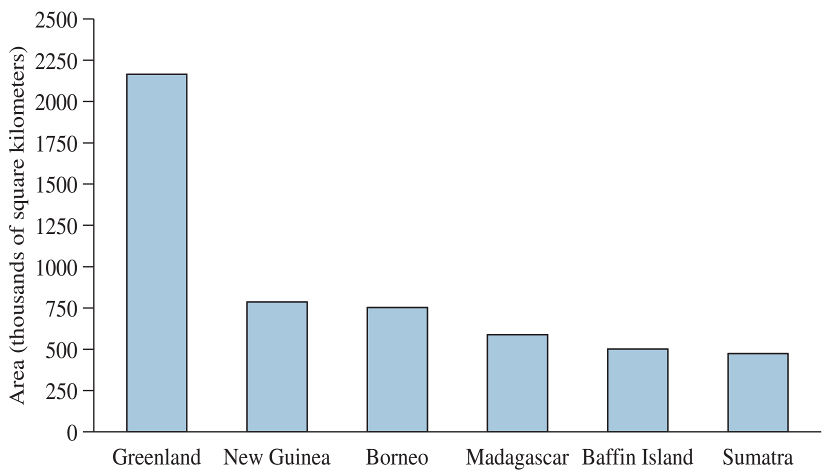
2. The following table presents a frequency distribution for the number of cars and light trucks sold in a recent month.

Type of Vehicle	Frequency
Small car	271,716
Midsize car	268,127
Luxury car	82,824
Minivan	56,772
SUV	151,703
Pickup truck	225,110
Crossover truck	416,104

Source: *Wall Street Journal*

a. Construct a bar graph.
b. Construct a relative frequency distribution.
c. Construct a relative frequency bar graph.
d. Construct a pie chart.

3. The following bar graph presents the areas (in thousands of square kilometers) of the six largest islands in the world.

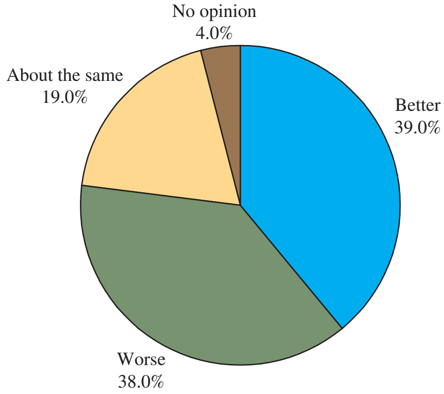


Source: CIA—*The World Factbook*

a. Which island is the largest in the world?
b. Someone says that Madagascar and Baffin Island together are larger than New Guinea. Is this correct? Explain how you can tell.

c. Approximately how large is Borneo?
d. Approximately how much larger is New Guinea than Sumatra?

4. The CBS News/New York Times poll asked a sample of people the following question: Do you think things in the United States five years from now will be better, worse, or about the same as they are today? The following pie chart presents the percentages of people who gave each response.



a. Which was the most common response?
b. What percentage of people said that things would be the same or worse in five years?
c. True or false: More than half of the people surveyed said that things would be the same or better in five years.

Answers are on page 48.

USING TECHNOLOGY

We use the data in Table 2.6 to illustrate the technology steps. Table 2.6 lists 20 responses to a survey question about the reason for visiting a local library.

Table 2.6

Study	Meet someone	Check out books	Meet someone	Check out books
Study	Study	Meet someone	Check out books	Check out books
Meet someone	Check out books	Study	Study	Check out books
Check out books	Study	Study	Meet someone	Study

EXCEL

Constructing a frequency distribution

Step 1. Enter the data in **Column A** with the label *Reason* in the topmost cell.

Step 2. Select **Insert**, then **Pivot Table**. Enter the range of cells that contain the data in the **Table/Range** field and click **OK**.

Step 3. In **Choose fields to add to report**, check Reason.

Step 4. Click on *Reason* and drag to the **Values** box. The result is shown in Figure A.

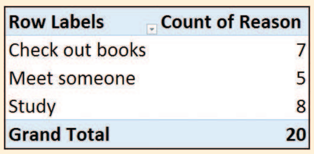


Figure A

Constructing bar graphs and pie charts

Step 1. Enter the categories in **Column A** and the frequencies or relative frequencies in **Column B**.

Step 2. Highlight the values in **Column A** and **Column B**, and select **Insert**. For a bar graph, select **Column**. For a pie chart, select **Pie**.

MINITAB

Constructing a frequency distribution

Step 1. Name your variable *Reason*, and enter the data into **Column C1**.

Step 2. Click on **Stat**, then **Tables**, then **Tally Individual Variables...**

Step 3. Double-click on the *Reason* variable, and check the **Counts** and **Percents** boxes.

Step 4. Press **OK** (Figure B).

Tally

Reason	Count	Percent
Check Out Books	7	35.00
Meet Someone	5	25.00
Study	8	40.00
N=	20	

Figure B

Constructing a bar graph

Step 1. Name your variable *Reason*, and enter the data into **Column C1**.

Step 2. Click on **Graph**, then **Bar Chart**. If given raw data as in Table 2.6, select **Bars Represent: Counts of Unique Values**. For the **Bar Chart type**, select **Simple**. Click **OK**. (If given a frequency distribution, select **Bars Represent: Values from a Table**.)

Step 3. Double-click on the *Reason* variable, and click on any of the options desired.

Step 4. Press **OK** (Figure C).

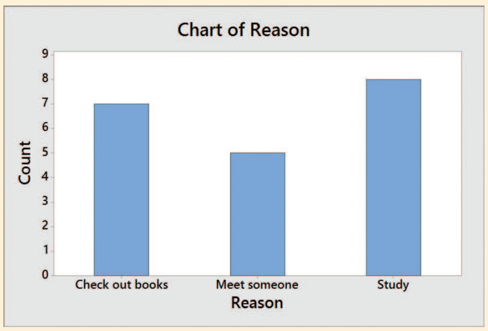


Figure C

Constructing a pie chart

Step 1. Name your variable *Reason*, and enter the data into **Column C1**.

Step 2. Click on **Graph**, then **Pie Chart**. If given raw data as in Table 2.6, select **Chart counts of unique values**, and click **OK**. (If given a frequency distribution, select **Chart Values from a Table**.)

Step 3. Double-click on the *Reason* variable, and click on any of the options desired.

Step 4. Press **OK** (Figure D).

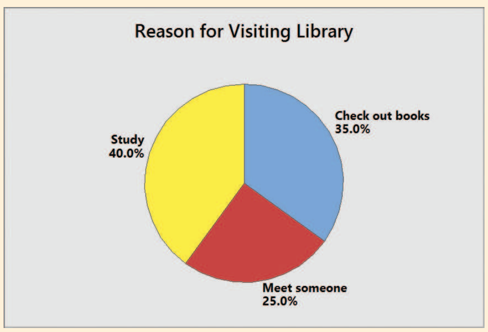


Figure D

SECTION 2.1 Exercises

Exercises 1–4 are the Check Your Understanding exercises located within the section.

Understanding the Concepts

In Exercises 5–8, fill in each blank with the appropriate word or phrase.

5. In a data set, the number of items that are in a particular category is called the ________ .

6. In a data set, the proportion of items that are in a particular category is called the ________ .

7. A ________ is a bar graph in which the bars are ordered by size.

8. A ________ is represented by a circle in which the sizes of the sectors match the relative frequencies of the categories.

In Exercises 9–12, determine whether the statement is true or false. If the statement is false, rewrite it as a true statement.

9. In a frequency distribution, the sum of all frequencies is less than the total number of observations.

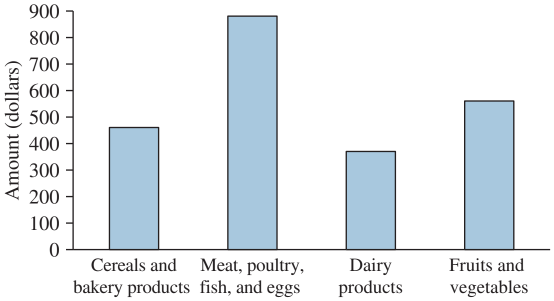
10. In a pie chart, if a category has a relative frequency of 30%, then its sector takes up 30% of the circle.

11. The relative frequency of a category is equal to the frequency divided by the sum of all frequencies.

12. In bar graphs and Pareto charts, the widths of the bars represent the frequencies or relative frequencies.

Practicing the Skills

13. The following bar graph presents the average amount a U.S. family spent, in dollars, on various food categories in a recent year.



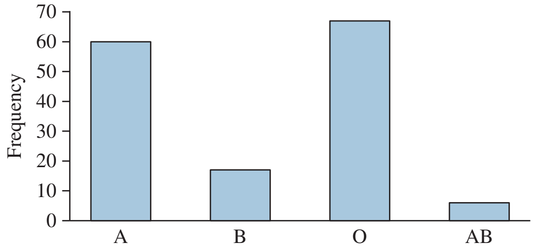
Source: Consumer Expenditure Survey

a. On which food category was the most money spent?

b. True or false: On the average, families spent more on cereals and bakery products than on fruits and vegetables.

c. True or false: Families spent more on animal products (meat, poultry, fish, eggs, and dairy products) than on plant products (cereals, bakery products, fruits, and vegetables).

14. The most common blood typing system divides human blood into four groups: A, B, O, and AB. The following bar graph presents the frequencies of these types in a sample of 150 blood donors.

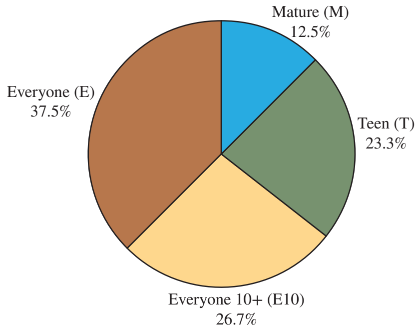


a. Which is the most frequent type?

b. True or false: More than half of the individuals in the sample had type O blood.

c. True or false: More than twice as many people had type A blood as had type B blood.

15. Following is a pie chart that presents the percentages of video games sold in each of four rating categories.



Source: Entertainment Software Association

a. Construct a relative frequency bar graph for these data.

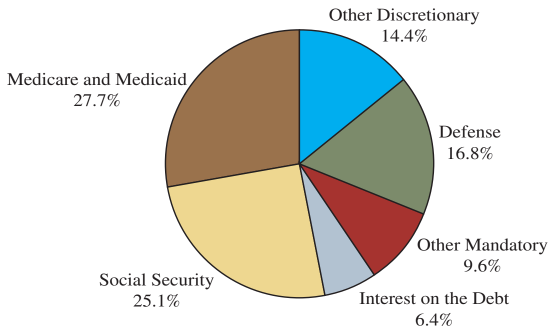
b. Construct a relative frequency Pareto chart for these data.

c. In which rating category are the most games sold?

d. True or false: More than twice as many T-rated games are sold as M-rated games.

e. True or false: Fewer than one in five games sold is an M-rated game.

16. Government spending: The following pie chart presents the percentages of the U.S. federal budget spent in various categories during a recent year.



Source: nationalpriorities.org

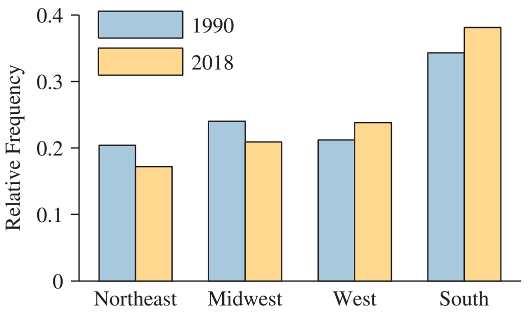
a. Construct a relative frequency bar graph for these data.

b. Construct a relative frequency Pareto chart for these data.

c. In which category was the largest amount spent?

d. Social Security, Medicare and Medicaid, and interest on the debt are considered to be mandatory spending because they fulfill promises made by the government ahead of time. Including other mandatory spending with these categories, what percentage of the spending was mandatory?

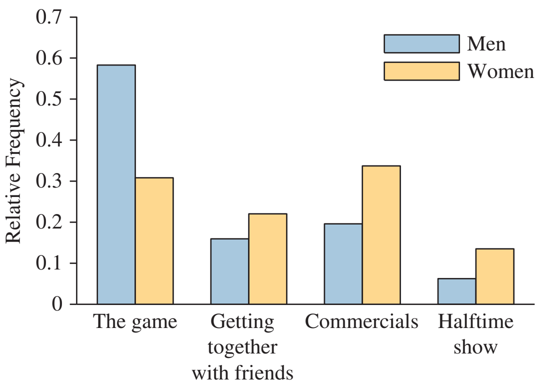
17. U.S. population: The following side-by-side bar graph presents the proportions of people residing in various geographic regions of the United States in 1990 and 2018.



Source: U.S. Census Bureau

a. Which regions increased as a proportion of the total from 1990 to 2018?
b. Which regions decreased as a proportion of the total from 1990 to 2018?
c. True or false: The South had the largest population in both 1990 and 2018.
d. True or false: All four regions had 20% or more of the population in 2018.

18. Super Bowl: The following side-by-side bar graph presents the results of a survey in which men and women were asked to name their favorite thing about watching the Super Bowl.



Source: BIGInsight

a. Which part of the Super Bowl do a greater proportion of men than women have as their favorite?
b. True or false: For both men and women, the smallest proportion have the halftime show as their favorite.
c. True or false: About twice as many men as women have the commercials as their favorite.
d. True or false: The proportion of men for whom the game or the half time show is their favorite is about the same as the proportion of women for whom the game or the commercials is their favorite.

Working with the Concepts

19. Phone sales: The following frequency distribution presents the number of phones (in millions) shipped in each quarter of each year from 2015 through 2018.

Quarter	Number Sold (in millions)
Jan.–Mar. 2015	334.4
Apr.–Jun. 2015	336.0
Jul.–Sep. 2015	359.7
Oct.–Dec. 2015	400.7
Jan.–Mar. 2016	332.2
Apr.–Jun. 2016	346.1
Jul.–Sep. 2016	363.4
Oct.–Dec. 2016	430.7
Jan.–Mar. 2017	344.4
Apr.–Jun. 2017	348.2
Jul.–Sep. 2017	377.8
Oct.–Dec. 2017	394.6
Jan.–Mar. 2018	344.3
Apr.–Jun. 2018	342.0
Jul.–Sep. 2018	355.2
Oct.–Dec. 2018	375.4

Source: Statista

a. Construct a frequency bar graph.
b. Construct a relative frequency distribution.
c. Construct a relative frequency bar graph.
d. True or false: In each year, the quarter with the largest sales was October to December.
e. True or false: Phone shipments increased each year from 2015 through 2018.

20. Popular video games: The following frequency distribution presents the number of copies sold at retail in the United States in 2018 for each of the ten best-selling video games.

Game	Sales (millions)
Spider-Man (PS4)	5.2
God of War (PS4)	5.0
FIFA 19 (PS4)	5.0
Monster Hunter: World (PS4)	4.4
Far Cry 5 (PS4)	3.7
Mario Kart 8 Deluxe (Switch)	3.3
Super Mario Odyssey (Switch)	2.6
Call of Duty: Black Ops III (PS4)	2.5
Legend of Zelda: Breath of the Wild (Switch)	2.2
Splatoon 2	2.1

Source: http://www.msn.com

a. Construct a frequency bar graph.
b. Construct a relative frequency distribution.
c. Construct a relative frequency bar graph.
d. True or false: More than 10% of the games sold were Spider-Man.

21. More Phones: Using the data in Exercise 19:
a. Construct a frequency distribution for the total number of phones sold in each of the four quarters Jan.–Mar., Apr.–Jun., Jul.–Sep., and Oct.–Dec.
b. Construct a frequency bar graph.
c. Construct a relative frequency distribution.
d. Construct a relative frequency bar graph.
e. Construct a pie chart.
f. True or false: More than half of phones were sold between October and December.

22. More video games: Using the data in Exercise 20:

a. Construct a frequency distribution that presents the total sales for each of the platforms among the top ten games.
b. Construct a frequency bar graph.
c. Construct a relative frequency distribution.
d. Construct a relative frequency bar graph.
e. Construct a pie chart.
f. True or false: More than half of the games sold among the top ten were for the PS4.

23. Hospital admissions: The following frequency distribution presents the five most frequent reasons for hospital admissions in U.S. community hospitals in a recent year.

Reason	Frequency (in thousands)
Congestive heart failure	990
Coronary atherosclerosis	1400
Heart attack	744
Infant birth	3800
Pneumonia	1200

Source: Agency for Health Care Policy and Research

a. Construct a frequency bar graph.
b. Construct a relative frequency distribution.
c. Construct a relative frequency bar graph.
d. Construct a relative frequency Pareto chart.
e. Construct a pie chart.
f. The categories coronary atherosclerosis, congestive heart failure, and heart attack refer to diseases of the circulatory system. True or false: There were more hospital admissions for infant birth than for diseases of the circulatory system.

January Smith/iStockphoto/Getty Images

24. World population: Following are the populations of the continents of the world (not including Antarctica) in the year 2019.

Continent	Population (in millions)
Africa	1320
Asia	4585
Europe	743
North America	366
Oceania	42
South America	432

Source: worldpopulationreview.com

a. Construct a frequency bar graph.
b. Construct a relative frequency distribution.
c. Construct a relative frequency bar graph.
d. Construct a relative frequency Pareto chart.
e. Construct a pie chart.
f. True or false: In the year 2019, more than half of the people in the world lived in Asia.
g. True or false: In the year 2019, there were more people in Europe than in North and South America combined.

25. Ages of video gamers: The Nielsen Company estimated the numbers of people in various gender and age categories who used a video game console. The results are presented in the following frequency distribution.

Gender and Age Group	Frequency (in millions)
Males 2–11	13.0
Females 2–11	10.1
Males 12–17	9.6
Females 12–17	6.2
Males 18–34	16.1
Females 18–34	11.6
Males 35–49	10.4
Females 35–49	9.3
Males 50+	3.5
Females 50+	3.9

Source: The Nielsen Company

a. Construct a frequency bar graph.
b. Construct a relative frequency distribution.
c. Construct a relative frequency bar graph.
d. Construct a pie chart.
e. True or false: More than half of video gamers are male.
f. True or false: More than 40% of video gamers are female.
g. What proportion of video gamers are 35 or over?

26. How secure is your job? In a survey, employed adults were asked how likely they thought it was that they would lose their jobs within the next year. The results are presented in the following frequency distribution.

Response	Frequency
Very likely	741
Fairly likely	859
Not too likely	3789
Not likely	9773

Source: General Social Survey

a. Construct a frequency bar graph.
b. Construct a relative frequency distribution.
c. Construct a relative frequency bar graph.
d. Construct a pie chart.
e. True or false: More than half of the people surveyed said that it was not likely that they would lose their job.
f. What proportion of the people in the survey said that it was very likely or fairly likely that they would lose their job?

27. Back up your data: In a survey commissioned by the Maxtor Corporation, U.S. computer users were asked how often they backed up their computer's hard drive. The following frequency distribution presents the results.

Response	Frequency
More than once per month	338
Once every 1–3 months	424
Once every 4–6 months	212
Once every 7–11 months	127
Once per year or less	311
Never	620

Source: The Maxtor Corporation

a. Construct a frequency bar graph.
b. Construct a relative frequency distribution.
c. Construct a relative frequency bar graph.
d. Construct a pie chart.
e. True or false: More than 30% of the survey respondents never back up their data.
f. True or false: Less than 50% of the survey respondents back up their data more than once per year.

28. Education levels: The following frequency distribution categorizes U.S. adults aged 18 and over by educational attainment in a recent year.

Educational Attainment	Frequency (in thousands)
None	834
1–4 years	1,764
5–6 years	3,618
7–8 years	4,575
9 years	4,068
10 years	4,814
11 years	11,429
High school graduate	70,441
Some college but no degree	45,685
Associate's degree (occupational)	9,380
Associate's degree (academic)	12,100
Bachelor's degree	43,277
Master's degree	16,625
Professional degree	3,099
Doctoral degree	3,191

Source: U.S. Census Bureau

a. Construct a frequency bar graph.
b. Construct a relative frequency distribution.
c. Construct a relative frequency bar graph.
d. Construct a frequency distribution with the following categories: 8 years or less, 9–11 years, High school graduate, Some college but no degree, College degree (Associate's or Bachelor's), Graduate degree (Master's, Professional, or Doctoral).
e. Construct a pie chart for the frequency distribution in part (d).
f. What proportion of people did not graduate from high school?

29. Twitter followers: The following frequency distribution presents the number of Twitter followers in 2019 for each of five well-known singers.

Singer	Followers (millions)
Katy Perry	107.3
Justin Bieber	105.4
Rihanna	90.7
Taylor Swift	83.2
Lady Gaga	78.5

Source: www.twittercounter.com

a. Construct a frequency bar graph.
b. Construct a relative frequency distribution.
c. Construct a relative frequency bar graph.
d. Construct a pie chart.
e. What proportion are following Taylor Swift?

30. Music sales: The following frequency distribution presents the sales, in millions of dollars, for several categories of music in the years 2017 and 2108.

	Sales	
Type of Music	2017	2018
Digital subscription	4092.1	5403.1
Digital streaming	1572.4	1963.7
Digital downloads	1330.7	1039.1
Physical	1495.5	1154.8

Source: Recording Industry Association of America

a. Construct a relative frequency distribution for the 2017 sales.
b. Construct a relative frequency distribution for the 2018 sales.
c. Construct a side-by-side relative frequency bar graph to compare the sales in 2017 and 2018.
d. True or false: Sales in every category increased from 2017 to 2018.

31. Keeping up with the Kardashians: The following frequency distribution presents the number of Twitter and Instagram followers in a recent year for five members of the Kardashian family.

	Followers (in millions)	
Kardashian	Twitter	Instagram
Kim	60.5	127
Kendall	27.7	104
Kylie	27.1	127
Khloé	27.0	86.9
Kourtney	24.3	73.5

Sources: twitter.com, www.cheatsheet.com

a. Construct a relative frequency distribution for the number of Twitter followers.
b. Construct a relative frequency distribution for the number of Instagram followers.
c. Construct a side-by-side relative frequency bar graph to compare the number of Twitter followers to the number of Instagram followers.
d. True or false: Each member of the family has more Instagram followers than Twitter followers.
e. True or false: Kim has more Twitter followers than her sisters Khloé and Kourtney combined.

32. Bought a new car lately? The following table presents the number of cars sold by several manufacturers in a recent month.

Manufacturer	Sales (in thousands)
General Motors	82.2
Toyota	96.6
Ford	68.4
Chrysler	26.6
Honda	69.9
Nissan	78.0
Hyundai	42.0
Others	159.0

Source: *Wall Street Journal*

a. Construct a frequency bar graph.
b. Construct a relative frequency distribution.

c. Construct a relative frequency bar graph.
d. Construct a pie chart.
e. What proportion of sales were for General Motors cars?

33. Bought a new truck lately? The following table presents the number of light trucks sold by several manufacturers in a recent month.

Manufacturer	Sales (in thousands)
General Motors	173.0
Toyota	101.6
Ford	170.6
Chrysler	167.9
Honda	68.8
Nissan	62.5
Hyundai	25.5
Others	121.0

Source: *Wall Street Journal*

a. Construct a frequency bar graph.
b. Construct a relative frequency distribution.
c. Construct a relative frequency bar graph.
d. Construct a pie chart.
e. True or false: More light trucks were sold by Chrysler than by Honda, Nissan, and Hyundai combined.

34. Happy Halloween: The following table presents proportions of people who get ideas for Halloween costumes from various sources. Is this a valid relative frequency distribution? Why or why not?

Source	Proportion
Twitter	0.048
Facebook	0.152
Friends and Family	0.237
Retail Stores	0.357
Magazines	0.193
Pinterest	0.071

Source: National Retail Federation

Extending the Concepts

35. Native languages: The following frequency distribution presents the number of households (in thousands) categorized by the language spoken at home, for the cities of New York and Los Angeles in a recent year. The Total column presents the numbers of households in both cities combined.

Language	New York	Los Angeles	Total
English	4098	1339	5437
Spanish	1870	1555	3425
Other Indo-European	1037	237	1274
Asian and Pacific Island	618	301	919

Source: U.S. Census Bureau

a. Construct a frequency bar graph for each city.
b. Construct a frequency bar graph for the total.
c. Construct a relative frequency bar graph for each city.
d. Construct a relative frequency bar graph for the total.
e. Explain why the heights of the bars for the frequency bar graph for the total are equal to the sums of the heights for the individual cities.
f. Explain why the heights of the bars for the relative frequency bar graph for the total are not equal to the sums of the heights for the individual cities.

36. Proportion of females: Following are the proportions of the United States population that is female for five age groups.

Age Group	Proportion
0–19	0.488
20–39	0.497
40–59	0.508
60–79	0.534
over 79	0.628

Source: U.S. Census Bureau

a. Is this a relative frequency table? Explain why or why not?
b. Would it be appropriate to construct a pie chart for these data? Why or why not?

Answers to Check Your Understanding Exercises for Section 2.1

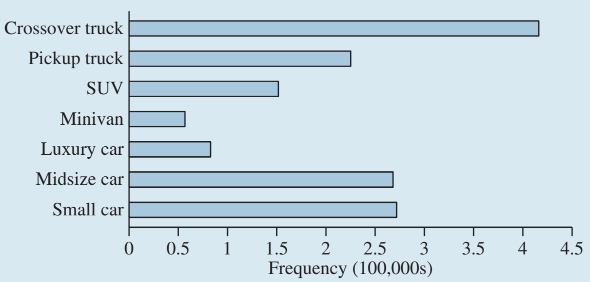
1. a.

Aircraft	Frequency
Twin	9
Single	10
Helicopter	2
Turboprop	5
Jet	4

b.

Aircraft	Frequency	Relative Frequency
Twin	9	0.300
Single	10	0.333
Helicopter	2	0.067
Turboprop	5	0.167
Jet	4	0.133

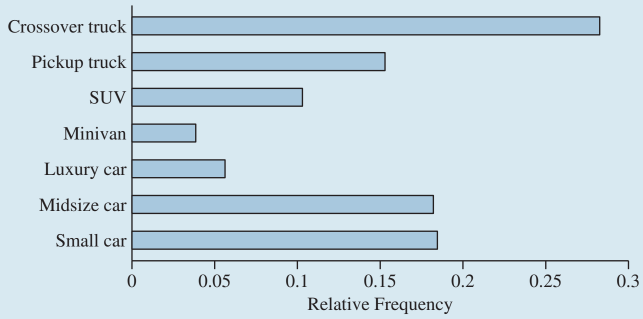
2. a.



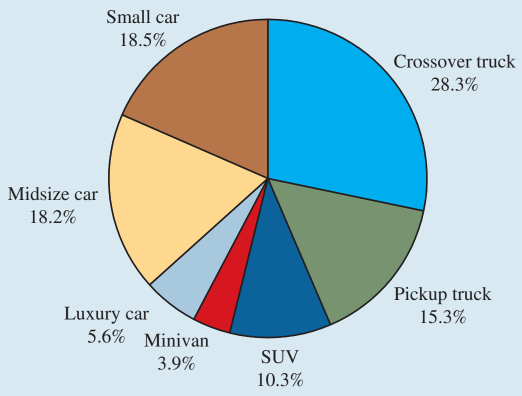
b.

Type of Vehicle	Frequency	Relative Frequency
Small car	271,716	0.185
Midsize car	268,127	0.182
Luxury car	82,824	0.056
Minivan	56,772	0.039
SUV	151,703	0.103
Pickup truck	225,110	0.153
Crossover truck	416,104	0.283

c.



d.



3. a. Greenland

b. Yes, the height of the bar for New Guinea is less than the heights of the bars for Madagascar and Baffin Island put together.

c. 750,000 **d.** 300,000

4. a. Better **b.** 57% **c.** True

SECTION 2.2 Frequency Distributions and Their Graphs

Objectives

1. Construct frequency distributions for quantitative data
2. Construct histograms
3. Determine the shape of a distribution from a histogram

Objective 1 Construct frequency distributions for quantitative data

Frequency Distributions for Quantitative Data

How much air pollution is caused by motor vehicles? This question was addressed in a study by Dr. Janet Yanowitz at the Colorado School of Mines. She studied the emissions of particulate matter, a form of pollution consisting of tiny particles, that has been associated with respiratory disease. The emissions for 65 vehicles, in units of grams of particles per gallon of fuel, are presented in Table 2.7.

Table 2.7 Particulate Emissions for 65 Vehicles

1.50	0.87	1.12	1.25	3.46	1.11	1.12	0.88	1.29	0.94	0.64	1.31	2.49
1.48	1.06	1.11	2.15	0.86	1.81	1.47	1.24	1.63	2.14	6.64	4.04	2.48
1.40	1.37	1.81	1.14	1.63	3.67	0.55	2.67	2.63	3.03	1.23	1.04	1.63
3.12	2.37	2.12	2.68	1.17	3.34	3.79	1.28	2.10	6.55	1.18	3.06	0.48
0.25	0.53	3.36	3.47	2.74	1.88	5.94	4.24	3.52	3.59	3.10	3.33	4.58

To summarize these data, we will construct a frequency distribution. Since these data are quantitative, there are no natural categories. We therefore divide the data into **classes**. The classes are intervals of equal width that cover all the values that are observed. For example, for the data in Table 2.7, we could choose the classes to be 0.00–0.99, 1.00–1.99, and so forth. We then count the number of observations that fall into each class, to obtain the class frequencies.

EXAMPLE 2.7

Construct a frequency distribution

Construct a frequency distribution for the data in Table 2.7, using the classes 0.00–0.99, 1.00–1.99, and so on.

Explain It Again

Frequency distributions for quantitative and qualitative data: Frequency distributions for quantitative data are just like those for qualitative data, except that the data are divided into classes rather than categories.

Solution

First we list the classes. We begin by noting that the smallest value in the data set is 0.25 and the largest is 6.64. We list classes until we get to the class that contains the largest value. The classes are 0.00–0.99, 1.00–1.99, 2.00–2.99, 3.00–3.99, 4.00–4.99, 5.00–5.99, and 6.00–6.99. Since the largest number in the data set is 6.64, these are enough classes.

Now we count the number of observations that fall into each class. The first class is 0.00–0.99. We count nine observations between 0.00 and 0.99 in Table 2.7. The next class is 1.00–1.99. We count 26 observations in this class. We repeat this procedure with classes 2.00–2.99, 3.00–3.99, 4.00–4.99, 5.00–5.99, and 6.00–6.99. The results are presented in Table 2.8. This is a frequency distribution for the data in Table 2.7.

Table 2.8 Frequency Distribution for Particulate Data

Class	Frequency
0.00–0.99	9
1.00–1.99	26
2.00–2.99	11
3.00–3.99	13
4.00–4.99	3
5.00–5.99	1
6.00–6.99	2

We can also construct a relative frequency distribution. As with qualitative data, the relative frequency of a class is the frequency of that class, divided by the sum of all the frequencies.

DEFINITION

The **relative frequency** of a class is given by

$$\text{Relative frequency} = \frac{\text{Frequency}}{\text{Sum of all frequencies}}$$

EXAMPLE 2.8

Construct a relative frequency distribution

Construct a relative frequency distribution for the data in Table 2.7, using the classes 0.00–0.99, 1.00–1.99, and so on.

Solution

The frequency distribution is presented in Table 2.8. We compute the sum of all the frequencies:

$$\text{Sum of all frequencies} = 9 + 26 + 11 + 13 + 3 + 1 + 2 = 65$$

We can now compute the relative frequency for each class. For the class 0.00–0.99, the frequency is 9. The relative frequency is therefore

$$\text{Relative frequency} = \frac{\text{Frequency}}{\text{Sum of all frequencies}} = \frac{9}{65} = 0.138$$

Table 2.9 is a relative frequency distribution for the data in Table 2.7. The frequencies are shown as well.

Table 2.9 Relative Frequency Distribution for Particulate Data

Class	Frequency	Relative Frequency
0.00–0.99	9	0.138
1.00–1.99	26	0.400
2.00–2.99	11	0.169
3.00–3.99	13	0.200
4.00–4.99	3	0.046
5.00–5.99	1	0.015
6.00–6.99	2	0.031

Choosing the classes

In Examples 2.7 and 2.8, we chose the classes to be 0.00–0.99, 1.00–1.99, and so on. There are many other choices we could have made. For example, we could have chosen the classes to be 0.00–1.99, 2.00–3.99, 4.00–5.99, and 6.00–7.99. As another example, we could have chosen them to be 0.00–0.49, 0.50–0.99, and so on, up to 6.50–6.99. We now define some of the terminology that we will use when discussing classes.

DEFINITION

- The **lower class limit** of a class is the smallest value that can appear in that class.
- The **upper class limit** of a class is the largest value that can appear in that class.
- The **class width** is the difference between consecutive lower class limits.

CAUTION

The class width is the difference between the lower limit and the lower limit of the next class, not the difference between the lower limit and the upper limit.

Class limits should be expressed with the same number of decimal places as the data. The data in Table 2.7 are rounded to two decimal places, so the class limits for these data are expressed with two decimal places as well.

EXAMPLE 2.9

Find the class limits and widths

Find the lower class limits, the upper class limits, and the class widths for the relative frequency distribution in Table 2.9.

Solution

The classes are 0.00–0.99, 1.00–1.99, and so on, up to 6.00–6.99. The lower class limits are therefore 0.00, 1.00, 2.00, 3.00, 4.00, 5.00, and 6.00. The upper class limits are 0.99, 1.99, 2.99, 3.99, 4.99, 5.99, and 6.99.

We find the class width for the first class by subtracting consecutive lower limits:

$$\begin{aligned}\text{Class width} &= \text{Lower limit for second class} - \text{Lower limit for first class}\\ &= 1.00 - 0.00\\ &= 1.00\end{aligned}$$

Similarly, we find that all the classes have a width of 1.

When constructing a frequency distribution, there is no one right way to choose the classes. However, there are some requirements that must be satisfied:

Requirements for Choosing Classes

- Every observation must fall into one of the classes.
- The classes must not overlap.
- The classes must be of equal width.
- There must be no gaps between classes. Even if there are no observations in a class, it must be included in the frequency distribution.

The following procedure will produce a frequency distribution whose classes meet these requirements.

Procedure for Constructing a Frequency Distribution for Quantitative Data

Step 1: Choose a class width.

Step 2: Choose a lower class limit for the first class. This should be a convenient number that is slightly less than the minimum data value.

Step 3: Compute the lower limit for the second class by adding the class width to the lower limit for the first class:

$$\text{Lower limit for second class} = \text{Lower limit for first class} + \text{Class width}$$

Step 4: Compute the lower limits for each of the remaining classes by adding the class width to the lower limit of the preceding class. Stop when the largest data value is included in a class.

Step 5: Count the number of observations in each class, and construct the frequency distribution.

EXAMPLE 2.10

Constructing a frequency distribution

Construct a frequency distribution for the data in Table 2.7, using a class width of 1.50.

Solution

Step 1: The class width is given to be 1.50.

Step 2: The smallest value in the data is 0.25. A convenient number that is smaller than 0.25 is 0.00. We will choose 0.00 to be the lower limit for the first class.

Step 3: The lower class limit for the second class is $0.00 + 1.50 = 1.50$.

Step 4: Continuing, the lower limits for the remaining classes are

$$1.50 + 1.50 = 3.00$$
$$3.00 + 1.50 = 4.50$$
$$4.50 + 1.50 = 6.00$$
$$6.00 + 1.50 = 7.50$$

Since the largest data value is 6.64, every data value is now contained in a class.

Step 5: We count the number of observations in each class to obtain the following frequency distribution.

Frequency Distribution for Particulate Data Using a Class Width of 1.5

Class	Frequency
0.00–1.49	28
1.50–2.99	18
3.00–4.49	15
4.50–5.99	2
6.00–7.49	2

Check Your Understanding

1. Using the data in Table 2.7, construct a frequency distribution with classes of width 0.5.

Answer is on page 66.

Computing the class width for a given number of classes

In Example 2.10, the first step in computing the frequency distribution was to choose a class width. Sometimes we begin by choosing an approximate number of classes instead. In these cases, we compute the class width as follows:

Step 1: Decide approximately how many classes to have.

Step 2: Compute the class width as follows:

$$\text{Class width} = \frac{\text{Largest data value} - \text{Smallest data value}}{\text{Number of classes}}$$

Step 3: Round the class width to a convenient value. It is usually better to round up.

Once the class width is determined, we proceed just as in the case where the class width is given. We choose a lower limit for the first class by choosing a convenient number that is slightly less than the minimum data value. We then compute the lower limits for the remaining classes, count the number of observations in each class, and construct the frequency distribution. Note that the actual number of classes may differ somewhat from the chosen number, because the class width is rounded and because the lower limit of the first class will generally be less than the smallest data value.

EXAMPLE 2.11

Computing the class width

Find the class width for a frequency distribution for the data in Table 2.7, if it is desired to have approximately seven classes.

Solution

Step 1: We will have approximately seven classes.

Step 2: The smallest data value is 0.25 and the largest is 6.64. We compute the class width:

$$\text{Class width} = \frac{6.64 - 0.25}{7} = 0.91$$

Step 3: We round 0.91 up to 1, since this is the nearest convenient number. We will use a class width of 1.

A reasonable choice for the lower limit of the first class is 0. This choice will give us the frequency distribution in Table 2.8.

Objective 2 Construct histograms

Histograms

Once we have a frequency distribution or a relative frequency distribution, we can put the information in graphical form by constructing a **histogram**. Histograms based on frequency distributions are called **frequency histograms**, and histograms based on relative frequency distributions are called **relative frequency histograms**. Histograms are related to bar graphs and are appropriate for quantitative data. A histogram is constructed by drawing a rectangle for each class. The heights of the rectangles are equal to the frequencies or the relative frequencies, and the widths are equal to the class width. The left edge of each rectangle corresponds to the lower class limit, and the right edge touches the left edge of the next rectangle.

EXAMPLE 2.12

Construct a histogram

Table 2.10 presents a frequency distribution and the relative frequency distribution for the particulate emissions data.

Construct a frequency histogram based on the frequency distribution in Table 2.10. Construct a relative frequency histogram based on the relative frequency distribution in Table 2.10.

Table 2.10 Frequency and Relative Frequency Distributions for Particulate Data

Class	Frequency	Relative Frequency
0.00–0.99	9	0.138
1.00–1.99	26	0.400
2.00–2.99	11	0.169
3.00–3.99	13	0.200
4.00–4.99	3	0.046
5.00–5.99	1	0.015
6.00–6.99	2	0.031

Solution

We construct a rectangle for each class. The first rectangle has its left edge at the lower limit of the first class, which is 0.00, and its right edge at the lower limit of the next class, which is 1.00. The second rectangle has its left edge at 1.00 and its right edge at the lower limit of the next class, which is 2.00, and so on.

For the frequency histogram, the heights of the rectangles are equal to the frequencies. For the relative frequency histogram, the heights of the rectangles are equal to the relative frequencies.

Figure 2.5 presents a frequency histogram, and Figure 2.6 presents a relative frequency histogram, for the data in Table 2.10. Note that the two histograms have the same shape. The only difference is the scale on the vertical axis.

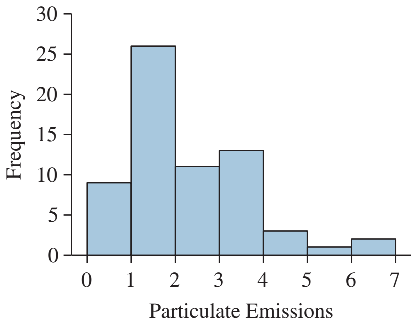


Figure 2.5 Frequency histogram for the frequency distribution in Table 2.10

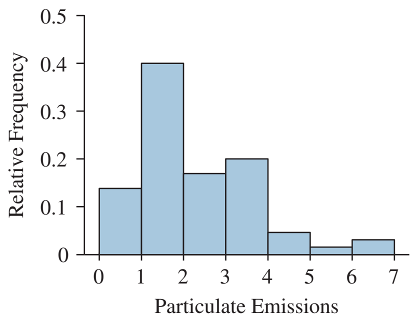


Figure 2.6 Relative frequency histogram for the relative frequency distribution in Table 2.10

Explain It Again

Choosing the number of classes: There is no single right way to choose classes for a histogram. Use your best judgment to construct a histogram with an appropriate amount of detail.

How should I choose the number of classes for a histogram?

There are no hard-and-fast rules for choosing the number of classes. In general, it is good to have more classes rather than fewer, but it is also good to have reasonably large frequencies in some of the classes. The following two principles can guide the choice:

- Too many classes produce a histogram with too much detail, so that the main features of the data are obscured.
- Too few classes produce a histogram lacking in detail.

Figures 2.7 and 2.8 (page 55) illustrate these principles. Figure 2.7 presents a histogram for the particulate data where the class width is 0.1. This narrow class width results in a

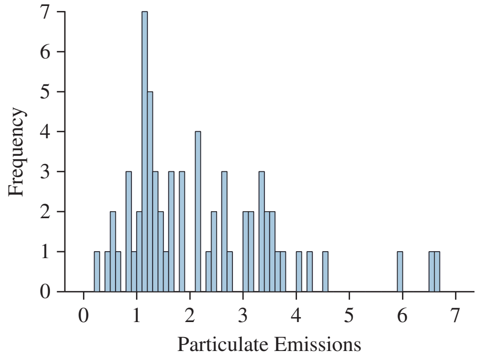


Figure 2.7 The class width is too narrow. The jagged appearance distracts from the overall shape of the data.

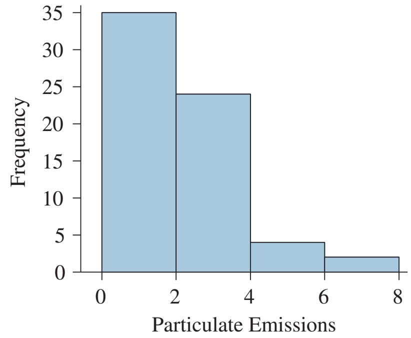


Figure 2.8 The class width is too wide. Only the most basic features of the data are visible.

large number of classes. The histogram has a jagged appearance, which distracts from the overall shape of the data. On the other extreme, Figure 2.8 presents a histogram for these data with a class width of 2.0. The number of classes is too small, so only the most basic features of the data are visible in this overly simple histogram.

Choosing a large number of classes will produce a narrow class width, and choosing a smaller number will produce a wider class width. It is appropriate to experiment with various choices for the number of classes, in order to find a good balance. The following guidelines are helpful.

Guidelines for Selecting the Number of Classes

- For many data sets, the number of classes should be at least 5 but no more than 20.
- For very large data sets, a larger number of classes may be appropriate.

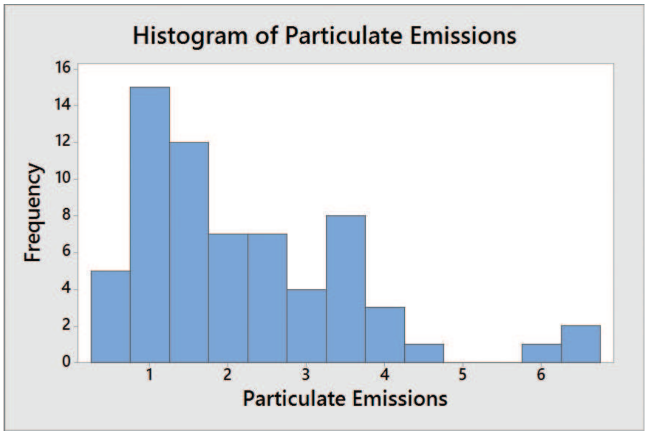
EXAMPLE 2.13

Constructing a histogram with technology

Use technology to construct a frequency histogram for the emissions data in Table 2.7 on page 49.

Solution

The following figure shows the histogram constructed in MINITAB. Note that MINITAB has chosen a class width of 0.5. With this class width, there are two empty classes. These show up as a gap that separates the last two rectangles on the right from the rest of the histogram.



Step-by-step instructions for constructing histograms with the TI-84 Plus and with MINITAB are given in the Using Technology section on pages 59 and 60.

Check Your Understanding

2. Following is a frequency distribution that presents the number of live births to women aged 15–44 in the state of Wyoming in a recent year.

Distribution of Births by Age of Mother

Age	Frequency
15–19	795
20–24	2410
25–29	2190
30–34	1208
35–39	499
40–44	109

Source: Wyoming Department of Health

a. List the lower class limits.
b. What is the class width?
c. Construct a frequency histogram.
d. Construct a relative frequency distribution.
e. Construct a relative frequency histogram.

Answers are on page 66.

Open-ended classes

It is sometimes necessary for the first class to have no lower limit or for the last class to have no upper limit. Such a class is called **open-ended**. Table 2.11 presents a frequency distribution for the number of deaths in the United States due to pneumonia in a recent year for various age groups. Note that the last age group is "85 and older," an open-ended class.

Table 2.11 Deaths Due to Pneumonia

Age	Number of Deaths
5–14	69
15–24	178
25–34	299
35–44	875
45–54	1872
55–64	3099
65–74	6283
75–84	17,775
85 and older	27,758

Source: U.S. Census Bureau

When a frequency distribution contains an open-ended class, a histogram cannot be drawn.

Histograms for discrete data

When data are discrete, we can construct a frequency distribution in which each possible value of the variable forms a class. Then we can draw a histogram in which each rectangle represents one possible value of the variable. Table 2.12 (page 57) presents the results of a hypothetical survey in which 1000 adult women were asked how many children they had. Number of children is a discrete variable, and in this data set, the values of this variable are 0 through 8. To construct a histogram, we draw rectangles of equal width, centered at the values of the variables. The rectangles should be just wide enough to touch. Figure 2.9 (page 57) presents a histogram.

Table 2.12 Women with a Given Number of Children

Number of Children	Frequency
0	435
1	175
2	222
3	112
4	38
5	9
6	7
7	0
8	2

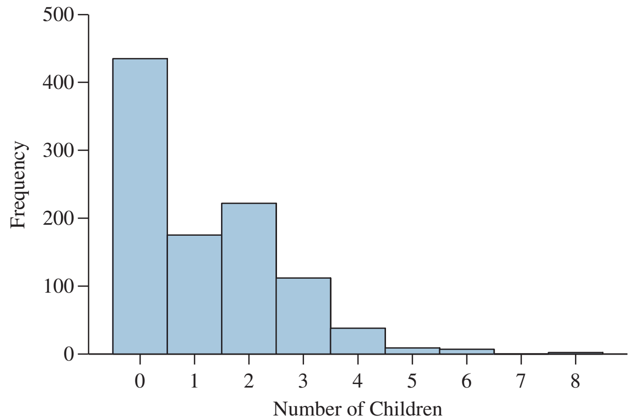


Figure 2.9 Histogram for data in Table 2.12

Objective 3 Determine the shape of a distribution from a histogram

Shapes of Histograms

The purpose of a histogram is to give a visual impression of the "shape" of a data set. Statisticians have developed terminology to describe some of the commonly observed shapes. A histogram is **skewed** if one side, or tail, is longer than the other. A histogram with a long right-hand tail is said to be **skewed to the right**, or **positively skewed**. A histogram with a long left-hand tail is said to be **skewed to the left**, or **negatively skewed**. A histogram is **symmetric** if its right half is a mirror image of its left half. Very few histograms are perfectly symmetric, but many are approximately symmetric. There are two special cases of symmetric histograms. A symmetric histogram with a peak in the middle is referred to as a **bell-shaped** histogram. A histogram in which all the classes have equal frequencies is said to be **uniformly distributed**. These terms apply to both frequency histograms and relative frequency histograms. Figure 2.10 presents some histograms for hypothetical samples. As another example, the histogram for particulate concentration, shown in Figure 2.5, is skewed to the right.

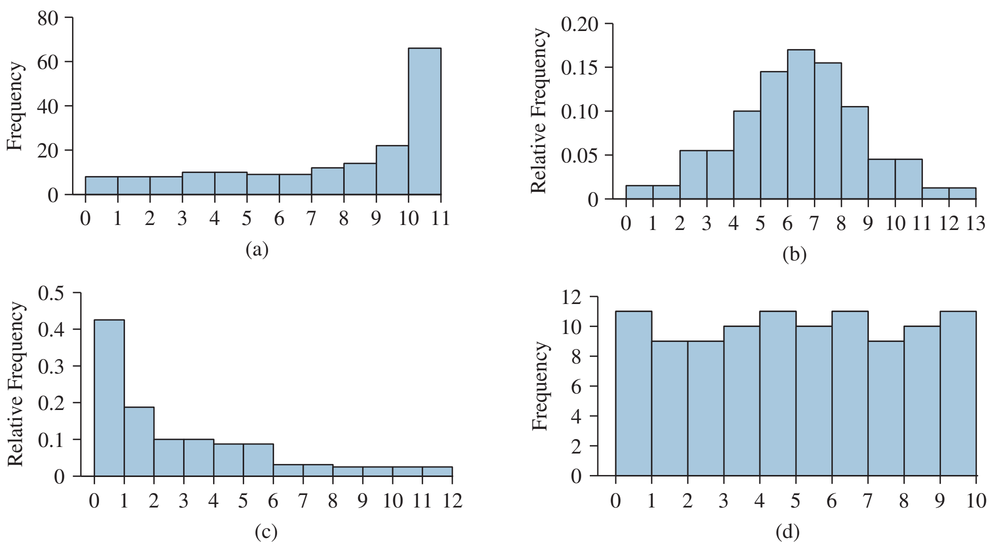


Figure 2.10 (a) A histogram skewed to the left. (b) A bell-shaped histogram. (c) A histogram skewed to the right. (d) An approximately uniformly distributed histogram.

The examples in Figure 2.10 are straightforward to categorize. In real life, the classification is not always clear-cut, and people may sometimes disagree on how to describe the shape of a particular histogram.

Modes

A peak, or high point, of a histogram is referred to as a **mode**. A histogram is **unimodal** if it has only one mode, and **bimodal** if it has two clearly distinct modes. In principle, a histogram can have more than two modes, but this does not happen often in practice. The histograms in Figure 2.10(a)–(c) are unimodal. Figure 2.11 presents a bimodal histogram for a hypothetical sample.

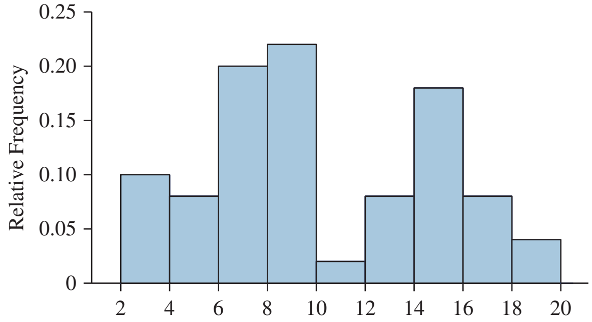


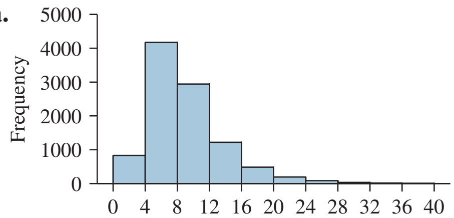
Figure 2.11 A bimodal histogram

As another example, it is reasonable to classify the histogram for particulate emissions, shown in Figure 2.5, as unimodal, with the rectangle above the class 1–2 as the only mode. While some might say that the rectangle above the class 3–4 is another mode, most would agree that it is too small a peak to count as a second mode.

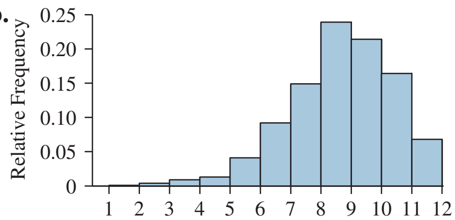
Check Your Understanding

3. Classify each of the following histograms as skewed to the left, skewed to the right, approximately bell-shaped, or approximately uniformly distributed.

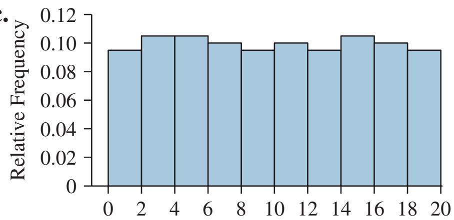
a.



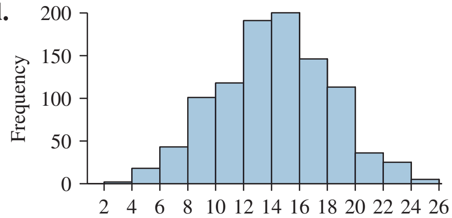
b.



c.

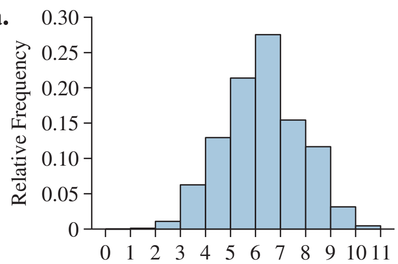


d.

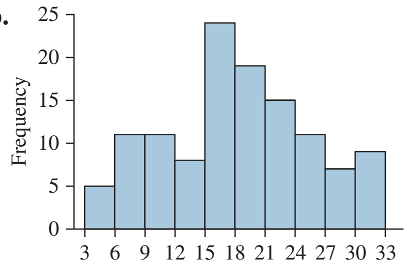


4. Classify each of the following histograms as unimodal or bimodal.

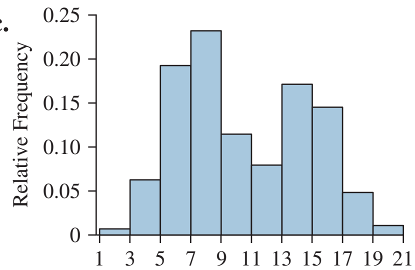
a.



b.



c.



Answers are on page 66.

USING TECHNOLOGY

We use the data in Table 2.7 to illustrate the technology steps.

TI-84 PLUS

Entering Data

Step 1. We enter the data into **L1** in the data editor. To clear out any data that may be in the list, press **STAT**, then **4: ClrList**, then enter **L1** by pressing **2nd, L1** (Figure A). Then press **ENTER**.

Step 2. Enter the data into **L1** in the data editor by pressing **STAT** then **1: Edit...** For the data in Table 2.7, we begin with **1.5, .87, 1.12, 1.25, 3.46, ...** (Figure B).

Figure A

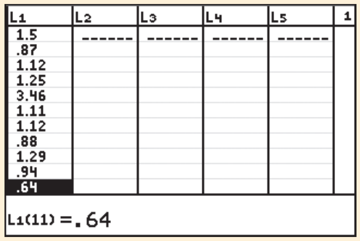


Figure B

Constructing a Histogram

Step 1. Press **2nd, Y=** to access the STAT PLOTS menu and select **Plot1** by pressing **1**.

Step 2. Select **On** and the histogram icon (Figure C).

Step 3. Press **WINDOW** and:

- Set **Xmin** to the lower class limit of the first class. We use 0 for our example.
- Set **Xmax** to the lower class limit of the class following the one containing the largest data value. We use 7.
- Set **Xscl** to the class width. We use 1.
- Set **Ymin** to 0.
- Set **Ymax** to a value greater than the largest frequency of all classes. We use 30.

Step 4. Press **GRAPH** to view the histogram (Figure D).

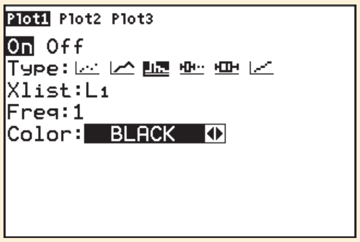


Figure C

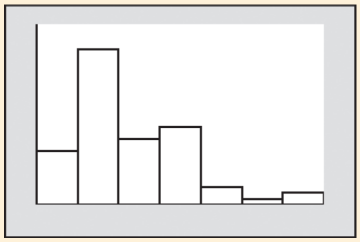


Figure D

EXCEL

Constructing a Histogram

Step 1. Enter the *Particulate Emissions* data from Table 2.7 in **Column A**.

Step 2. Press **Data**, then **Data Analysis**. Select **Histogram** and click **OK**.

Step 3. Enter the range of cells that contain the data in the **Input Range** field, and check the **Chart Output** box.

Step 4. Click **OK**.

MINITAB

Constructing a Histogram

Step 1. Name your variable *Particulate Emissions* and enter the data from Table 2.7 into Column **C1**.

Step 2. Click on **Graph**. Select **Histogram**. Choose the **Simple** option. Press **OK**.

Step 3. Double-click on the *Particulate Emissions* variable and press **OK** (Figure E).

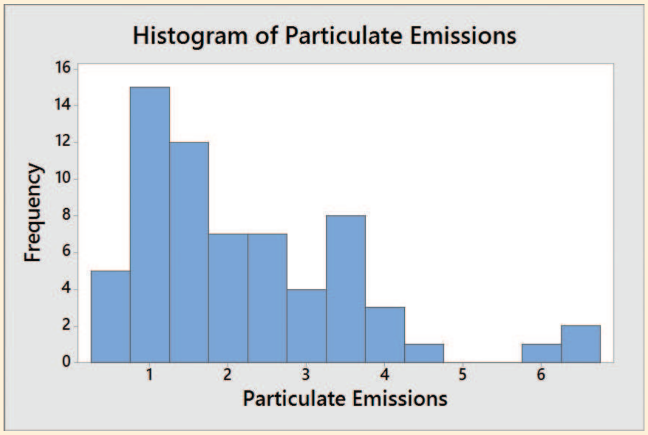


Figure E

SECTION 2.2 Exercises

Exercises 1–4 are the Check Your Understanding exercises located within the section.

Understanding the Concepts

In Exercises 5–8, fill in each blank with the appropriate word or phrase.

5. When the right half of a histogram is a mirror image of the left half, the histogram is ________ .
6. A histogram is skewed to the left if its ________ tail is longer than its ________ tail.
7. A histogram is ________ if it has two clearly distinct modes.
8. The ________ of a class is the sum of the frequencies of that class and all previous classes.

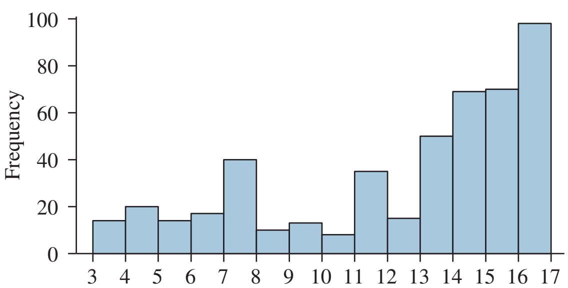
In Exercises 9–12, determine whether the statement is true or false. If the statement is false, rewrite it as a true statement.

9. In a frequency distribution, the class width is the difference between the upper and lower class limits.
10. The number of classes used has little effect on the shape of the histogram.
11. There is no one right way to choose the classes for a frequency distribution.
12. A mode occurs at the peak of a histogram.

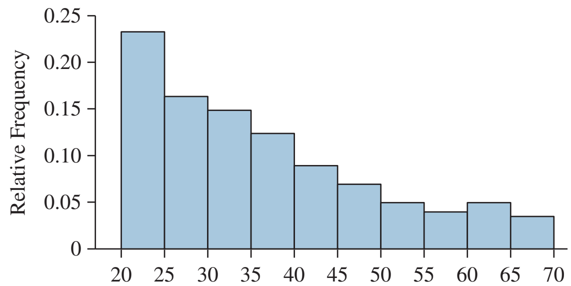
Practicing the Skills

In Exercises 13–16, classify the histogram as skewed to the left, skewed to the right, bell-shaped, or approximately uniform.

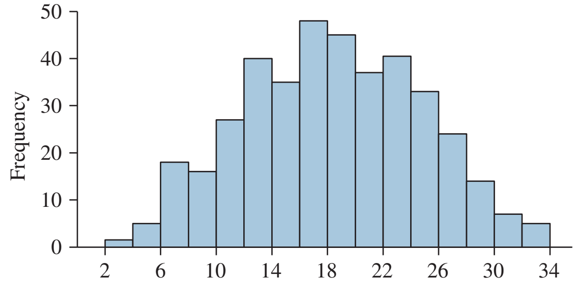
13.



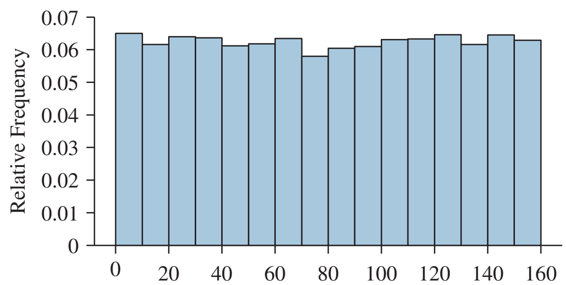
14.



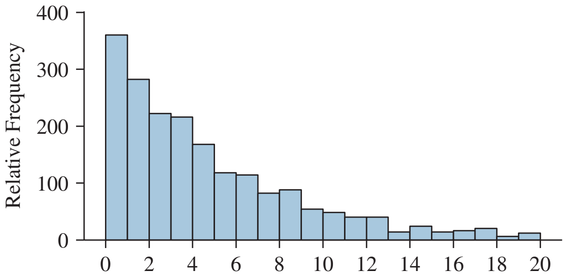
15.



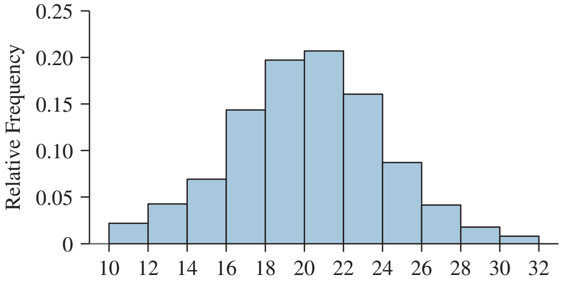
16.



17.

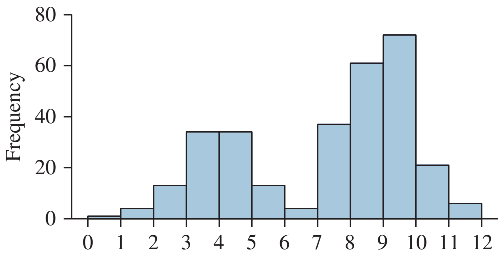


18.

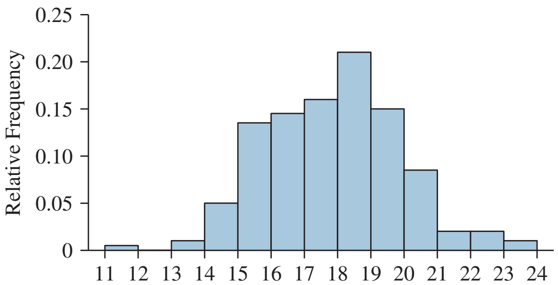


In Exercises 19 and 20, classify the histogram as unimodal or bimodal.

19.

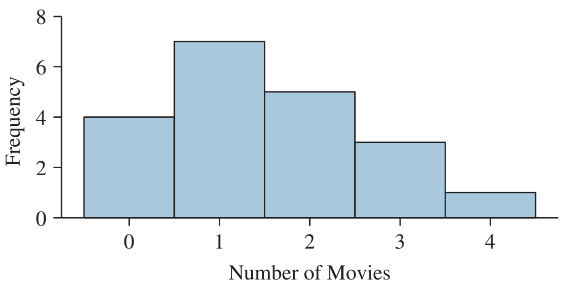


20.



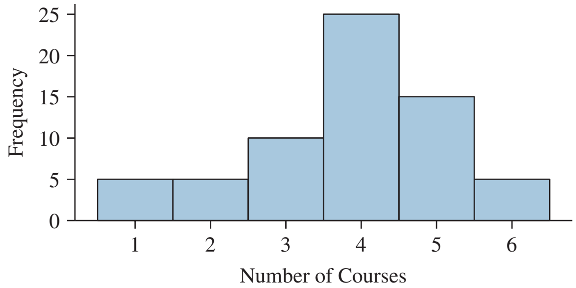
Working with the Concepts

21. Movies: A sample of people were asked how many movies they had seen in a movie theater during the past month. Following is a frequency histogram of their responses.



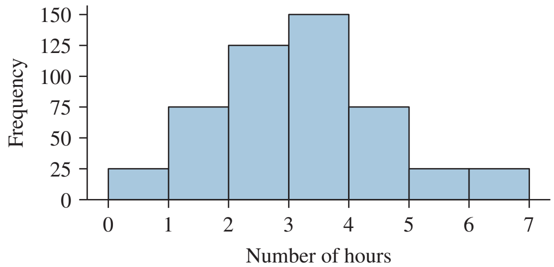
a. What is the most frequent number of movies seen?
b. How many people were in the sample?
c. How many people saw more than two movies?
d. What percentage of the people saw fewer than two movies?

22. College courses: The following frequency histogram presents the number of courses taken by a sample of students in a certain semester.



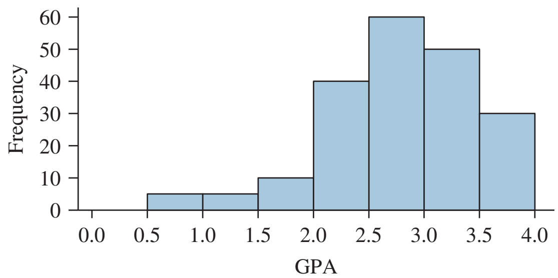
a. What is the most frequent number of courses taken?
b. How many students were in the sample?
c. How many students were taking more than four courses?
d. What percentage of the students were taking three or fewer courses?

23. Time on social media: A survey was conducted to determine the average number of hours per week people spend on social media. The following frequency histogram presents the results.



a. Which class has the highest frequency?
b. How many people were in the sample?
c. How many people spend more than 5 hours per week on social media?
d. What percentage of the people spend less than 2 hours per week on social media?

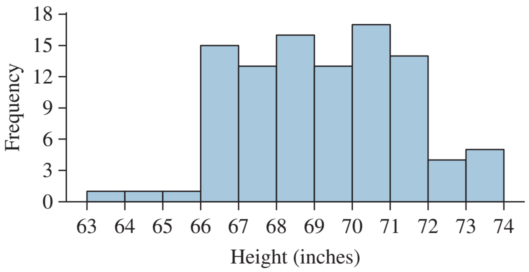
24. What's your GPA? The following frequency histogram presents the GPAs of a sample of students at a certain college.



a. Which class has the highest frequency?
b. How many students were in the sample?
c. How many students had GPAs above 3.0?

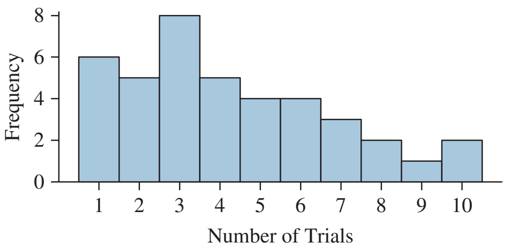
d. What percentage of the students had GPAs between 2.0 and 3.0?

25. Student heights: The following frequency histogram presents the heights, in inches, of a random sample of 100 male college students.



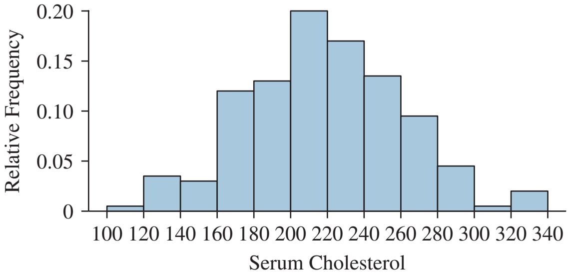
a. How many classes are there?
b. What is the class width?
c. Which class has the highest frequency?
d. What percentage of students are more than 72 inches tall?
e. Is the histogram most accurately described as skewed to the right, skewed to the left, or approximately symmetric?

26. Trained rats: Forty rats were trained to run a maze. The following frequency histogram presents the numbers of trials it took each rat to learn the maze.

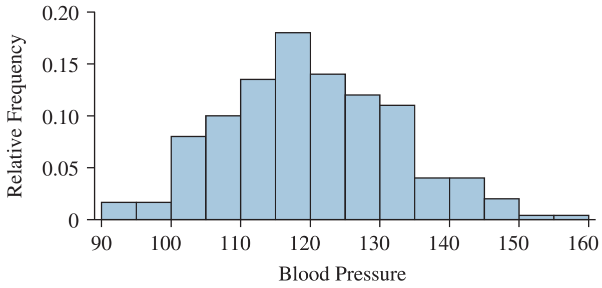


a. What is the most frequent number of trials?
b. How many rats learned the maze in three trials or less?
c. How many rats took nine trials or more to learn the maze?
d. Is the histogram most accurately described as skewed to the right, skewed to the left, or approximately symmetric?

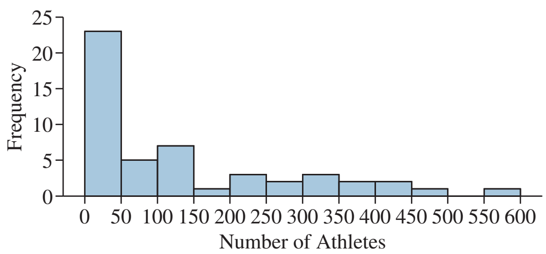
27. Cholesterol: The following histogram shows the distribution of serum cholesterol level (in milligrams per deciliter) for a sample of men. Use the histogram to answer the following questions:
a. Is the percentage of men with cholesterol levels above 240 closest to 30%, 50%, or 70%?
b. In which interval are there more men: 240–260 or 280–340?



28. Blood pressure: The following histogram shows the distribution of systolic blood pressure (in millimeters of mercury) for a sample of women. Use the histogram to answer the following questions:
a. Is the percentage of women with blood pressures above 120 closest to 25%, 50%, or 75%?
b. In which interval are there more women: 130–135 or 140–150?

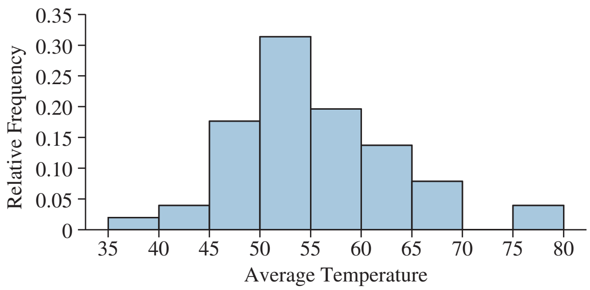


29. Olympic athletes: The following frequency histogram presents the number of athletes sent to a recent Olympic games by the 50 most populous countries.



a. True or false: More than half of the countries sent fewer than 50 athletes.
b. True or false: Fewer than 15 countries sent more than 300 athletes.
c. How many countries sent 300 or more athletes?
d. The United States sent 554 athletes. Did any other country send more than 500?
e. Is the histogram most accurately described as skewed to the right, skewed to the left, or approximately symmetric?

30. How's the weather? The following relative frequency histogram presents the average temperatures, in °F, for each of the 50 states of the United States, plus the District of Columbia. Use the histogram to answer the following questions.



a. In which interval are the largest number of states?
b. Is the percentage of states with average temperatures above 60° closest to 30%, 40%, or 50%?

c. In which interval are there more states, 55–60 or 60–70?

31. Skewed which way? For which of the following data sets would you expect a histogram to be skewed to the right? For which would it be skewed to the left?
a. The lengths of the words in a book
b. Dates of coins in circulation
c. Scores of students on an easy exam

32. Skewed which way? For which of the following data sets would you expect a histogram to be skewed to the right? For which would it be skewed to the left?
a. Annual incomes for residents of a town
b. Amounts of time taken by students on a one-hour exam
c. Ages of residents of a town

33. Batting average: The following frequency distribution presents the batting averages of Major League Baseball players who had 300 or more plate appearances during a recent season.

Batting Average	Frequency
0.180–0.199	4
0.200–0.219	14
0.220–0.239	41
0.240–0.259	59
0.260–0.279	58
0.280–0.299	51
0.300–0.319	29
0.320–0.339	10
0.340–0.359	1

Source: sports.espn.go.com

a. How many classes are there?
b. What is the class width?
c. What are the class limits?
d. Construct a frequency histogram.
e. Construct a relative frequency distribution.
f. Construct a relative frequency histogram.
g. What percentage of players had batting averages of 0.300 or more?
h. What percentage of players had batting averages less than 0.220?

34. Batting average: The following frequency distribution presents the batting averages of Major League Baseball players in both the American League and the National League who had 300 or more plate appearances during a recent season.

Batting Average	American League Frequency	National League Frequency
0.180–0.199	2	2
0.200–0.219	7	7
0.220–0.239	21	20
0.240–0.259	30	29
0.260–0.279	26	32
0.280–0.299	21	30
0.300–0.319	12	17
0.320–0.339	5	5
0.340–0.359	0	1

Source: sports.espn.go.com

a. Construct a frequency histogram for the American League.
b. Construct a frequency histogram for the National League.
c. Construct a relative frequency distribution for the American League.
d. Construct a relative frequency distribution for the National League.
e. Construct a relative frequency histogram for the American League.
f. Construct a relative frequency histogram for the National League.
g. What percentage of American League players had batting averages of 0.300 or more?
h. What percentage of National League players had batting averages of 0.300 or more?
i. Compare the relative frequency histograms. What is the main difference between the distributions of batting averages in the two leagues?

35. Time spent playing video games: A sample of 200 college freshmen was asked how many hours per week they spent playing video games. The following frequency distribution presents the results.

Number of Hours	Frequency
1.0–3.9	25
4.0–6.9	34
7.0–9.9	48
10.0–12.9	29
13.0–15.9	23
16.0–18.9	17
19.0–21.9	13
22.0–24.9	7
25.0–27.9	3
28.0–30.9	1

a. How many classes are there?
b. What is the class width?
c. What are the class limits?
d. Construct a frequency histogram.
e. Construct a relative frequency distribution.
f. Construct a relative frequency histogram.
g. What percentage of students play video games less than 10 hours per week?
h. What percentage of students play video games 19 or more hours per week?

36. Murder, she wrote: The following frequency distribution presents the number of murders (including negligent manslaughter) per 100,000 population for each U.S. city with population over 250,000 in a recent year.

Murder Rate	Frequency
0.0–4.9	21
5.0–9.9	23
10.0–14.9	12
15.0–19.9	6
20.0–24.9	5
25.0–29.9	0
30.0–34.9	2
35.0–39.9	2
40.0–44.9	0
45.0–49.9	0
50.0–54.9	2

Source: Federal Bureau of Investigation

a. How many classes are there?
b. What is the class width?
c. What are the class limits?
d. Construct a frequency histogram.

e. Construct a relative frequency distribution.
f. Construct a relative frequency histogram.
g. What percentage of cities had murder rates less than 10 per 100,000 population?
h. What percentage of cities had murder rates of 30 or more per 100,000 population?

37. BMW prices: The following table presents the manufacturer's suggested retail price (in $1000s) for base models and styles of BMW automobiles.

35.3	37.3	41.1	43.1	45.0	47.0
53.1	55.1	44.8	46.8	51.4	53.4
57.7	60.4	51.2	53.2	86.5	89.5
95.6	102.7	58.9	45.8	47.8	50.4
52.4	54.0	56.0	69.2	77.7	103.1
157.7	69.9	76.9	73.4	80.4	63.7

Source: Car Finder

a. Construct a frequency distribution using a class width of 10, and using 30 as the lower class limit for the first class.
b. Construct a frequency histogram from the frequency distribution in part (a).
c. Construct a relative frequency distribution using the same class width and lower limit for the first class.
d. Construct a relative frequency histogram.
e. Are the histograms unimodal or bimodal?
f. Repeat parts (a)–(d), using a class width of 20, and using 30 as the lower class limit for the first class.
g. Do you think that class widths of 10 and 20 are both reasonably good choices for these data, or do you think that one choice is much better than the other? Explain your reasoning.

38. Geysers: The geyser Old Faithful in Yellowstone National Park alternates periods of eruption, which typically last from 1.5 to 4 minutes, with periods of dormancy, which are considerably longer. The following table presents the durations, in minutes, of 60 dormancy periods that occurred during a recent year.

91	99	99	83	99	85	90	96	88	93
88	88	92	116	59	101	90	71	103	97
82	91	89	89	94	94	61	96	66	105
90	93	88	92	86	93	95	83	90	99
89	94	90	95	93	105	96	92	101	91
94	92	94	86	88	99	90	99	84	92

a. Construct a frequency distribution using a class width of 5, and using 55 as the lower class limit for the first class.
b. Construct a frequency histogram from the frequency distribution in part (a).
c. Construct a relative frequency distribution using the same class width and lower limit for the first class.
d. Construct a relative frequency histogram.
e. Are the histograms skewed to the left, skewed to the right, or approximately symmetric?
f. Repeat parts (a)–(d), using a class width of 10, and using 50 as the lower class limit for the first class.
g. Do you think that class widths of 5 and 10 are both reasonably good choices for these data, or do you think that one choice is much better than the other? Explain your reasoning.

39. Hail to the chief: There have been 58 presidential inaugurations in U.S. history. At each one, the president has made an inaugural address. Following are the number of words spoken in each of these addresses.

1431	135	2321	1730	2166	1177	1211	3375
4472	2915	1128	1176	3843	8460	4809	1090
3336	2831	3637	700	1127	1339	2486	2979
1686	4392	2015	3968	2218	984	5434	1704
1526	3329	4055	3672	1880	1808	1359	559
2273	2459	1658	1366	1507	2128	1803	1229
2427	2561	2320	1598	2155	1592	2071	2395
2096	1433						

Source: The American Presidency Project

a. Construct a frequency distribution with approximately five classes.
b. Construct a frequency histogram from the frequency distribution in part (a).
c. Construct a relative frequency distribution using the same classes as in part (a).
d. Construct a relative frequency histogram from this relative frequency distribution.
e. Are the histograms skewed to the left, skewed to the right, or approximately symmetric?
f. Construct a frequency distribution with approximately nine classes.
g. Repeat parts (b)–(d), using the frequency distribution constructed in part (f).
h. Do you think that five and nine classes are both reasonably good choices for these data, or do you think that one choice is much better than the other? Explain your reasoning.

40. Internet radio: The following table presents the number of hours a sample of 40 subscribers listened to Pandora Radio in a given week.

52	18	2	20	9	9	11	6	18	6
4	12	9	16	10	37	15	18	8	23
4	3	17	19	12	20	11	14	10	37
21	36	17	3	23	28	19	20	29	12

a. Construct a frequency distribution with approximately eleven classes.
b. Construct a frequency histogram from this frequency distribution.
c. Construct a relative frequency distribution using the same classes.
d. Construct a relative frequency histogram from this relative frequency distribution.
e. Are the histograms skewed to the left, skewed to the right, or approximately symmetric?
f. Construct a frequency distribution with approximately four classes.
g. Repeat parts (b)–(d), using the frequency distribution constructed in part (f).
h. Do you think that four and eleven classes are both reasonably good choices for these data, or do you think that one choice is much better than the other? Explain your reasoning.

41. Brothers and sisters: Thirty students in a first-grade class were asked how many siblings they have. Following are the results.

1	1	2	1	2	3	7	1	1	5
1	1	3	0	1	1	1	2	5	0
0	1	2	2	4	2	2	3	3	4

a. Construct a frequency histogram.
b. Construct a relative frequency histogram.
c. Are the histograms skewed to the left, skewed to the right, or approximately symmetric?

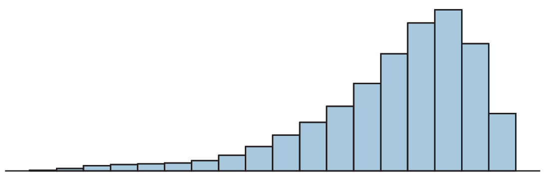
42. Cough, cough: The following table presents the number of days a sample of patients reported a cough lasting from an acute cough illness.

16	20	21	20	19	17	18	12	22
17	16	24	15	13	21	16	21	20
20	19	18	16	19	20	21	21	14
16	16	14	17	18	14	15	20	20
13	15	18	20	19	21	19	18	20

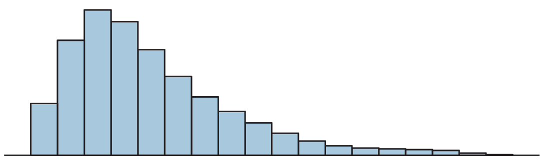
Based on data from *Annals of Family Medicine*

a. Construct a frequency histogram.
b. Construct a relative frequency histogram.
c. Are the histograms skewed to the left, skewed to the right, or approximately symmetric?

43. Which histogram is which? One of the following histograms represents the age at death from natural causes (heart attack, cancer, etc.), and the other represents the age at death from accidents. Which represents the age at death from accidents? How can you tell?

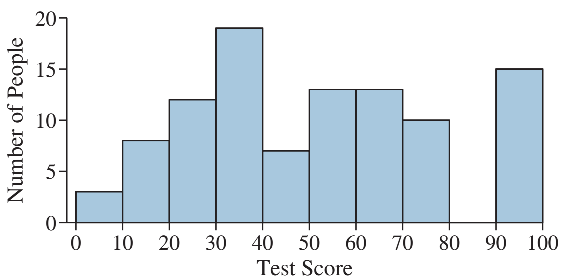


(A)



(B)

44. Test scores: In a certain city, applicants for engineering jobs are given an exam, and the ones with the top 15 scores are hired. The following frequency histogram presents the scores for the applicants on a recent exam. The examiners were charged with rigging the exam. Describe the evidence that supports this charge.



45. No histogram possible: A company surveyed 100 employees to find out how far they travel in their commute to work. The results are presented in the following frequency distribution.

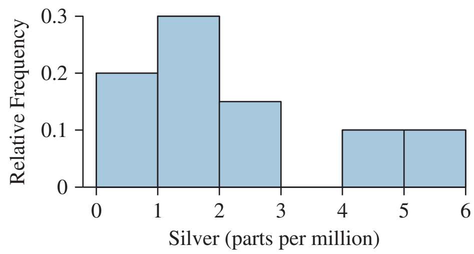
Distance in Miles	Frequency
0.0–4.9	18
5.0–9.9	26
10.0–14.9	15
15.0–19.9	13
20.0–24.9	12
25.0–29.9	9
30 or more	7

Explain why it is not possible to construct a histogram for this data set.

46. Histogram possible? Refer to Exercise 45. Suppose you found out that none of the employees traveled more than 34 miles. Would it be possible to construct a histogram? If so, construct a histogram. If not, explain why not.

Extending the Concepts

47. Silver ore: The following histogram presents the amounts of silver (in parts per million) found in a sample of rocks. One rectangle from the histogram is missing. What is its height?



48. Classes of differing widths: Consider the following relative frequency distribution for the data in Table 2.7, in which the classes have differing widths.

Class	Frequency	Relative Frequency
0.00–0.99	9	0.138
1.00–1.49	19	0.292
1.50–1.99	7	0.108
2.00–2.99	11	0.169
3.00–3.99	13	0.200
4.00–6.99	6	0.092

a. Compute the class width for each of the classes.
b. Construct a relative frequency histogram. Compare it to the relative frequency histogram in Figure 2.6, in which the classes all have the same width. Explain why using differing widths gives a distorted picture of the data.
c. The *density* of a class is the relative frequency divided by the class width. For each class, divide the relative frequency by the class width to obtain the density.
d. Construct a histogram in which the height of each rectangle is equal to the density of the class. This is called a *density histogram*.
e. Compare the density histogram to the relative frequency histogram in Figure 2.6, in which the classes all have the same width. Explain why differing class widths in a density histogram do not distort the data.

49. Frequencies and relative frequencies: The following relative frequency histogram presents the hourly wages for employees at a certain company.

Following are three tables showing hourly wages and the number of employees earning that wage. For each table, say whether the histogram is a correct relative frequency histogram for that table.

Wage	Number of Employees
12	10
14	5
17	10
18	10
22	10
26	3
28	2

(A)

Wage	Number of Employees
11	9
16	3
18	7
19	2
23	6
26	2
29	1

(B)

Wage	Number of Employees
14	10
16	5
17	10
18	10
22	10
26	3
28	2

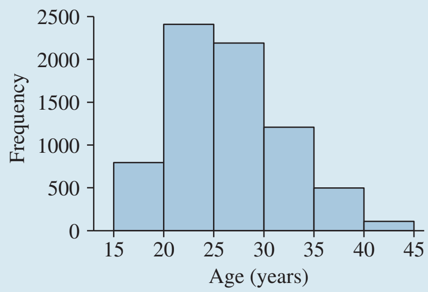
(C)

Answers to Check Your Understanding Exercises for Section 2.2

1.

Class	Frequency
0.00–0.49	2
0.50–0.99	7
1.00–1.49	19
1.50–1.99	7
2.00–2.49	7
2.50–2.99	4
3.00–3.49	9
3.50–3.99	4
4.00–4.49	2
4.50–4.99	1
5.00–5.49	0
5.50–5.99	1
6.00–6.49	0
6.50–6.99	2

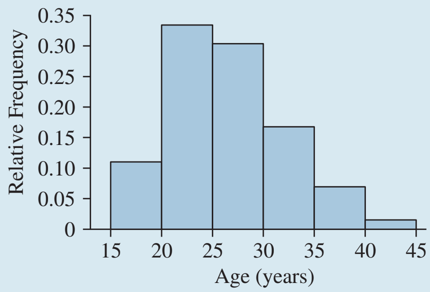
2. a. 15, 20, 25, 30, 35, 40 **b.** 5 **c.**



d.

Age	Frequency	Relative Frequency
15–19	795	0.110
20–24	2410	0.334
25–29	2190	0.304
30–34	1208	0.168
35–39	499	0.069
40–44	109	0.015

e.



3. a. Skewed right **b.** Skewed left **c.** Approximately uniform **d.** Bell-shaped

4. a. Unimodal **b.** Unimodal **c.** Bimodal

SECTION 2.3 More Graphs for Quantitative Data

Objectives

1. Construct stem-and-leaf plots
2. Construct dotplots
3. Construct time-series plots

Histograms and other graphs that are based on frequency distributions can be used to summarize both small and large data sets. For small data sets, it is sometimes useful to have a summary that is more detailed than a histogram. In this section, we describe some commonly used graphs that provide more detailed summaries of smaller data sets. These graphs illustrate the shape of the data set, while allowing every value in the data set to be seen.

Objective 1 Construct stem-and-leaf plots

Stem-and-Leaf Plots

Stem-and-leaf plots are a simple way to display small data sets. For example, Table 2.13 presents the U.S. Census Bureau data for the percentage of the population aged 65 and over for each state and the District of Columbia.

In a stem-and-leaf plot, the rightmost digit is the leaf, and the remaining digits form the stem. For example, the stem for Alabama is 14, and the leaf is 1. We construct a stem-and-leaf plot for the data in Table 2.13 by using the following three-step process:

Table 2.13 Percentage of Population Aged 65 and Over, by State

Alabama	14.1	Alaska	8.1	Arizona	13.9
Arkansas	14.3	California	11.5	Colorado	10.7
Connecticut	14.4	Delaware	14.1	District of Columbia	11.5
Florida	17.8	Georgia	10.2	Hawaii	14.3
Idaho	12.0	Illinois	12.4	Indiana	12.7
Iowa	14.9	Kansas	13.4	Kentucky	13.1
Louisiana	12.6	Maine	15.6	Maryland	12.2
Massachusetts	13.7	Michigan	12.8	Minnesota	12.4
Mississippi	12.8	Missouri	13.9	Montana	15.0
Nebraska	13.8	Nevada	12.3	New Hampshire	12.6
New Jersey	13.7	New Mexico	14.1	New York	13.6
North Carolina	12.4	North Dakota	15.3	Ohio	13.7
Oklahoma	13.8	Oregon	13.0	Pennsylvania	15.5
Rhode Island	14.1	South Carolina	13.6	South Dakota	14.6
Tennessee	13.3	Texas	10.5	Utah	9.0
Vermont	14.3	Virginia	12.4	Washington	12.2
West Virginia	16.0	Wisconsin	13.5	Wyoming	14.0

Source: U.S. Census Bureau

Step 1: Make a vertical list of all the stems in increasing order, and draw a vertical line to the right of this list. The smallest stem in Table 2.13 is 8, belonging to Alaska, and the largest is 17, belonging to Florida. The list of stems is shown in Figure 2.12(a).

Step 2: Go through the data set, and for each value, write its leaf next to its stem. For example, the first value is 14.1, for Alabama. We write a "1" next to the stem 14. The next value is 8.1 for Alaska, so we write a "1" next to the stem 8. When we are finished, we have the result shown in Figure 2.12(b).

Step 3: For each stem, arrange its leaves in increasing order. The result is the stem-and-leaf plot, shown in Figure 2.12(c).

8	
9	
10	
11	
12	
13	
14	
15	
16	
17	

(a)

8	1
9	0
10	7 2 5
11	5 5
12	0 4 7 6 2 8 4 8 3 6 4 4 2
13	9 4 1 7 9 8 7 6 7 8 0 6 3 5
14	1 3 4 1 3 9 1 1 6 3 0
15	6 0 3 5
16	0
17	8

(b)

8	1
9	0
10	2 5 7
11	5 5
12	0 2 2 3 4 4 4 4 6 6 7 8 8
13	0 1 3 4 5 6 6 7 7 7 8 8 9 9
14	0 1 1 1 1 3 3 3 4 6 9
15	0 3 5 6
16	0
17	8

(c)

Figure 2.12 Steps in the construction of a stem-and-leaf plot

Rounding data for a stem-and-leaf plot

Table 2.14 presents the particulate emissions for 65 vehicles. The first digits range from 0 to 6, and we would like to construct a stem-and-leaf plot with these digits as the stems. The problem is that this leaves two digits for the leaf, but the leaf must consist of only one digit. The solution to this problem is to round the data so that there will be only one digit for the leaf. Table 2.15 presents the particulate data rounded to two digits.

Table 2.14 Particulate Emissions for 65 Vehicles

1.50	0.87	1.12	1.25	3.46	1.11	1.12	0.88	1.29	0.94	0.64	1.31	2.49
1.48	1.06	1.11	2.15	0.86	1.81	1.47	1.24	1.63	2.14	6.64	4.04	2.48
1.40	1.37	1.81	1.14	1.63	3.67	0.55	2.67	2.63	3.03	1.23	1.04	1.63
3.12	2.37	2.12	2.68	1.17	3.34	3.79	1.28	2.10	6.55	1.18	3.06	0.48
0.25	0.53	3.36	3.47	2.74	1.88	5.94	4.24	3.52	3.59	3.10	3.33	4.58

Table 2.15 Particulate Emissions for 65 Vehicles, Rounded to Two Digits

1.5	0.9	1.1	1.3	3.5	1.1	1.1	0.9	1.3	0.9	0.6	1.3	2.5
1.5	1.1	1.1	2.2	0.9	1.8	1.5	1.2	1.6	2.1	6.6	4.0	2.5
1.4	1.4	1.8	1.1	1.6	3.7	0.6	2.7	2.6	3.0	1.2	1.0	1.6
3.1	2.4	2.1	2.7	1.2	3.3	3.8	1.3	2.1	6.6	1.2	3.1	0.5
0.3	0.5	3.4	3.5	2.7	1.9	5.9	4.2	3.5	3.6	3.1	3.3	4.6

We now follow the three-step process to obtain the stem-and-leaf plot. The result is shown in Figure 2.13.

0	3 5 5 6 6 9 9 9 9
1	0 1 1 1 1 1 1 2 2 2 2 3 3 3 3 4 4 5 5 5 6 6 6 8 8 9
2	1 1 1 2 4 5 5 6 7 7 7
3	0 1 1 1 3 3 4 5 5 5 6 7 8
4	0 2 6
5	9
6	6 6

Figure 2.13 Stem-and-leaf plot for the data in Table 2.15

Split stems

Sometimes one or two stems contain most of the leaves. When this happens, we often use two or more lines for each stem. The plot is then called a **split stem-and-leaf plot**. We will use the data in Table 2.16 to illustrate the method. These data consist of scores on a final examination in a statistics class, arranged in order.

Table 2.16 Scores on a Final Examination

58	66	68	70	70	71	71	72	73	73
75	76	78	78	79	80	80	80	81	82
82	82	82	83	84	86	86	86	87	88
89	89	89	90	92	93	95	97		

Figure 2.14 presents a stem-and-leaf plot for these data, using the stems 5, 6, 7, 8, and 9.

5	8
6	6 8
7	0 0 1 1 2 3 3 5 6 8 8 9
8	0 0 0 1 2 2 2 2 3 4 6 6 6 7 8 9 9 9
9	0 2 3 5 7

Figure 2.14 Stem-and-leaf plot for the data in Table 2.16

Most of the leaves are on two stems, 7 and 8. For this reason, the stem-and-leaf plot does not reveal much detail about the data. To remedy this situation, we will assign each stem two lines on the plot instead of one. Leaves with values 0–4 will go on the first line, and leaves with values 5–9 will go on the second line. So, for example, we will do the following with the stem 7:

7 | 0 0 1 1 2 3 3 5 6 8 8 9   will become   7 | 0 0 1 1 2 3 3
7 | 5 6 8 8 9

The split stem-and-leaf plot is shown in Figure 2.15. Note that every stem is given two lines, even those that have only a few leaves. Each stem in a split stem-and-leaf plot must receive the same number of lines.

CAUTION

In a split stem-and-leaf plot, each stem must be given the same number of lines.

5	
5	8
6	
6	6 8
7	0 0 1 1 2 3 3
7	5 6 8 8 9
8	0 0 0 1 2 2 2 2 3 4
8	6 6 6 7 8 9 9 9
9	0 2 3
9	5 7

Figure 2.15 Split stem-and-leaf plot for the data in Table 2.16

Check Your Understanding

1. **Weights of college students:** The following table presents weights in pounds for a group of male college freshmen.

136	163	157	195	150	149	151	155	163	145
124	124	156	148	195	192	133	129	160	158
166	155	171	157	182	124	160	172	161	143

a. List the stems for a stem-and-leaf plot.
b. For each item in the data set, write its leaf next to its stem.
c. Rearrange the leaves in numerical order to create a stem-and-leaf plot.

Answers are on page 78.

Back-to-back stem-and-leaf plots

When two data sets have values similar enough so that the same stems can be used, we can compare their shapes with a **back-to-back stem-and-leaf plot**. In a back-to-back stem-and-leaf plot, the stems go down the middle. The leaves for one of the data sets go off to the right, and the leaves for the other go off to the left.

EXAMPLE 2.14

Constructing a back-to-back stem-and-leaf plot

In Table 2.15, we presented particulate emissions for 65 vehicles. In a related experiment carried out at the Colorado School of Mines, particulate emissions were measured for 35 vehicles driven at high altitude. Table 2.17 presents the results. Construct a back-to-back stem-and-leaf plot to compare the emission levels of vehicles driven at high altitude with those of vehicles driven at sea level.

Table 2.17 Particulate Emissions for 35 Vehicles Driven at High Altitude

8.9	4.4	3.6	4.4	3.8	2.4	3.8	5.3	5.8	2.9	4.7	1.9	9.1
8.7	9.5	2.7	9.2	7.3	2.1	6.3	6.5	6.3	2.0	5.9	5.6	5.6
1.5	6.5	5.3	5.6	2.1	1.1	3.3	1.8	7.6				

Solution

Figure 2.16 presents the results. It is clear that vehicles driven at high altitude tend to produce higher emissions.

High Altitude		Sea Level
	0	3 5 5 6 6 9 9 9 9
9 8 5 1	1	0 1 1 1 1 1 1 2 2 2 2 3 3 3 3 4 4 5 5 5 6 6 6 8 8 9
9 7 4 1 1 0	2	1 1 1 2 4 5 5 6 7 7 7
8 8 6 3	3	0 1 1 1 3 3 4 5 5 5 6 7 8
7 4 4	4	0 2 6
9 8 6 6 6 3 3	5	9
5 5 3 3	6	6 6
6 3	7	
9 7	8	
5 2 1	9	

Figure 2.16 Back-to-back stem-and-leaf plots comparing the emissions in vehicles driven at high altitude with emissions from vehicles driven at sea level

Objective 2 Construct dotplots

Dotplots

A **dotplot** is a graph that can be used to give a rough impression of the shape of a data set. It is useful when the data set is not too large, and when there are some repeated values. As an example, Table 2.18 presents the number of children of each of the presidents of the United States and their wives.

Table 2.18 Numbers of Children of U.S. Presidents and Their Wives

0	2	10	2	5	3	6	2	2	4	1
5	4	15	3	4	5	3	2	3	4	2
6	0	0	0	8	3	3	6	2	4	2
0	4	6	4	7	2	0	1	2	6	5

Figure 2.17 presents a dotplot for the data in Table 2.18. For each value in the data set, a vertical column of dots is drawn, with the number of dots in the column equal to the number of times the value appears in the data set. The dotplot gives a good indication of where the values are concentrated and where the gaps are. For example, it is immediately apparent from Figure 2.17 that the most frequent number of children is 2, and only four presidents had more than 6. (John Tyler holds the record with 15.)

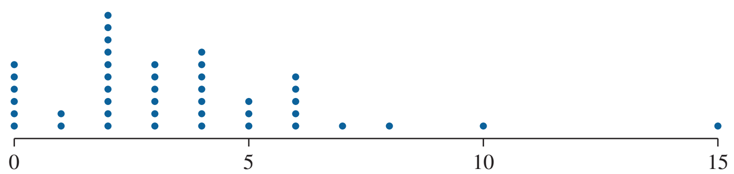


Figure 2.17 Dotplot for the data in Table 2.18

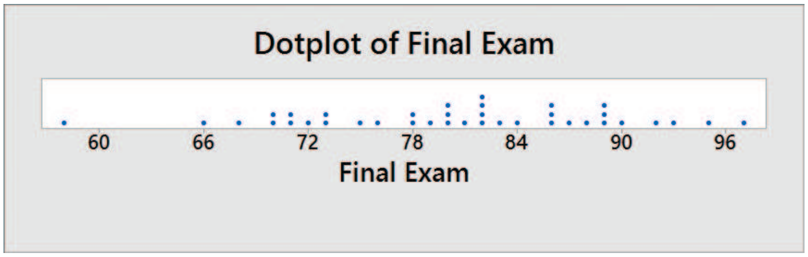
EXAMPLE 2.15

Constructing a dotplot with technology

Use technology to construct a dotplot for the exam score data in Table 2.16 on page 68.

Solution

The following figure shows the dotplot constructed in MINITAB. Step-by-step instructions for constructing dotplots with MINITAB are given in the Using Technology section on page 73.



Objective 3 Construct time-series plots

Time-Series Plots

A **time-series plot** may be used when the data consist of values of a variable measured at different points in time. As an example, we consider the Dow Jones Industrial Average, which reflects the prices of 30 large stocks. Table 2.19 presents the closing value of the Dow Jones Industrial Average at the end of each year from 2005 to 2018.

Table 2.19 Dow Jones Industrial Average

Year	Average
2005	10,717.50
2006	12,463.15
2007	13,264.82
2008	8,776.39
2009	10,428.05
2010	11,557.51
2011	12,217.56
2012	13,104.14
2013	16,576.66
2014	17,823.07
2015	17,425.03
2016	19,762.60
2017	24,719.22
2018	23,327.46

In a time-series plot, the horizontal axis represents time, and the vertical axis represents the value of the variable we are measuring. We plot the values of the variable at each of the times, then connect the points with straight lines. Example 2.16 shows how.

EXAMPLE 2.16

Constructing a time-series plot

Construct a time-series plot for the data in Table 2.19.

Solution

Step 1: Label the horizontal axis with the times at which measurements were made.

Step 2: Plot the value of the Dow Jones Industrial Average for each year.

Step 3: Connect the points with straight lines.

The result is shown in Figure 2.18 (page 72). It is clear that the average generally increased from 2005 to 2007, dropped sharply in 2008, generally increased from 2008 to 2017, and dropped in 2018.

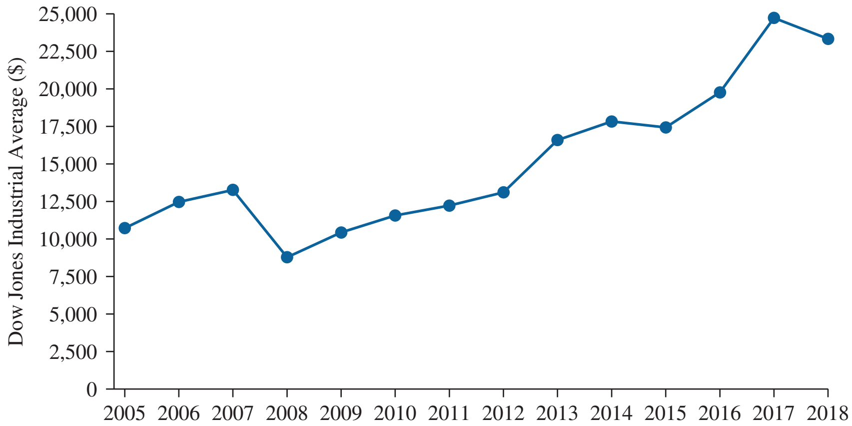
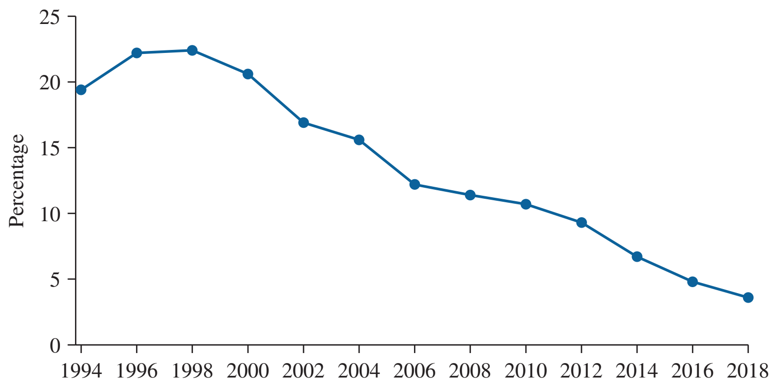


Figure 2.18

Check Your Understanding

2. The following time-series plot presents the percentage of high school seniors who smoked cigarettes every two years from 1994 through 2018.



Source: Department of Health and Human Services

a. In what year was the percentage the lowest?
b. What was the first year that the percentage was below 15%?
c. True or false: Since 1994, the percentage has never been lower than 5%.
d. During what period of time was the percentage decreasing?
e. During what period of time was the percentage increasing?

Answers are on page 78.

USING TECHNOLOGY

We use the data in Tables 2.16 and 2.19 to illustrate the technology steps.

EXCEL

Constructing a time-series plot

Step 1. Enter the data from Table 2.19 into **Columns A** and **B** (Figure A).
Step 2. Select the **Insert** tab and click on the **Line Graph** icon.
Step 3. Click on the **Select Data** icon. For the **Horizontal (Category) Axis Labels**, click on **Edit** and select the year values in **Column A** (Figure B).
Step 4. Click **OK** (Figure C).

	A	B
1	**Year**	**Average**
2	2005	10,717.50
3	2006	12,463.15
4	2007	13,264.82
5	2008	8,776.39
6	2009	10,428.05
7	2010	11,557.51
8	2011	12,217.56
9	2012	13,104.14

Figure A

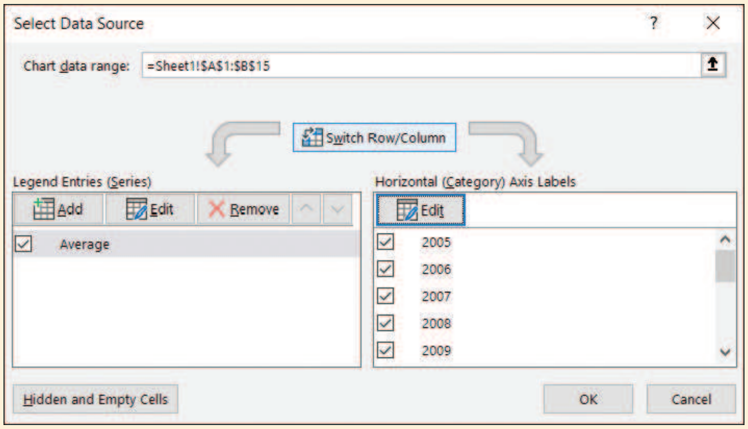


Figure B

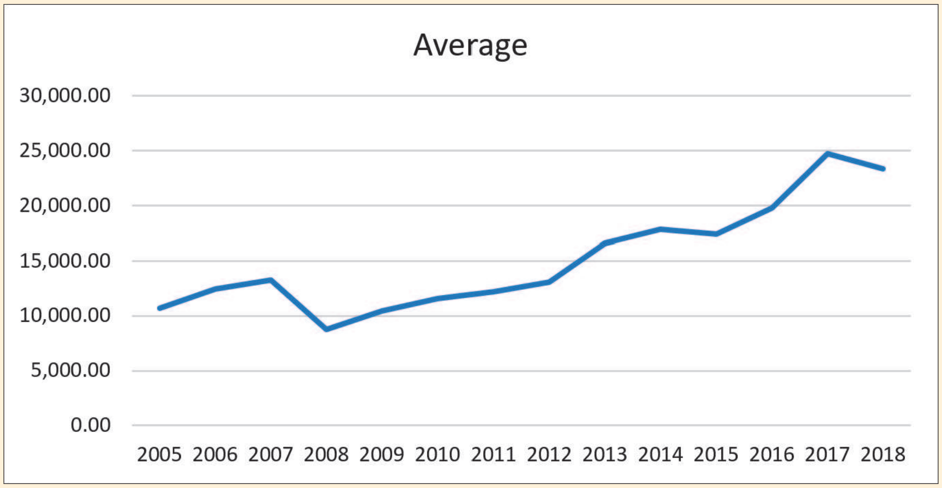


Figure C

MINITAB

Constructing a stem-and-leaf plot and dotplot

Step 1. Name your variable *Final Exam*, and enter the data from Table 2.16 into **Column C1**.

Step 2. Click on **Graph**. Select **Stem-and-Leaf** or **Dotplot**. For **Dotplot**, choose the **Simple** option. Press **OK**.

Step 3. Double-click on the *Final Exam* variable and press **OK** (See Figures D and E).

Stem-and-leaf of Scores N = 38

1	5	8
1	6	
3	6	68
10	7	0011233
15	7	56889
(10)	8	0001222234
13	8	66678999
5	9	023
2	9	57

Leaf Unit = 1

Figure D

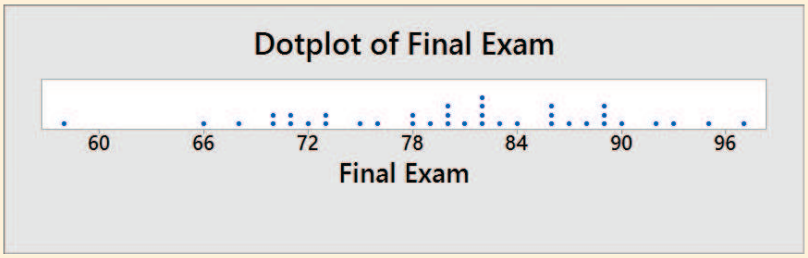


Figure E

SECTION 2.3 Exercises

Exercises 1 and 2 are the Check Your Understanding exercises located within the section.

Understanding the Concepts

In Exercises 3–6, fill in each blank with the appropriate word or phrase.

3. In a stem-and-leaf plot, the rightmost digit of each data value is the ________ .

4. In a back-to-back stem-and-leaf plot, each of the two data sets plotted must have the same ________ .

5. A ________ is useful when the data consist of values measured at different points in time.

6. In a time-series plot, the horizontal axis represents ________ .

In Exercises 7–10, determine whether the statement is true or false. If the statement is false, rewrite it as a true statement.

7. Stem-and-leaf plots and dotplots provide a simple way to display data for small data sets.

8. In a stem-and-leaf plot, each stem must be a single digit.

9. In a dotplot, the number of dots in a vertical column represents the number of times a certain value appears in a data set.

10. In a time-series plot, the vertical axis represents time.

Practicing the Skills

11. Construct a stem-and-leaf plot for the following data.

57	20	27	16	11	12	29	39	45	52	58	15
46	27	22	21	15	50	16	45	20	55	12	31

12. Construct a stem-and-leaf plot for the following data, in which the leaf represents the hundredths digit.

5.03	4.99	4.95	5.01	4.99	5.03	4.91	5.25	4.80
5.24	4.94	5.04	5.17	4.81	5.22	4.92	5.05	4.89
5.19	5.17	5.25	5.14	5.10	4.94	5.19	4.99	

13. List the data in the following stem-and-leaf plot. The leaf represents the ones digit.

3	0 0 1 2
3	5 6 7 7 9
4	2 3 4
4	5 6 7 7 7 7 8 8 9
5	0 1 1 1 2 2 2 2 4
5	6 7 8 8 9
6	1 3
6	

14. List the data in the following stem-and-leaf plot. The leaf represents the tenths digit.

14	4 6 8 9
15	1 2 2 4 5 7 7 8
16	0 1 1 1 2 3 7 7 9
17	
18	2 3 8

15. Construct a dotplot for the data in Exercise 11.

16. Construct a dotplot for the data in Exercise 12.

Working with the Concepts

17. BMW prices: The following table presents the manufacturers suggested retail price (in $1000s) for 2020 base models and styles of BMW automobiles.

35.3	37.3	41.1	43.1	45.0	47.0
53.1	55.1	44.8	46.8	51.4	53.4
57.7	60.4	51.2	53.2	86.5	89.5
95.6	102.7	58.9	45.8	47.8	50.4
52.4	54.0	56.0	69.2	77.7	103.1
157.7	69.9	76.9	73.4	80.4	63.7

Source: Car Finder

a. Round the data to the nearest whole number (round .5 up) and construct a stem-and-leaf plot, using the numbers 3 through 15 as the stems.

b. Repeat part (a), but split the stems, using two lines for each stem.

c. Which stem-and-leaf plot do you think is more appropriate for these data, the one in part (a) or the one in part (b)? Why?

18. How's the weather? The following table presents the daily high temperatures for the city of Macon, Georgia, in degrees Fahrenheit, for the winter months of January and February, 2019.

73	66	64	66	63	72	70	72	57	54	55	50
52	50	61	61	61	73	55	46	55	46	73	63
52	57	54	61	50	48	54	64	64	68	68	72
72	81	64	51	48	55	75	61	64	66	75	66
66	54	52	70	82	59	73	66	72	70	64	

a. Construct a stem-and-leaf plot, using the digits 4, 5, 6, 7, and 8 as the stems.

b. Repeat part (a), but split the stems, using two lines for each stem.

c. Which stem-and-leaf plot do you think is more appropriate for these data, the one in part (a) or the one in part (b)? Why?

19. Air pollution: The following table presents amounts of particulate emissions for 65 vehicles. These data also appear in Table 2.15.

1.5	0.9	1.1	1.3	3.5	1.1	1.1	0.9	1.3	0.9	0.6	1.3	2.5
1.5	1.1	1.1	2.2	0.9	1.8	1.5	1.2	1.6	2.1	6.6	4.0	2.5
1.4	1.4	1.8	1.1	1.6	3.7	0.6	2.7	2.6	3.0	1.2	1.0	1.6
3.1	2.4	2.1	2.7	1.2	3.3	3.8	1.3	2.1	6.6	1.2	3.1	0.5
0.3	0.5	3.4	3.5	2.7	1.9	5.9	4.2	3.5	3.6	3.1	3.3	4.6

a. Construct a split stem-and-leaf plot in which each stem appears twice, once for leaves 0–4 and again for leaves 5–9.

b. Compare the split stem-and-leaf plot to the plot in Figure 2.13. Comment on the advantages and disadvantages of the split stem-and-leaf plot for these data.

20. Blood pressure: The following table presents systolic blood pressures for 50 adults.

98	122	114	123	142	174	130	136	109	139
103	119	114	119	117	92	124	107	115	167
116	120	128	122	117	113	120	118	109	113
156	130	119	128	114	127	104	118	111	121
94	118	110	125	121	101	102	110	99	106

a. Construct a stem-and-leaf plot, using the numbers 9 through 17 as the stems.

b. Repeat part (a), but split the stems, using two lines for each stem.

c. Which stem-and-leaf plot do you think is more appropriate for these data, the one in part (a) or the one in part (b)? Why?

21. Tennis and golf: Following are the ages of the winners of the men's Wimbledon tennis championship and the Masters golf championship for the years 1972 through 2019.

Ages of Wimbledon Winners

25	27	21	31	20	21	22	23	24	22	29	24
25	17	18	22	22	21	24	22	22	21	22	23
26	25	26	27	28	31	21	21	22	23	24	25
22	27	24	24	30	26	27	28	29	35	31	32

Ages of Masters Winners

32	36	38	35	33	27	42	27	23	31	28	26
32	27	46	28	30	31	32	33	32	35	28	43
38	23	41	33	37	25	26	32	31	29	33	31
28	39	39	26	33	32	35	21	28	37	27	43

a. Construct back-to-back split stem-and-leaf plots for these data sets.

b. How do the ages of Wimbledon champions differ from the ages of Masters champions?

22. Pass the popcorn: Following are the running times (in minutes) for the top 15 domestic grossing movies of all time rated PG-13 and the top 15 domestic grossing movies of all time rated R.

Movies Rated PG-13

Star Wars: The Force Awakens	136
Avengers: Endgame	182
Avatar	162
Black Panther	135
Avengers: Infinity War	160
Titanic	194
Jurassic World	124
Marvel's The Avengers	143
Star Wars: The Last Jedi	153
The Dark Knight	152
Rogue One: A Star Wars Story	133
Avengers: Age of Ultron	141
The Dark Knight Rises	164
Captain Marvel	125
The Hunger Games: Catching Fire	146

Source: Box Office Mojo

Movies Rated R

The Passion of the Christ	126
Deadpool	106
American Sniper	132
It	135
Deadpool 2	134
The Matrix Reloaded	138
The Hangover	96
The Hangover Part II	102
Beverly Hills Cop	105
The Exorcist	122
Logan	141
Ted	106
Saving Private Ryan	170
A Star is Born	134
300	117

Source: Box Office Mojo

a. Construct back-to-back stem-and-leaf plots for these data sets.

b. Do the running times of R-rated movies differ greatly from the running times of movies rated PG or PG-13, or are they roughly similar?

23. More weather: Construct a dotplot for the data in Exercise 18. Are there any gaps in the data?

24. Safety first: Following are the numbers of hospitals in each of the 50 U.S. states plus the District of Columbia that won Patient Safety Excellence Awards. Construct a dotplot for these data and describe its shape.

2	0	9	3	24	6	1	0	1	14	3
0	2	10	10	11	3	1	4	0	5	12
5	12	0	3	4	0	0	0	5	1	7
11	2	15	3	5	20	1	2	1	5	16
2	0	8	6	0	8	0				

25. Looking for a job: The following table presents the U.S. unemployment rate for each of the years 1996 through 2019.

Year	Unemployment	Year	Unemployment
1996	5.4	2008	5.8
1997	4.9	2009	9.3
1998	4.5	2010	9.6
1999	4.2	2011	8.9
2000	4.0	2012	8.1
2001	4.7	2013	7.4
2002	5.8	2014	6.1
2003	6.0	2015	5.4
2004	5.5	2016	4.9
2005	5.1	2017	4.4
2006	4.6	2018	3.9
2007	4.6	2019	3.8

Source: National Bureau of Labor Statistics

a. Construct a time-series plot of the unemployment rate.

b. For which periods of time was the unemployment rate increasing? For which periods was it decreasing?

26. Vacant apartments: The following table presents the percentage of U.S. residential rental units that were vacant during June and December of each year from 2011 through 2018.

Quarter	Vacancy Rate	Quarter	Vacancy Rate
Jun. 2011	9.2	Jun. 2015	6.8
Dec. 2011	9.4	Dec. 2015	7.0
Jun. 2012	8.6	Jun. 2016	6.7
Dec. 2012	8.7	Dec. 2016	6.9
Jun. 2013	8.2	Jun. 2017	7.3
Dec. 2013	8.2	Dec. 2017	6.9
Jun. 2014	7.5	Jun. 2018	6.8
Dec. 2014	7.0	Dec. 2018	6.8

Source: Current Population Survey

a. Construct a time-series plot for these data.
b. From 2011 through 2014, the proportion of Americans who owned a home declined. What was the trend in the vacancy rate during this time period?

27. Military spending: The following table presents the amount spent, in billions of dollars, on national defense by the U.S. government every other year for the years 1951 through 2019. The amounts are adjusted for inflation, and represent 2019 dollars.

Year	Spending	Year	Spending
1951	524.4	1987	617.2
1953	554.5	1989	593.0
1955	429.4	1991	583.8
1957	456.8	1993	494.7
1959	450.7	1995	448.0
1961	456.2	1997	427.2
1963	490.9	1999	434.6
1965	466.1	2001	461.4
1967	593.4	2003	614.0
1969	596.9	2005	660.3
1971	497.0	2007	748.8
1973	455.4	2009	790.6
1975	421.1	2011	756.8
1977	446.0	2013	634.3
1979	442.4	2015	647.5
1981	495.2	2017	677.6
1983	592.4	2019	686.1
1985	634.1		

Source: Department of Defense

a. Construct a time-series plot for these data.
b. The plot covers seven decades, from the 1950s through the 2010s. During which of these decades did national defense spending increase, and during which decades did it decrease?
c. The United States fought in the Korean War, which ended in 1953. What effect did the end of the war have on military spending after 1953?
d. During the period 1965–1968, the United States steadily increased the number of troops in Vietnam from 23,000 at the beginning of 1965 to 537,000 at the end of 1968. Beginning in 1969, the number of Americans in Vietnam was steadily reduced, with the last of them leaving in 1975. How is this reflected in the national defense spending from 1965 to 1975?

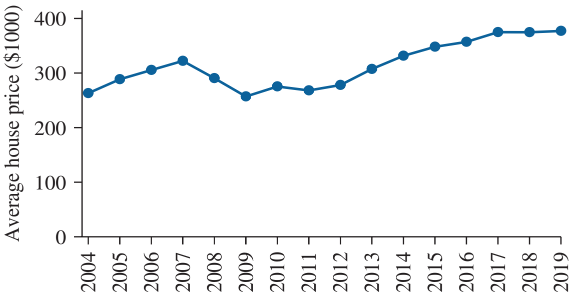
28. College students: The following table presents the numbers of male and female students (in thousands) enrolled in college in the United States as undergraduates for each of the years 2000 through 2019.

Year	Male	Female	Year	Male	Female
2000	5778	7377	2010	7836	10246
2001	6004	7711	2011	7823	10254
2002	6192	8065	2012	7715	10021
2003	6227	8253	2013	7660	9815
2004	6340	8441	2014	7586	9707
2005	6409	8555	2015	7502	9544
2006	6514	8671	2016	7417	9458
2007	6728	8876	2017	7347	9413
2008	7067	9299	2018	7372	9441
2009	7563	9901	2019	7399	9478

Source: National Center for Educational Statistics

a. Construct a time-series plot for the male enrollment; then on the same axes, construct a time-series plot for the female enrollment.
b. Which grew faster from 2000–2010, male enrollment or female enrollment?

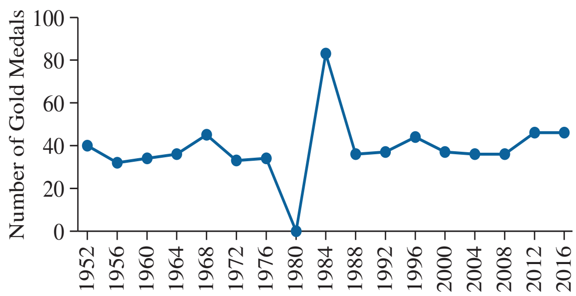
29. House prices: The following time-series plot presents the average price of houses sold in the United States during the first quarter of each of the years 2004–2019.



Source: Federal Reserve Bank of St. Louis

a. Estimate the average house price in 2005.
b. Was the average price in 2006 greater than, less than, or about the same as the average price in 2013?
c. True or false: The average price in 2019 exceeded the average price in 2004 by more than $100,000.
d. In 2008, an economic downturn known as the Great Recession occurred. What was the effect on the average house price?

30. Going for gold: The following time-series plot presents the number of Summer Olympic events in which the United States won a gold medal in each Olympic year from 1952 through 2016.



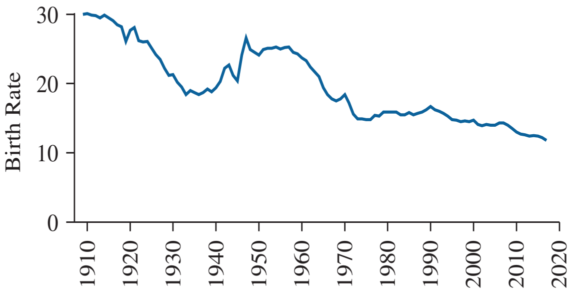
Source: Wikipedia

a. In one year, the United States did not participate in Summer games that were held in Moscow, in protest of the invasion of Afghanistan by the Soviet Union. Which year was this?

b. In 1984, the Soviet Union did not participate in the Summer games held in Los Angeles, citing "undisguised threats" against their athletes. Estimate the number of gold medals won by the United States in that year.

c. Other than 1980 and 1984, has the number of gold medals won by the United States been generally increasing, generally decreasing, or staying about the same?

31. Birth rates: The following time-series plot presents the birth rate (number of births per 1000 population) for the United States for the years 1909–2018.



Sources: Centers for Disease Control, Statista, Infoplease

a. Estimate the highest birth rate since 1909.

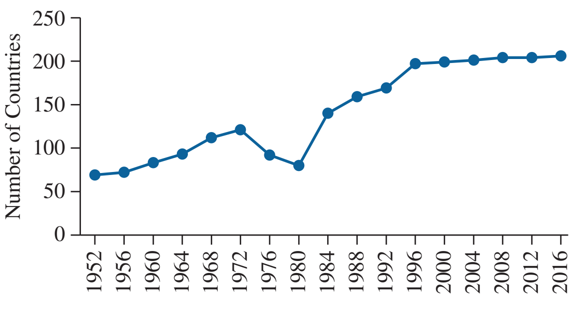
b. Was the first year that the birth rate was below 20 closer to 1930, 1950, or 1970?

c. The 1930s were the time of the Great Depression, during which many Americans were financially distressed. Was the birth rate during this period greater than, less than, or about the same as in the 1920s and 1940s?

d. In 1945 World War II ended, and many soldiers returned home to their families. What was the effect on the birth rate? (*Hint:* This period is known as the "baby boom.")

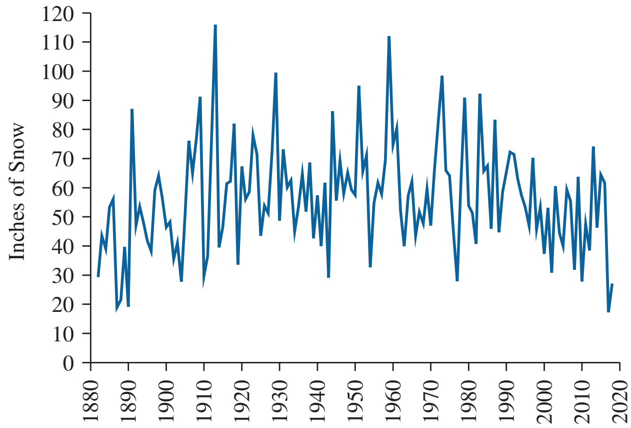
e. True or false: The birth rate in 2018 was lower than at any point in the previous 100 years.

32. More gold: The following time-series plot presents the number of countries participating in the Summer Olympic games in each Olympic year from 1952 through 2016.



Refer to Exercise 30. Someone says "Although the number of gold medals won by the United States didn't change much from 1952 to 1972, the performance of the United States steadily improved during that period." Which feature of the plot of the number of participating countries justifies that statement?

33. Let's go skiing: The following time-series plot presents the number of inches of snow falling in Denver each year from 1882 through 2018.



Source: National Weather Service

a. Estimate the greatest annual snowfall ever recorded in Denver.

b. Was the year of the greatest annual snowfall closest to 1900, 1910, or 1920?

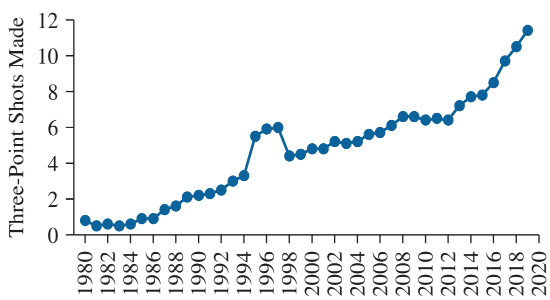
c. Was the amount of snowfall in the years 2000–2010 greater than, less than, or about equal to the snowfall in most other years?

d. True or false: The year with the least snowfall ever recorded in Denver was in the 2010s.

e. True or false: It usually snows more than 80 inches per year in Denver.

samot/ 123RF

34. Three-point shot: The following time-series plot presents the average number of three-point shots made in a National Basketball Association game for seasons ending in 1980 through 2019.



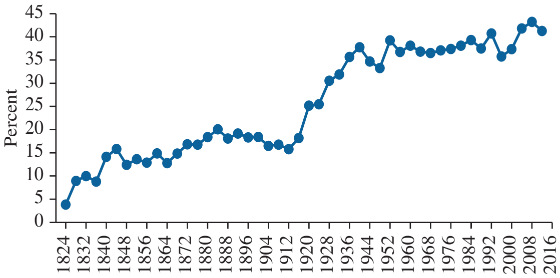
Source: Basketball-Reference.com

a. In 1997 the average number of three-point shots per game was greater than six for the first time. What was the next year that it was greater than six?

b. True or false: Since the year 2000, the average number of three-point shots made per game has increased every year.

c. In 1995 the distance from the three-point line to the basket was reduced from 23 feet 9 inches to 22 feet. In 1998 the distance was restored to 23 feet 9 inches. What was the effect of these rule changes?

35. Vote: The following time-series plot presents the percentage of the total U.S. population that voted in each presidential election from 1824 through 2016.



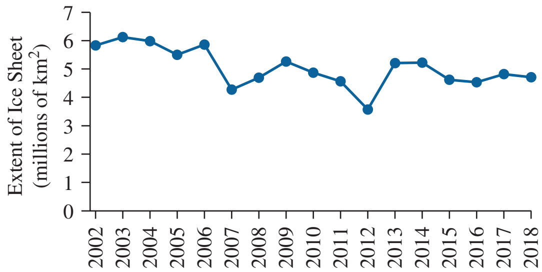
a. In 1824, the only people eligible to vote were white men who owned property. Approximately what percentage of the total population voted in 1824?

b. In 1828, voting privileges were extended to all white men, whether or not they owned property. What was the effect on the percentage of the population that voted?

c. During the 1890s, many Southern states passed laws that effectively prevented most African Americans and many poor whites from voting. What was the effect on the percentage of the population that voted?

d. In 1920, women obtained the right to vote. What was the effect on the percentage of the population that voted?

36. Arctic ice sheet: The following table presents the extent of ice coverage (in millions of square kilometers) in the Arctic region in September of each year from 2002 through 2018.



Source: National Snow & Ice Data Center

a. What was the first year that the coverage dropped below 5 million square kilometers?

b. What was the first year that the coverage dropped below 4 million square kilometers?

c. True or false: The coverage has been less than 5 million square kilometers in every year since 2011.

d. True or false: The coverage has decreased in every year since 2011.

Extending the Concepts

37. Elections: In U.S. presidential elections, each of the 50 states casts a number of electoral votes equal to its number of senators (2) plus its number of members of the House of Representatives. In addition, the District of Columbia casts three electoral votes. Following are the numbers of electoral votes cast for president for each of the 50 states and the District of Columbia in the election of 2020.

9	3	11	6	55	9	7	3	29	16	4
4	20	11	6	6	8	8	4	10	11	16
10	6	10	3	5	6	4	14	5	29	15
3	18	7	7	20	4	9	3	11	38	6
3	13	12	5	10	3	3				

a. Construct a split stem-and-leaf plot for these data, using two lines for each stem.

b. Construct a frequency histogram, with the classes chosen so that there are two classes for each stem.

c. Explain why the stem-and-leaf plot and the histogram have the same shape.

Answers to Check Your Understanding Exercises for Section 2.3

1. a.

12	
13	
14	
15	
16	
17	
18	
19	

b.

12	4494
13	63
14	9583
15	70156857
16	330601
17	12
18	2
19	552

c.

12	4449
13	36
14	3589
15	01556778
16	001336
17	12
18	2
19	255

2. a. 2018 **b.** 2006 **c.** False **d.** 1998–2018 **e.** 1994–1998

SECTION 2.4 Graphs Can Be Misleading

Objectives

1. Understand how improper positioning of the vertical scale can be misleading
2. Understand the area principle for constructing statistical graphs
3. Understand how three-dimensional graphs can be misleading

Statistical graphs, when properly used, are powerful forms of communication. Unfortunately, when graphs are improperly used, they can misrepresent the data and lead people to draw incorrect conclusions. We discuss here three of the most common forms of misrepresentation: incorrect position of the vertical scale, incorrect sizing of graphical images, and misleading perspective for three-dimensional graphs.

Objective 1 Understand how improper positioning of the vertical scale can be misleading

Positioning the Vertical Scale

Table 2.20 is a distribution of the number of passengers, in millions, at Denver International Airport in each year from 2012 through 2018.

Antonio Saba

Table 2.20 Passenger Traffic at Denver International Airport

Year	Number of Passengers (in millions)
2012	53.2
2013	52.6
2014	53.5
2015	54.0
2016	58.3
2017	61.4
2018	64.5

Source: Denver International Airport

In order to get a better picture of the data, we can make a bar graph. Figures 2.19 and 2.20 present two different bar graphs of the same data. Figure 2.19 presents a clear picture of the data. We can see that the number of passengers has been fairly steady, with just a slight increase from 2012 through 2018. Now imagine that someone was eager to persuade us that passenger traffic had increased greatly since 2012. If they were to show us Figure 2.19, we wouldn't be convinced. So they might show us a misleading picture like Figure 2.20 instead. Figure 2.20 gives the impression of a truly dramatic increase.

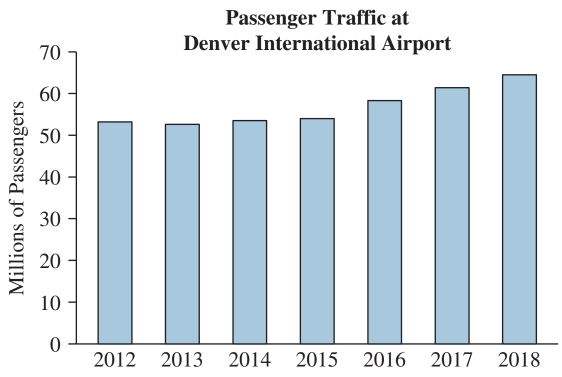


Figure 2.19 The bottom of the bars is at zero. This bar graph gives a correct impression of the data.

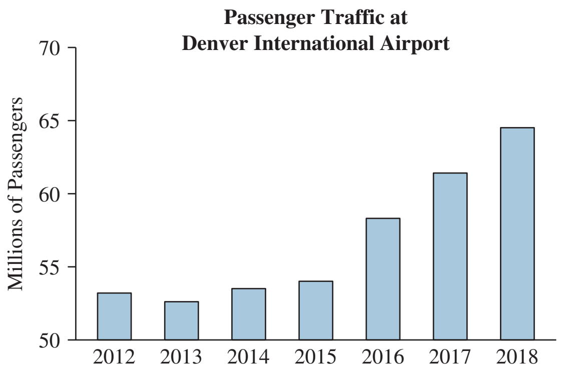


Figure 2.20 The bottom of the bars is not at zero. This bar graph exaggerates the differences between the bars.

Figures 2.19 and 2.20 are based on the same data. Why do they give such different impressions? The reason is that the baseline (the value corresponding to the bottom of the bars) is at zero in Figure 2.19, but not at zero in Figure 2.20. This exaggerates the differences between the bars. For example, in Figure 2.20, the bar for the year 2018 is more than three times as long as the bar for the year 2012, but the actual increase in passenger traffic is much less than that.

This sort of misleading information can be created with time-series plots as well. Figures 2.21 and 2.22 (page 80) present two different time-series plots of the data. In Figure 2.21, the baseline is at zero, so an accurate impression is given. In Figure 2.22, the baseline is larger than zero, so the rate of increase is exaggerated.

> When a graph or plot represents how much or how many of something, check the baseline. If it isn't at zero, the graph may be misleading.

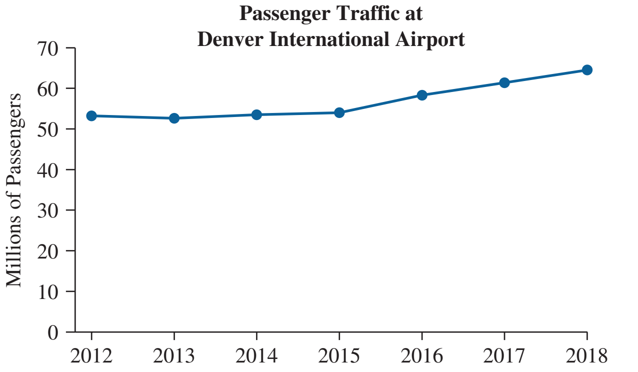


Figure 2.21 The baseline is at zero. This plot gives an accurate picture of the data.

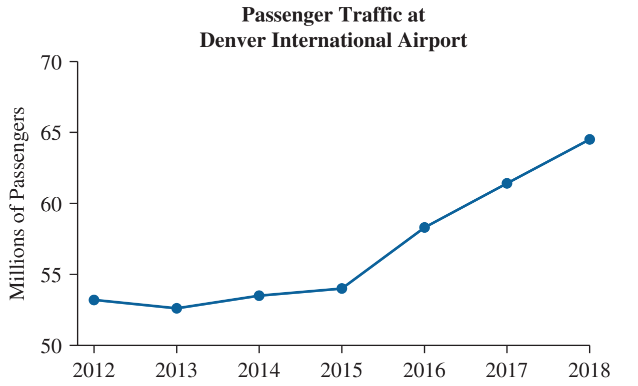


Figure 2.22 The baseline is not at zero. This plot exaggerates the rate of increase.

Objective 2 Understand the area principle for constructing statistical graphs

The Area Principle

We often use images to compare amounts. Larger images correspond to greater amounts. To use images properly in this way, we must follow a rule known as the **area principle**.

> **The Area Principle**
>
> When amounts are compared by constructing an image for each amount, the *areas* of the images must be proportional to the amounts. For example, if one amount is twice as much as another, its image should have twice as much area as the other image.
>
> When the area principle is violated, the images give a misleading impression of the data.

Bar graphs, when constructed properly, follow the area principle. The reason is that all the bars have the same width; only their height varies. Therefore, the areas of the bars are proportional to the amounts. For example, Figure 2.23 presents a bar graph that illustrates a comparison of the cost of jet fuel in 2016 and 2019. In 2016, the cost of jet fuel was \$0.93 per gallon, and in 2019 it had increased to \$1.78 per gallon.

The bars in the bar graph differ in only one dimension—their height. The widths are the same. For this reason, the bar graph presents an accurate comparison of the two prices. The price in 2016 is about half the price in 2019, and the area of the bar for 2016 is about half the area of the bar for 2019.

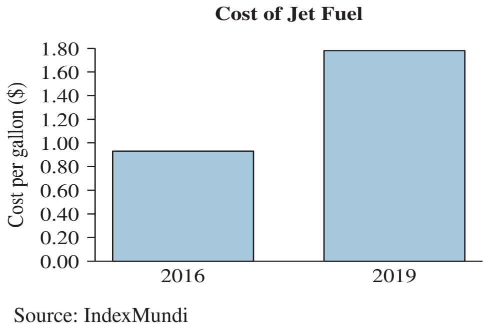


Figure 2.23 Price per gallon of jet fuel in 2016 and 2019. The bar graph accurately represents the difference.

Unfortunately, people often mistakenly vary both dimensions of an image when making a comparison. This exaggerates the difference. Following is a comparison of the cost of jet fuel in the years 2016 and 2019 that uses a picture of an airplane to illustrate the difference.

The pictures of the planes make the difference appear much larger than the correctly drawn bar graph does. The reason is that both the height and the width of the airplane have approximately doubled. Thus the area of the larger plane is about four times the area of the smaller plane. This graph violates the area principle and gives a misleading impression of the comparison.

Check Your Understanding

1. The population of country A is twice as large as the population of country B. True or false: If images are used to represent the populations, both the height and width of the image for country A should be twice as large as the height and width of the image for country B.
2. If the baseline of a bar graph or time-series plot is not at zero, then the differences may appear to be ________ than they actually are. (*Choices: larger, smaller*)

Answers are on page 85.

Objective 3 Understand how three-dimensional graphs can be misleading

Three-Dimensional Graphs and Perspective

The bar graph in Figure 2.23 presents an accurate picture of the prices of jet fuel in the years 2016 and 2019. Newspapers and magazines often prefer to present three-dimensional bar graphs, because they are visually more impressive. Unfortunately, in order to make the tops of the bars visible, these graphs are often drawn as though the reader is looking down on them. This can make the bars look shorter than they really are.

Figure 2.24 presents a three-dimensional bar graph of the sort often seen in publications. The data are the same as in Figure 2.23: The price in 2016 is \$0.93, and the price in 2019 is \$1.78. However, because you are looking down on the bars, they appear shorter than they really are.

Beware of three-dimensional bar graphs. If you can see the tops of the bars, they may look shorter than they really are.

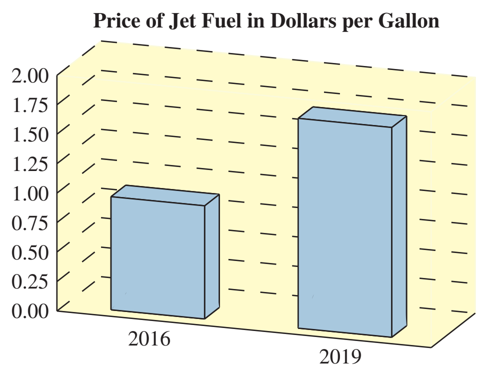


Figure 2.24 Price per gallon of jet fuel in 2016 and 2019. The bars appear shorter than they really are, because you are looking down at them.

SECTION 2.4 Exercises

Exercises 1 and 2 are the Check Your Understanding exercises located within the section.

Understanding the Concepts

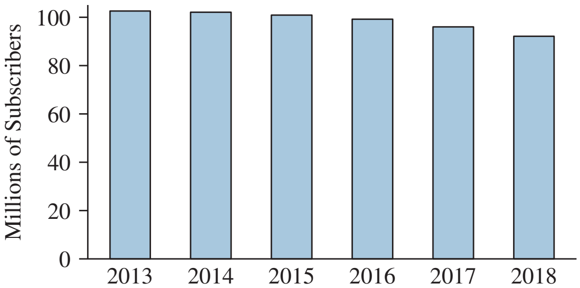
In Exercises 3 and 4, fill in each blank with the appropriate word or phrase.

3. A plot that represents how much of something there is may be misleading if the baseline is not at ________ .

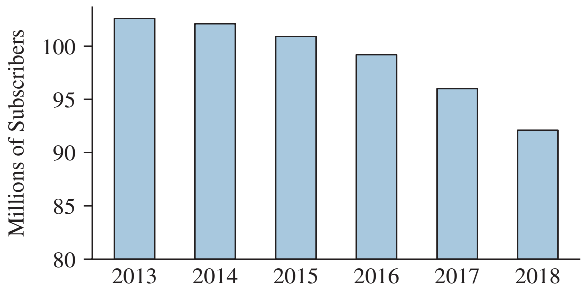
4. The area principle says that when images are used to compare amounts, the areas of the images should be ________ to the amounts.

Working with the Concepts

5. Cable subscriptions decline: The number of cable television subscribers has been declining in recent years. Following are two bar graphs that illustrate the decline. (Source: Business Insider)



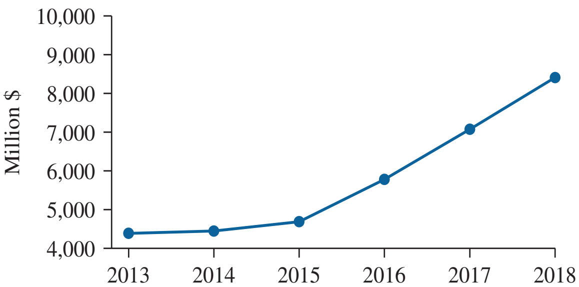
(A)



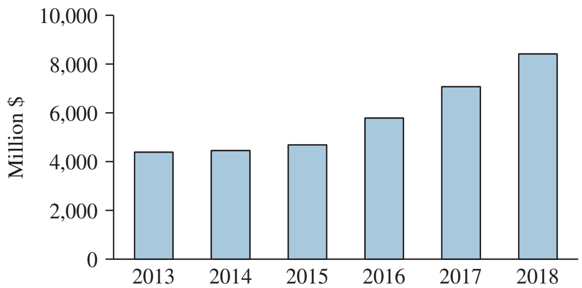
(B)

Choose one of the following options, and explain why it is correct:
(i) Graph A presents an accurate picture, and graph B exaggerates the decline.
(ii) Graph B presents an accurate picture, and graph A understates the decline.

6. Music sales: The following time-series plot and bar graph both present the sales of digital music for the years 2013–2018. Which of the graphs presents the more accurate picture? Why?



(A)

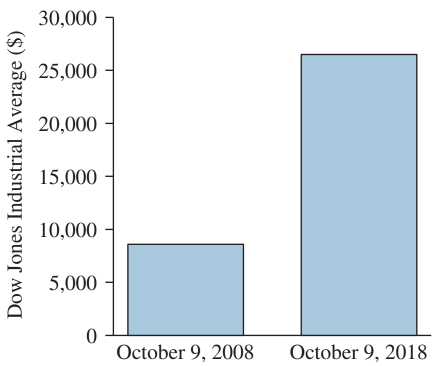


(B)

7. Stock market prices: The Dow Jones Industrial Average reached its lowest point in recent history on October 9, 2008, when it closed at $8,579. Ten years later, on October 9, 2018, the average had risen to $26,486.78. Which of the following graphs accurately represents the magnitude of the increase? Which one exaggerates it?

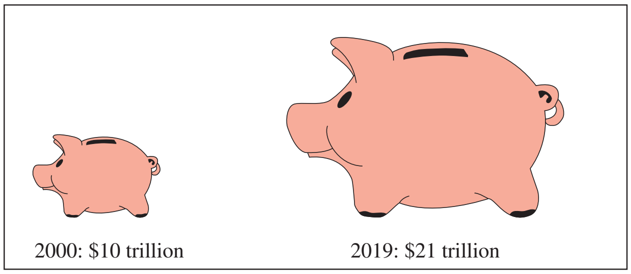
McGraw Hill Education/Ken Cavanagh

(A)

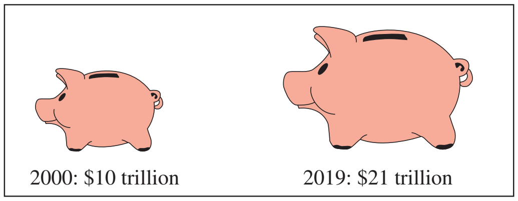


(B)

8. Gross domestic product: The gross domestic product (GDP) of the United States is the total value of all goods and services produced in the country. In 2000, the GDP was $10.0 trillion. In 2019, the GDP was $21.0 trillion, slightly more than twice as much. Which of the following graphs compares these totals more accurately, and why? (Source: St. Louis Federal Reserve)

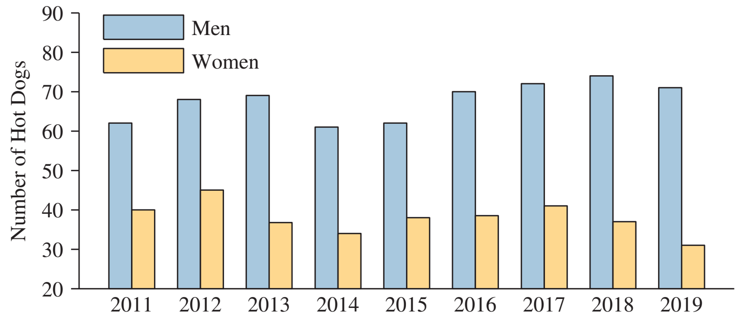


(A)

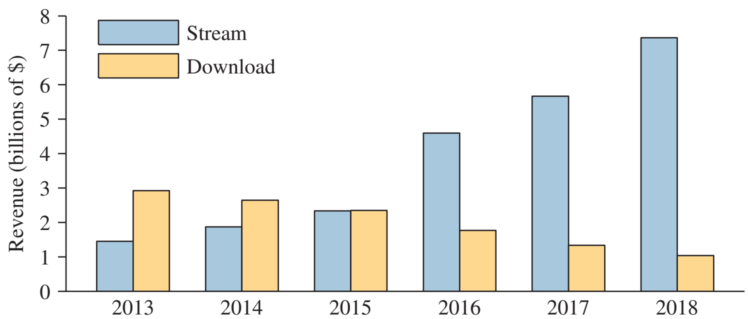


(B)

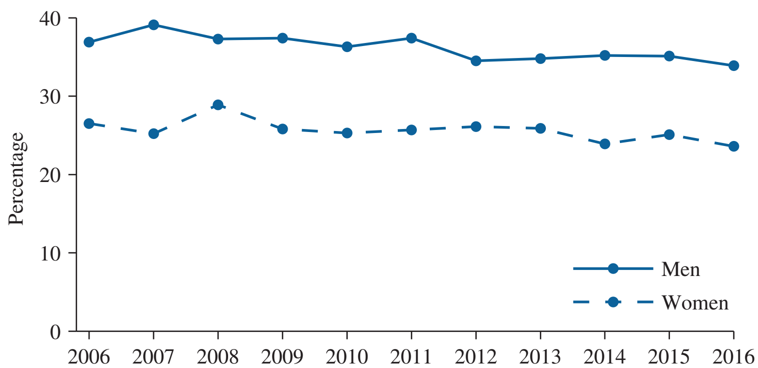
9. I'll take mine with mustard: The following bar graph presents the number of hot dogs eaten by the men's and women's winner of Nathans Famous Hot Dog eating championship for the years 2011–2019. Does the graph present an accurate picture of the difference between the men's and women's winners? Or is it misleading? Explain



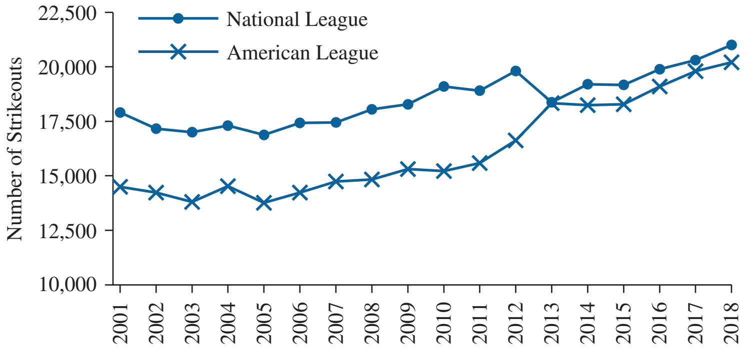
10. Stream or download? The following bar graph presents the revenue (in billions of $) for the music industry from music streaming (including subscriptions) and music downloading for the years 2013–2018. Does the graph present an accurate picture of the differences in revenue from these two sources? Or is it misleading? Explain



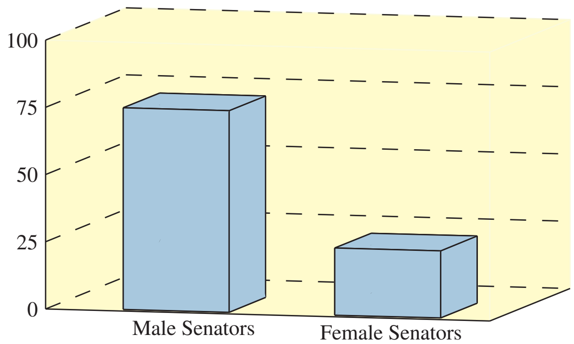
11. Heart disease: The following plot contains two time-series graphs presenting the percentages of men and women over the age of 65 who have been diagnosed with heart disease for the years 2006–2016. Does the plot present an accurate picture of the differences between the percentages for men and women? Or is it misleading? Explain?



12. Strike three: The plot contains two time-series graphs presenting the number of strikeouts in both the American and National League for the years 2001–2018. There have been more strikeouts in the National League in each of those years. Does the plot present an accurate picture of the differences in the numbers of strikeouts? Or is it misleading? Explain.



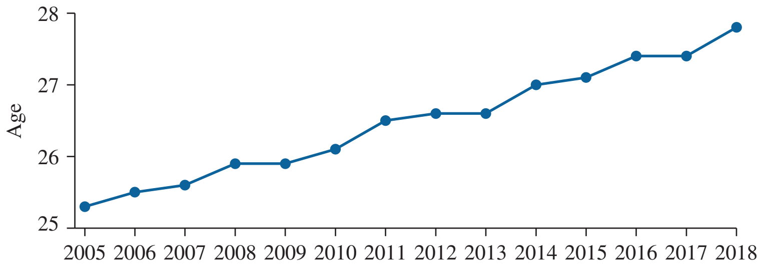
13. Female senators: Of the 100 members of the United States Senate recently, 75 were men and 25 were women. The following three-dimensional bar graph attempts to present this information.



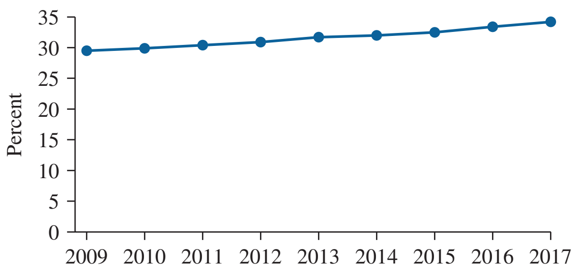
a. Explain how this graph is misleading.

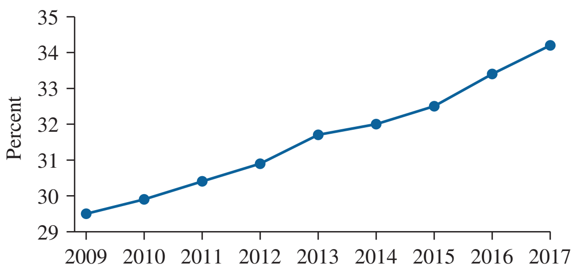
b. Construct a graph (not necessarily three-dimensional) that presents this information accurately.

14. Age at marriage: Data compiled by the U.S. Census Bureau suggests that the age at which women first marry has increased over time. The following time-series plot presents the average age at which women first marry for the years 2005–2018. Does the plot present an accurate picture of the increase, or is it misleading? Explain.



15. College degrees: Both of the following time-series plots present the percentage of U.S. adults who have earned college degrees for the years 2009–2017.



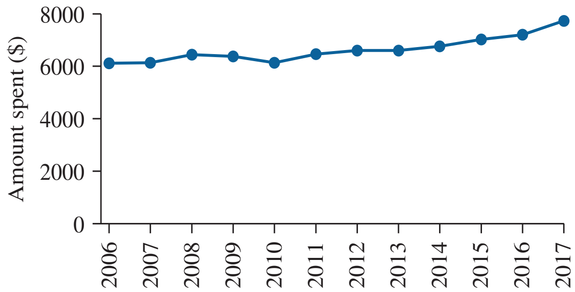


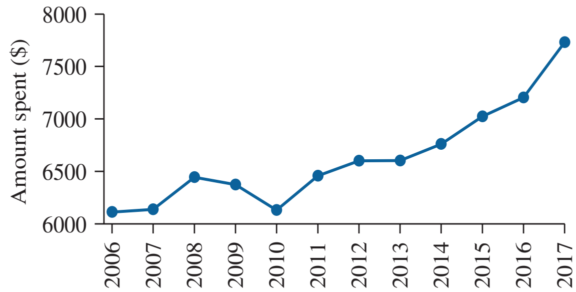
Which of the following statements is true and why?

i. The percentage of U.S. adults with college degrees increased slightly between 2009 and 2017.

ii. The percentage of U.S. adults with college degrees increased considerably between 2009 and 2017.

16. Food expenditures: Both of the following time-series plots present the average amount spent on food by U.S. residents for the years 2006 through 2017. (Source: U.S. Department of Agriculture)





Which of the following statements is true and why?

(i) The amount spent on food increased considerably between 2006 and 2017.

(ii) The amount spent on food increased slightly between 2006 and 2017.

Extending the Concepts

17. Manipulating the *y*-axis: For the data in Table 2.20:

a. Construct a bar graph in which the *y*-axis is labeled from 0 to 100.

b. Compare this bar graph with the bar graphs in Figures 2.19 and 2.20. Does this bar graph tend to make the difference seem smaller than the other bar graphs do?

c. Which of the three bar graphs do you think presents the most accurate picture of the data? Why?

Answers to Check Your Understanding Exercises for Section 2.4

1. False **2.** Larger

Chapter 2 Summary

Section 2.1: The first step in summarizing qualitative data is to construct a frequency distribution or relative frequency distribution. Then a bar graph or pie chart can be constructed. Bar graphs can illustrate either frequencies or relative frequencies. Side-by-side bar graphs can be used to compare two qualitative data sets that have the same categories.

Section 2.2: Frequency distributions and relative frequency distributions are also used to summarize quantitative data. Histograms are graphical summaries that illustrate frequency distributions and relative frequency distributions, allowing us to visualize the shape of a data set. Histograms can show us whether a data set is skewed, bell-shaped, or uniformly distributed, and whether it is unimodal or bimodal.

Section 2.3: Stem-and-leaf plots and dotplots are useful summaries for small data sets. They have an advantage over histograms: They allow every point in the data set to be seen. Back-to-back stem-and-leaf plots can be used to compare the shapes of two data sets. Time-series plots illustrate how the value of a variable has changed over time.

Section 2.4: To avoid constructing a misleading graph, be sure to start the vertical scale at zero. When images are used to compare amounts, the area principle should be followed. This principle states that the areas of the images should be proportional to the amounts. Three-dimensional bar graphs are often misleading, because the bars look shorter than they really are.

Vocabulary and Notation

area principle 80
back-to-back stem-and-leaf plot 69
bar graph 38
bell-shaped 57
bimodal 57
class 49
class width 51
dotplot 70
frequency 36
frequency distribution 36
frequency histogram 53
histogram 53
lower class limit 51
mode 57
negatively skewed 57
open-ended class 56
Pareto chart 39
pie chart 40
positively skewed 57
relative frequency 37
relative frequency distribution 37
relative frequency histogram 53
side-by-side bar graph 40
skewed 57

Chapter Quiz

1. Following is the list of letter grades for students in an algebra class: A, B, F, A, C, C, A, B, D, F, D, A, A, B, C, F, B, D, C, A, A, A, F, B, C, A, C. Construct a frequency distribution for these data.
2. Construct a relative frequency distribution for the data in Exercise 1.
3. Construct a frequency bar graph for the data in Exercise 1.
4. Construct a pie chart for the data in Exercise 1.
5. The first class in a relative frequency distribution is 2.0–4.9, and there are six classes. Find the remaining five classes. What is the class width?
6. True or false: A histogram can have more than one mode.
7. A sample of 100 students was asked how many hours per week they spent studying. The following frequency distribution shows the results. Construct a frequency histogram for these data.

Number of Hours	Frequency
1.0–4.9	14
5.0–8.9	34
9.0–12.9	29
13.0–16.9	15
17.0–20.9	8

8. Construct a relative frequency histogram for the data in Exercise 7.
9. List the data in the following stem-and-leaf plot. The leaf represents the ones digit.

1	1 1 5 5 9 9 9
2	2 2 3 5 7 8
3	0 0 8
4	4 5 7 8
5	0 1 3 3 5 6 8

10. Following are the prices (in dollars) for a sample of coffee makers.
19 22 29 68 35 37 28 22 41 39 28
Construct a stem-and-leaf plot for these data.
11. Following are the prices (in dollars) for a sample of espresso makers.
99 50 31 65 50 99 70 40 25 56 30 77
Construct a stem-and-leaf plot for these data.
12. Construct back-to-back stem-and-leaf plots for the data in Exercises 10 and 11.
13. Construct a dotplot for the data in Exercise 10.
14. The following table presents the percentage of Americans who use a phone exclusively, with no landline phone, for the years 2014–2017. Construct a time-series plot for these data.

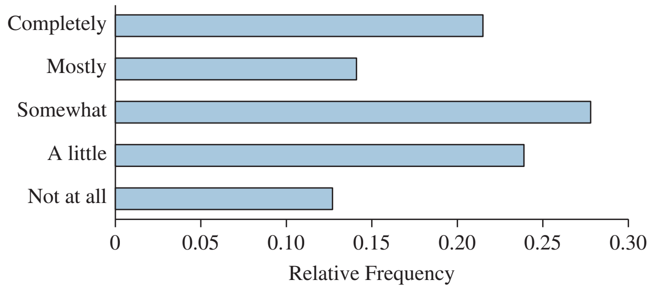
Time Period	Percent
January–June 2014	43.1
July–December 2014	44.1
January–June 2015	46.7
July–December 2015	47.7
January–June 2016	49.0
July–December 2016	50.5
January–June 2017	52.0
July–December 2017	53.3

Source: National Health Interview Survey

15. According to the area principle, if one amount is twice as much as another, its image should have ________ as much area as the other image.

Review Exercises

1. **Trust your doctor:** The General Social Survey recently surveyed people to ask, "How much would you trust your doctor to put your health above costs?" The following relative frequency bar graph presents the results.



Source: General Social Survey

   a. Which was the most frequently given answer?
   b. True or false: Less than one-fourth of the respondents said that they trusted their doctor completely.
   c. True or false: More than half of the respondents said that they trusted their doctor either a little or not at all.
   d. A total of 2719 people responded to this question. True or false: More than 500 of them said that they completely trusted or mostly trusted their doctor.

2. **Phones:** The following relative frequency distribution presents the U.S. market share, in percent, of various phone companies in a recent year.

Company	Market Share
Apple	47
Samsung	22
LG	12
Motorola	6
Others	13

Source: Counterpoint Research

   a. Construct a relative frequency bar graph.
   b. Construct a pie chart.
   c. True or false: There are more than twice as many Apple users than Samsung users.

3. **Poverty rates:** The following table presents the percentage of people who lived in poverty in the various regions of the United States in the years 2014 and 2017.

Region	Percent in 2014	Percent in 2017
Northeast	12.6	12.0
Midwest	13.0	12.7
South	16.5	14.8
West	15.2	12.9

Source: United States Census Bureau

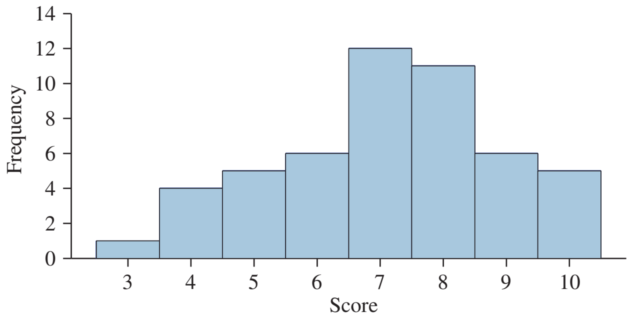
   a. Construct a side-by-side bar graph for these data.
   b. True or false: The poverty rate was lower in 2017 than in 2014 in each region.
   c. Which region had the greatest decrease?

4. **Do your homework:** The National Survey of Student Engagement asked a sample of college freshmen how often they came to class without completing their assignments. Following are the results:

Response	Percent
Never	35
Sometimes	48
Often	12
Very often	5

   a. Construct a relative frequency bar graph.
   b. Construct a pie chart.
   c. True or false: More than half of the students reported that they sometimes come to class without completing their assignments.

5. Quiz scores: The following frequency histogram presents the scores on a recent statistics quiz in a class of 50 students.



a. What is the most frequent score?
b. How many students scored less than 6?
c. What percentage of students scored 10?
d. Is the histogram more accurately described as unimodal or as bimodal?

6. House freshmen: Newly elected members of the U.S. House of Representatives are referred to as "freshmen." The following frequency distribution presents the number of freshmen elected in each election from 1912 to 2016.

Number of Freshmen	Frequency
20–39	2
40–59	15
60–79	10
80–99	14
100–119	7
120–139	3
140–159	1
160–179	1

Source: Library of Congress

a. How many classes are there?
b. What is the class width?
c. What are the class limits?
d. Construct a frequency histogram.
e. Construct a relative frequency distribution.
f. Construct a relative frequency histogram.

7. More freshmen: For the data in Exercise 6:
a. In what percentage of elections were 100 or more freshmen elected?
b. In what percentage of elections were fewer than 60 freshmen elected?

8. Royalty: Following are the ages at death for all English and British monarchs since 1066.

59	40	67	58	56	28	41	49	65	68
43	64	33	46	35	49	40	12	32	52
55	15	42	69	58	48	54	67	51	49
67	76	81	67	71	81	68	70	77	56

a. Construct a frequency distribution with approximately eight classes.
b. Construct a frequency histogram based on this frequency distribution.
c. Construct a relative frequency distribution with approximately eight classes.
d. Construct a relative frequency histogram based on this frequency distribution.

9. More royalty: Construct a stem-and-leaf plot for the data in Exercise 8.

10. Presidents: Following are the ages at deaths for all U.S. presidents.

67	83	90	73	85	68	78	80	53	65
71	79	56	77	64	74	66	49	63	57
70	67	58	71	60	57	67	72	60	63
46	90	78	88	64	81	93	93		

a. Construct a frequency distribution with a class width of 5 and a lower limit of 45 for the first class.
b. Construct a frequency histogram based on this frequency distribution.
c. Construct a relative frequency distribution with a class width of 5 and a lower limit of 45 for the first class.
d. Construct a relative frequency histogram based on this frequency distribution.

11. Royalty and presidents: For the data in Exercises 8 and 10:

a. Construct a back-to-back stem-and-leaf plot.

b. Construct a back-to-back stem-and-leaf plot with split stems.

c. Which plot do you think is more appropriate for these data?

12. Dotplot: Construct a dotplot for the data in Exercise 10.

13. Pandora vs. Spotify: Following are the numbers of subscribers (in millions) to the music streaming services Pandora and Spotify for the years 2013 through 2019:

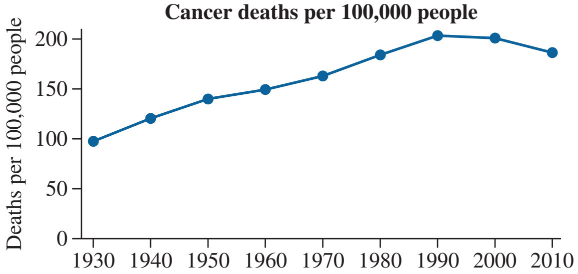
Year	Pandora	Spotify
2013	65.6	6.0
2014	76.2	10.0
2015	81.5	20.0
2016	81.1	30.0
2017	76.7	52.0
2018	72.3	75.0
2019	66.0	100.0

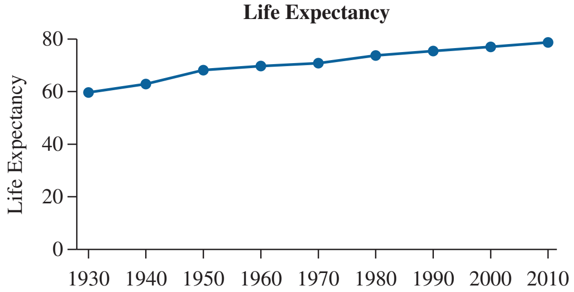
a. Construct a time-series plot for the number of Pandora subscribers.

b. Construct a time-series plot for the number of Spotify subscribers.

c. Describe the trends in the number of subscribers for both services.

14. Cancer rates: Cancer has long been the second most common cause of death (behind heart disease) in the United States. Cancer risk increases with age, so that older people are much more likely to be diagnosed with cancer than younger people. The following time-series plots present the number of deaths from cancer per 100,000 people, and the life expectancy, for the years 1930–2010.



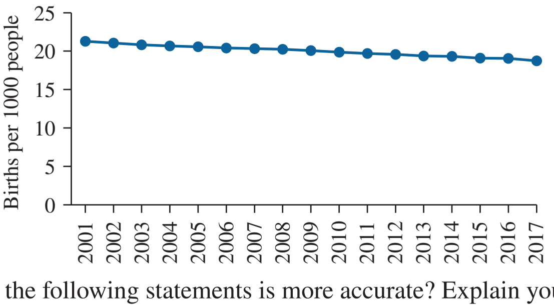


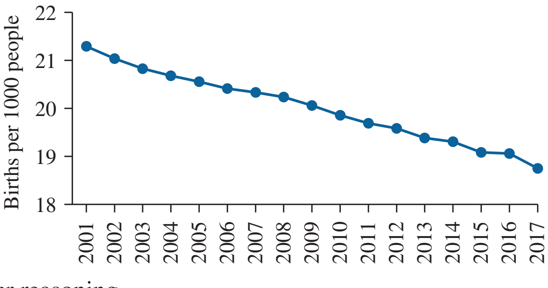
a. In which year was the death rate from cancer the highest? In which year was it the lowest?

b. Has life expectancy been increasing or decreasing during the years 1930–2010?

c. Treatments for cancer have been improving since 1930. Yet the death rate from cancer increased during the period 1930–1990. How can this be explained?

15. Falling birth rate: The following time-series plots both present estimates for the number of births per 1000 people worldwide for the years 2001–2017. (Source: The World Bank)





Which of the following statements is more accurate? Explain your reasoning.

(i) The birth rate decreased somewhat between 2001–2017.

(ii) The birth rate decreased dramatically between 2001–2017.

Write About It

1. Explain why the frequency bar graph and the relative frequency bar graph for a data set have a similar appearance.
2. In what ways do frequency distributions for qualitative data differ from those for quantitative data?
3. Provide an example of a data set whose histogram you would expect to be skewed to the right. Explain why you would expect the histogram to be skewed to the right.
4. Time-series data are discrete when observations are made at regularly spaced time intervals. The time-series data sets in this chapter are all discrete. Time-series data are continuous when there are observations at extremely closely spaced intervals that are connected to provide values at every instant of time. An example of continuous time-series data is an electrocardiogram. Provide some examples of time-series data that are discrete and some that are continuous.

In-Class Activities

1. **Shapes of histograms:** Each student tosses five dice and adds the numbers. Then each student tosses two dice and multiplies the numbers. Construct histograms for the results of each experiment. What are their shapes?
2. **Different graphs:** Collect data on variables for each student. Potential variables include height, number of siblings, number of pets, and number of classes currently taken. For each variable, construct a dotplot, a stem-and-leaf plot, and a histogram. Discuss the advantages and disadvantages of each graph.
3. **Misleading graphs:** Look through the internet, newspapers, or magazines for misleading graphs. Explain how they are misleading. Then find some that present accurate comparisons, and explain why you believe they are accurate.

Case Study: Do Hybrid Cars Get Better Gas Mileage?

In the chapter introduction, we presented gas mileage data for 2019 model year hybrid and small non-hybrid cars. We will use histograms and back-to-back stem-and-leaf plots to compare the mileages between these two groups of cars. The following tables present the mileages, in miles per gallon.

Mileage Ratings for 2019 Hybrid Cars							
40	28	21	43	23	56	22	28
28	48	31	41	19	26	39	48
26	46	19	30	55	23	29	41
33	21	42	50	22	42	50	29
26	29	34	46	52	42	30	41
43	58	25	25	44	44	22	19
27	49	21	23	52	41	24	52
24	23	46					

Source: www.fueleconomy.gov

Mileage Ratings for 2019 Small Non-hybrid Cars							
33	32	28	32	34	33	27	25
31	26	30	33	32	32	29	27
33	30	29	31	30	37	34	30
35	34	36	33	35	33	39	36
32	36	22	35	36	33	28	31
27	28	35	27	31	32	29	34
30	33	29	32	32	30	32	29
26	30	32	29	31	31	32	33

Source: www.fueleconomy.gov

1. Construct a frequency distribution for the hybrid cars with a class width of 2.
2. Explain why a class width of 2 is too narrow for these data.
3. Construct a relative frequency distribution for the hybrid cars with a class width of 3, where the first class has a lower limit of 18.
4. Construct a histogram based on this relative frequency distribution. Is the histogram unimodal or bimodal? Describe the skewness, if any, in these data.
5. Construct a frequency distribution for the non-hybrid cars with an appropriate class width.
6. Using this class width, construct a relative frequency distribution for the non-hybrid cars.
7. Construct a histogram based on this relative frequency distribution. Is the histogram unimodal or bimodal? Describe the skewness, if any, in these data.
8. Compare the histogram for the hybrid cars with the histogram for the non-hybrid cars. For which cars do the mileages vary more?
9. Construct a back-to-back stem-and-leaf plot for these data, using two lines for each stem. Which do you think illustrates the comparison better, the histograms or the back-to-back stem-and-leaf plot? Why?

Ryan McVay/Getty Images

chapter 4

Probability

Introduction

How likely is it that you will live to be 100 years old? The following table, called a *life table*, can be used to answer this question.

United States Life Table, Total Population

Age Interval	Proportion Surviving	Age Interval	Proportion Surviving
0–10	0.99123	50–60	0.94010
10–20	0.99613	60–70	0.86958
20–30	0.99050	70–80	0.70938
30–40	0.98703	80–90	0.42164
40–50	0.97150	90–100	0.12248

Source: Centers for Disease Control and Prevention

The column labeled "Proportion Surviving" presents the proportion of people alive at the beginning of an age interval who will still be alive at the end of the age interval. For example, among those currently age 20, the proportion who will still be alive at age 30 is 0.99050, or 99.050%. With an understanding of some basic concepts of probability, one can use the life table to compute the probability that a person of a given age will still be alive a given number of years from now. Life insurance companies use this information to determine how much to charge for life insurance policies. In the case study at the end of the chapter, we will use the life table to study some further questions that can be addressed with the methods of probability.

This chapter presents an introduction to probability. Probability is perhaps the only branch of knowledge that owes its existence to gambling. In the seventeenth century, owners of gambling houses hired some of the leading mathematicians of the time to calculate the chances that players would win certain gambling games. Later, people realized that many real-world problems involve chance as well, and since then the methods of probability have been used in almost every area of knowledge.

SECTION 4.1 Basic Concepts of Probability

Objectives

1. Construct sample spaces
2. Compute and interpret probabilities
3. Approximate probabilities by using the Empirical Method

At the beginning of a football game, a coin is tossed to decide which team will get the ball first. There are two reasons for using a coin toss in this situation. First, it is impossible to predict which team will win the coin toss, because there is no way to tell ahead of time whether the coin will land heads or tails. The second reason is that in the long run, over the course of many football games, we know that the home team will win about half of the tosses and the visiting team will win about half. In other words, although we don't know what the outcome of a single coin toss will be, we do know what the outcome of a long series of tosses will be—they will come out about half heads and half tails.

A coin toss is an example of a **probability experiment**. A probability experiment is one in which we do not know what any individual outcome will be, but we do know how a long series of repetitions will come out. Another familiar example of a probability experiment is the rolling of a die. A die has six faces; the faces have from one to six dots. We cannot predict which face will turn up on a single roll of a die, but, assuming the die is evenly balanced (not loaded), we know that in the long run, each face will turn up one-sixth of the time.

The *probability* of an event is the proportion of times that the event occurs in the long run. So, for a "fair" coin, that is, one that is equally likely to come up heads as tails, the probability of heads is $1/2$ and the probability of tails is $1/2$.

DEFINITION

The **probability** of an event is the proportion of times the event occurs in the long run, as a probability experiment is repeated over and over again.

The South African mathematician John Kerrich carried out a famous study that illustrates the idea of the long-run proportion. Kerrich was in Denmark when World War II broke out and spent the war interned in a prisoner-of-war camp. To pass the time, he carried out a series of probability experiments, including one in which he tossed a coin 10,000 times and recorded each toss as a head or a tail.

Figure 4.1 (page 161) summarizes a computer-generated re-creation of Kerrich's study, in which the proportion of heads is plotted against the number of tosses. For example, it turned out that after 5 tosses, 3 heads had appeared, so the proportion of heads was $3/5 = 0.6$. After 100 tosses, 49 heads had appeared, so the proportion of heads was $49/100 = 0.49$. After 10,000 tosses, the proportion of heads was 0.4994, which is very close to the true probability of 0.5. The figure shows that the proportion varies quite a bit within the first few tosses, but the proportion settles down very close to 0.5 as the number of tosses becomes larger.

The fact that the long-run proportion approaches the probability is called the *law of large numbers*.

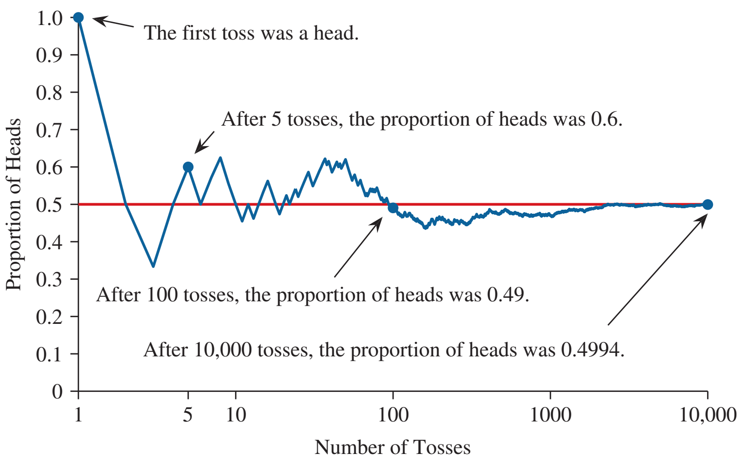


Figure 4.1 As the number of tosses increases, the proportion of heads fluctuates around the true probability of 0.5 and gets closer to 0.5. The horizontal axis is not drawn to scale.

Explain It Again

Law of large numbers: The law of large numbers is another way to state our definition of probability.

Law of Large Numbers

The **law of large numbers** says that as a probability experiment is repeated again and again, the proportion of times that a given event occurs will approach its probability.

Objective 1 Construct sample spaces

Probability Models

To study probability formally, we need some basic terminology. The collection of all the possible outcomes of a probability experiment is called a *sample space*.

DEFINITION

A **sample space** contains all the possible outcomes of a probability experiment.

EXAMPLE 4.1

Describe sample spaces

Describe a sample space for each of the following experiments.

a. The toss of a coin
b. The roll of a die
c. Selecting a student at random from a list of 10,000 students at a large university
d. Selecting a simple random sample of 100 students from a list of 10,000 students

Solution

a. There are two possible outcomes for the toss of a coin: Heads and Tails. So a sample space is {Heads, Tails}.

b. There are six possible outcomes for the roll of a die: the numbers from 1 to 6. So a sample space is {1, 2, 3, 4, 5, 6}.

c. Each of the 10,000 students is a possible outcome for this experiment, so the sample space consists of the 10,000 students.

d. This sample space consists of every group of 100 students that can be chosen from the population of 10,000—in other words, every possible simple random sample of size 100. This is a huge number of outcomes; it can be written approximately as a 6 followed by 241 zeros. This is larger than the number of atoms in the universe.

We are often concerned with occurrences that consist of several outcomes. For example, when rolling a die, we might be interested in the possibility of rolling an odd number. Rolling an odd number corresponds to the collection of outcomes {1, 3, 5} from the sample space {1, 2, 3, 4, 5, 6}. In general, a collection of outcomes of a sample space is called an *event*.

Notation Roundup

P(A) denotes the probability of the event *A*.

DEFINITION

An **event** is an outcome or a collection of outcomes from a sample space.

Once we have a sample space for an experiment, we need to specify the probability of each event. This is done with a *probability model*. We use the letter "*P*" to denote probabilities. So, for example, if we toss a coin, we denote the probability that the coin lands heads by "*P*(Heads)."

DEFINITION

A **probability model** for a probability experiment consists of a sample space, along with a probability for each event.

Notation: If A denotes an event, the probability of the event A is denoted $P(A)$.

Objective 2 Compute and interpret probabilities

Probability models with equally likely outcomes

In many situations, the outcomes in a sample space are equally likely. For example, when we toss a coin, we usually assume that the two outcomes "Heads" and "Tails" are equally likely. We call such a coin a *fair* coin. Similarly, a fair die is one in which the numbers from 1 to 6 are equally likely to turn up. When the outcomes in a sample space are equally likely, we can use a simple formula to determine the probability of events.

Explain It Again

Fair and unfair: A fair coin or die is one for which all outcomes are equally likely. An unfair coin or die is one for which some outcomes are more likely than others.

Computing Probabilities with Equally Likely Outcomes

If a sample space has n **equally likely outcomes**, and an event A has k outcomes, then

$$P(A) = \frac{\text{Number of outcomes in } A}{\text{Number of outcomes in the sample space}} = \frac{k}{n}$$

EXAMPLE 4.2

Compute the probability of an event

A fair die is rolled. Find the probability that an odd number comes up.

Solution

The sample space has six equally likely outcomes: {1, 2, 3, 4, 5, 6}. The event of an odd number has three outcomes: {1, 3, 5}. The probability is

$$P(\text{odd number}) = \frac{3}{6} = \frac{1}{2}$$

EXAMPLE 4.3

Compute the probability of an event

In the Georgia Cash-4 Lottery game, a winning number between 0000 and 9999 is chosen at random, with all the possible numbers being equally likely. What is the probability that all four digits of the winning number are the same?

Solution

The outcomes in the sample space are the numbers from 0000 to 9999, so there are 10,000 equally likely outcomes in the sample space. There are 10 outcomes for which all the digits are the same: 0000, 1111, and so on up to 9999. The probability is

$$P(\text{all four digits the same}) = \frac{10}{10{,}000} = 0.001$$

The law of large numbers states that the probability of an event is the long-run proportion of times that the event occurs. An event that never occurs, even in the long run, has a probability of 0. This is the smallest probability an event can have. An event that occurs every time has a probability of 1. This is the largest probability an event can have.

Explain It Again

Rules for the value of a probability: A probability can never be negative, and a probability can never be greater than 1.

SUMMARY

The probability of an event is always between 0 and 1. In other words, for any event A, $0 \leq P(A) \leq 1$.

If A cannot occur, then $P(A) = 0$.

If A is certain to occur, then $P(A) = 1$.

EXAMPLE 4.4

Computing probabilities

A penny, a nickel, and a dime are tossed. Denoting a head by H and a tail by T, we can denote these three tosses in order. For example, HTH means the penny landed heads, the nickel landed tails, and the dime landed heads. There are eight possible outcomes: HHH, HHT, HTH, HTT, THH, THT, TTH, and TTT. Assume these outcomes are equally likely.

a. What is the probability that there are exactly two heads?

b. What is the probability that all three tosses are the same?

Solution

a. Of the eight equally likely outcomes, the three outcomes HHT, HTH, and THH correspond to having two heads. Therefore

$$P(\text{Two heads}) = \frac{3}{8}$$

b. Of the eight equally likely outcomes, the two outcomes HHH and TTT correspond to having all tosses the same. Therefore

$$P(\text{All three tosses are the same}) = \frac{2}{8} = \frac{1}{4}$$

Check Your Understanding

1. In Example 4.4, what is the probability that the penny comes up heads?
2. In Example 4.4, what is the probability that the penny and the dime come up the same?

Answers are on page 169.

EXAMPLE 4.5

Constructing a sample space

Cystic fibrosis is a disease of the mucous glands whose most common sign is progressive damage to the respiratory system and digestive system. This disease is inherited, as follows. A certain gene may be of type *A* or type *a*. Every person has two copies of the gene—one inherited from the person's mother, one from the person's father. If both copies are *a*, the person will have cystic fibrosis. Assume that a mother and father both have genotype *Aa*, that is, one gene of each type. Assume that each copy is equally likely to be transmitted to their child. What is the probability that the child will have cystic fibrosis?

Solution

Most of the work in solving this problem is in constructing the sample space. We'll do this in two ways. First, the tree diagram in Figure 4.2 shows that there are four possible outcomes. In the tree diagram, the first two branches indicate the two possible outcomes, *A* and *a*, for the mother's gene. Then for each of these outcomes there are two branches indicating the possible outcomes for the father's gene. An alternate method is to construct a table like Table 4.1.

Table 4.1

Mother's Gene	Father's Gene	Child's Genotype
A	*A*	*AA*
A	*a*	*Aa*
a	*A*	*aA*
a	*a*	*aa*

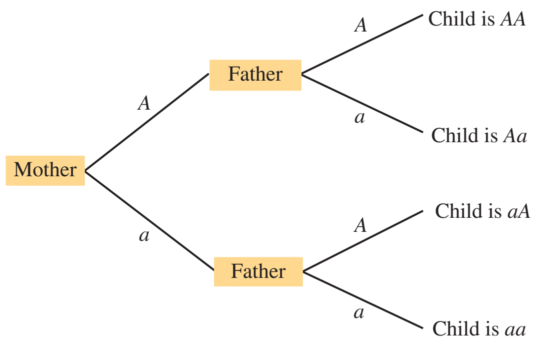


Figure 4.2 Tree diagram illustrating the four outcomes for the child's genotype

We can use either the table or the tree to list the outcomes. Listing the mother's gene first, the four outcomes are *AA*, *Aa*, *aA*, and *aa*. For one of the four outcomes, *aa*, the child will have cystic fibrosis. Therefore, the probability of cystic fibrosis is $1/4$.

Check Your Understanding

3. A penny and a nickel are tossed. Each is a fair coin, which means that heads and tails are equally likely.

a. Construct a sample space containing equally likely outcomes. Each outcome should specify the results for both coins.

b. Find the probability that one coin comes up heads and the other comes up tails.

Answers are on page 169.

Sampling from a population is a probability experiment

In Section 1.1, we learned that statisticians collect data by drawing samples from populations. Sampling an individual from a population is a probability experiment. The population is the sample space, and the members of the population are equally likely outcomes. For this reason, the ideas of probability are fundamental to statistics.

EXAMPLE 4.6

Computing probabilities involving sampling

There are 10,000 families in a certain town. They are categorized by their type of housing as follows.

Own a house	4753
Own a condo	1478
Rent a house	912
Rent an apartment	2857

A pollster samples a single family at random from this population.

a. What is the probability that the sampled family owns a house?

b. What is the probability that the sampled family rents?

Solution

a. The sample space consists of the 10,000 households. Of these, 4753 own a house, so the probability that the sampled family owns a house is

$$P(\text{Owns a house}) = \frac{4753}{10{,}000} = 0.4753$$

b. The number of families who rent is $912 + 2857 = 3769$. Therefore, the probability that the sampled family rents is

$$P(\text{Rents}) = \frac{3769}{10{,}000} = 0.3769$$

In practice, of course, the pollster would sample many people, not just one. In fact, statisticians use the basic ideas of probability to draw conclusions about populations by studying samples drawn from them. In later chapters of this book, we will see how this is done.

Unusual events

As the name implies, an unusual event is one that is not likely to happen—in other words, an event whose probability is small. There are no hard-and-fast rules as to just how small a probability needs to be before an event is considered unusual, but 0.05 is commonly used.

Explain It Again

Unusual events: The cutoff value for the probability of an unusual event can be any small value that seems appropriate for a specific situation. The most commonly used value is 0.05.

SUMMARY

An **unusual event** is one whose probability is small.

Sometimes people use the cutoff 0.05; that is, they consider any event whose probability is less than 0.05 to be unusual. But there are no hard-and-fast rules about this.

EXAMPLE 4.7

Determine whether an event is unusual

In a college of 5000 students, 150 are math majors. A student is selected at random and turns out to be a math major. Is this an unusual event?

Solution

The sample space consists of 5000 students, each of whom is equally likely to be chosen. The event of choosing a math major consists of 150 students. Therefore,

$$P(\text{Math major is chosen}) = \frac{150}{5000} = 0.03$$

Since the probability is less than 0.05, then by the most commonly applied rule, this would be considered an unusual event.

Objective 3 Approximate probabilities by using the Empirical Method

Approximating Probabilities with the Empirical Method

The law of large numbers says that if we repeat a probability experiment a large number of times, then the proportion of times that a particular outcome occurs is likely to be close to the true probability of the outcome. The **Empirical Method** consists of repeating an experiment a large number of times and using the proportion of times an outcome occurs to approximate the probability of the outcome.

EXAMPLE 4.8

Approximate the probability that a newborn baby is low birth weight

Explain It Again

The Empirical Method is only approximate: The Empirical Method does not give us the exact probability. But the larger the number of replications of the experiment, the more reliable the approximation will be.

The Centers for Disease Control reports that in a recent year there were 313,752 births in the United States to low birth weight babies (less than 2500 grams), while 3,477,960 were to babies with weight greater than 2500 grams. Approximate the probability that a newborn baby is low birth weight.

Solution

We compute the number of times the experiment has been repeated:

3,477,960	not low birth weight
+ 313,752	low birth weight
= 3,791,712	births

The proportion of births that are low birth weight is

$$\frac{313{,}752}{3{,}791{,}712} = 0.0827$$

We approximate $P(\text{low birth weight}) \approx 0.0827$.

Example 4.8 is based on a very large number (3,791,712) of replications. The law of large numbers says that the proportion of outcomes approaches the true probability as the number of replications becomes large. For a number this large, we can be virtually certain that the proportion 0.0827 is very close to the true probability. Of course, changes in nutrition and other environmental factors can affect the probability in the future.

Check Your Understanding

4. There are 100,000 voters in a city. A pollster takes a simple random sample of 1000 of them and finds that 513 support a bond issue to support the public library and 487 oppose it. Estimate the probability that a randomly chosen voter in this city supports the bond issue.

Answer is on page 169.

SECTION 4.1 Exercises

Exercises 1–4 are the Check Your Understanding exercises located within the section.

Understanding the Concepts

In Exercises 5–8, fill in each blank with the appropriate word or phrase.

5. If an event cannot occur, its probability is ________.

6. If an event is certain to occur, its probability is ________.

7. The collection of all possible outcomes of a probability experiment is called a ________.

8. An outcome or collection of outcomes from a sample space is called an ________.

In Exercises 9–12, determine whether the statement is true or false. If the statement is false, rewrite it as a true statement.

9. The law of large numbers states that as a probability experiment is repeated, the proportion of times that a given outcome occurs will approach its probability.

10. If A denotes an event, then the sample space is denoted by $P(A)$.

11. The Empirical Method can be used to calculate the exact probability of an event.

12. For any event A, $0 \le P(A) \le 1$.

Practicing the Skills

In Exercises 13–18, assume that a fair die is rolled. The sample space is {1, 2, 3, 4, 5, 6}, and all the outcomes are equally likely.

13. Find $P(2)$.

14. Find P(Even number).

15. Find P(Less than 3).

16. Find P(Greater than 2).

17. Find $P(7)$.

18. Find P(Less than 10).

19. A fair coin has probability 0.5 of coming up heads.
- **a.** If you toss a fair coin twice, are you certain to get one head and one tail?
- **b.** If you toss a fair coin 100 times, are you certain to get 50 heads and 50 tails?
- **c.** As you toss the coin more and more times, will the proportion of heads approach 0.5?

20. Roulette wheels in Nevada have 38 pockets. They are numbered 0, 00, and 1 through 36. On each spin of the wheel, a ball lands in a pocket, and each pocket is equally likely.
- **a.** If you spin a roulette wheel 38 times, is it certain that each number will come up once?
- **b.** If you spin a roulette wheel 3800 times, is it certain that each number will come up 100 times?
- **c.** As the wheel is spun more and more times, will the proportion of times that each number comes up approach 1/38?

In Exercises 21–24, assume that a coin is tossed twice. The coin may not be fair. The sample space consists of the outcomes {HH, HT, TH, TT}.

21. Is the following a probability model for this experiment? Why or why not?

Outcome	HH	HT	TH	TT
Probability	0.55	0.42	0.31	0.25

22. Is the following a probability model for this experiment? Why or why not?

Outcome	HH	HT	TH	TT
Probability	0.36	0.24	0.24	0.16

23. Is the following a probability model for this experiment? Why or why not?

Outcome	HH	HT	TH	TT
Probability	0.09	0.21	0.21	0.49

24. Is the following a probability model for this experiment? Why or why not?

Outcome	HH	HT	TH	TT
Probability	0.33	0.46	−0.18	0.4

Working with the Concepts

25. How probable is it? Someone computes the probabilities of several events. The probabilities are listed on the left, and some verbal descriptions are listed on the right. Match each probability with the best verbal description. Some descriptions may be used more than once.

Probability		Verbal Description
(a)	0.50	i. This event is certain to happen.
(b)	0.00	ii. This event is as likely to happen as not.
(c)	0.90	
(d)	1.00	iii. This event may happen, but it isn't likely.
(e)	0.10	
(f)	−0.25	iv. This event is very likely to happen, but it isn't certain.
(g)	0.01	
(h)	2.00	v. It would be unusual for this event to happen.
		vi. This event cannot happen.
		vii. Someone made a mistake.

26. Do you know SpongeBob? According to a survey by Nickelodeon TV, 88% of children under 13 in Germany recognized a picture of the cartoon character SpongeBob SquarePants. What is the probability that a randomly chosen German child recognizes SpongeBob?

27. Who will you vote for? In a survey of 500 likely voters in a certain city, 275 said that they planned to vote to reelect the incumbent mayor.
- **a.** What is the probability that a surveyed voter plans to vote to reelect the mayor?
- **b.** Interpret this probability by estimating the percentage of all voters in the city who plan to vote to reelect the mayor.

28. Job satisfaction: In a poll conducted by the General Social Survey, 497 out of 1769 people said that their main satisfaction in life comes from their work.
- **a.** What is the probability that a person who was polled finds his or her main satisfaction in life from work?
- **b.** Interpret this probability by estimating the percentage of all people whose main satisfaction in life comes from their work.

29. True–false exam: A section of an exam contains four true–false questions. A completed exam paper is selected at random, and the four answers are recorded.
- **a.** List all 16 outcomes in the sample space.
- **b.** Assuming the outcomes to be equally likely, find the probability that all the answers are the same.
- **c.** Assuming the outcomes to be equally likely, find the probability that exactly one of the four answers is "True."
- **d.** Assuming the outcomes to be equally likely, find the probability that two of the answers are "True" and two of the answers are "False."

30. A coin flip: A fair coin is tossed three times. The outcomes of the three tosses are recorded.
- **a.** List all eight outcomes in the sample space.
- **b.** Assuming the outcomes to be equally likely, find the probability that all three tosses are "Heads."
- **c.** Assuming the outcomes to be equally likely, find the probability that the tosses are all the same.
- **d.** Assuming the outcomes to be equally likely, find the probability that exactly one of the three tosses is "Heads."

31. Empirical Method: A coin is tossed 400 times and comes up heads 180 times. Use the Empirical Method to approximate the probability that the coin comes up heads.

32. Empirical Method: A die is rolled 600 times. On 85 of those rolls, the die comes up 6. Use the Empirical Method to approximate the probability that the die comes up 6.

33. Pitching: During a recent season, pitcher Clayton Kershaw threw 2515 pitches. Of these, 1303 were fastballs, 12 were changeups, 385 were curveballs, and 815 were sliders.

a. What is the probability that Clayton Kershaw throws a fastball?

b. What is the probability that Clayton Kershaw throws a breaking ball (curveball or slider)?

34. More pitching: During a recent season, pitcher Jon Lester threw 3727 pitches. Of these, 1449 were thrown with no strikes on the batter, 1168 were thrown with one strike, and 1110 were thrown with two strikes.

a. What is the probability that a Jon Lester pitch is thrown with no strikes?

b. What is the probability that a Jon Lester pitch is thrown with fewer than two strikes?

35. Risky drivers: An automobile insurance company divides customers into three categories: good risks, medium risks, and poor risks. Assume that of a total of 11,217 customers, 7792 are good risks, 2478 are medium risks, and 947 are poor risks. As part of an audit, one customer is chosen at random.

a. What is the probability that the customer is a good risk?

b. What is the probability that the customer is not a poor risk?

Gino Santa Maria/Shutterstock

36. Pay your bills: A company audit showed that of 875 bills that were sent out, 623 were paid on time, 155 were paid up to 30 days late, 78 were paid between 31 and 90 days late, and 19 were paid after 90 days. One bill is selected at random.

a. What is the probability that the bill was paid on time?

b. What is the probability that the bill was paid late?

37. Roulette: A Nevada roulette wheel has 38 pockets. Eighteen of them are red, eighteen are black, and two are green. Each time the wheel is spun, a ball lands in one of the pockets, and each pocket is equally likely.

a. What is the probability that the ball lands in a red pocket?

b. If you bet on red on every spin of the wheel, you will lose more than half the time in the long run. Explain why this is so.

38. More roulette: Refer to Exercise 37.

a. What is the probability that the ball lands in a green pocket?

b. If you bet on green on every spin of the wheel, you will lose more than 90% of the time in the long run. Explain why this is so.

39. Get an education: The General Social Survey asked 32,201 people how much confidence they had in educational institutions. The results were as follows.

Response	Number
A great deal	10,040
Some	17,890
Hardly any	4,271
Total	32,201

a. What is the probability that a sampled person has either some or a great deal of confidence in educational institutions?

b. Assume this is a simple random sample from a population. Use the Empirical Method to estimate the probability that a person has a great deal of confidence in educational institutions.

c. If we use a cutoff of 0.05, is it unusual for someone to have hardly any confidence in educational institutions?

40. How many kids? The General Social Survey asked 46,349 women how many children they had. The results were as follows.

Number of Children	Number of Women
0	12,656
1	7,438
2	11,290
3	7,143
4	3,797
5	1,811
6	916
7	522
8 or more	776
Total	46,349

a. What is the probability that a sampled woman has two children?

b. What is the probability that a sampled woman has fewer than three children?

c. Assume this is a simple random sample of U.S. women. Use the Empirical Method to estimate the probability that a U.S. woman has more than five children.

d. Using a cutoff of 0.05, is it unusual for a woman to have no children?

41. Hospital visits: According to the Agency for Healthcare Research and Quality, there were 409,706 hospital visits for asthma-related illnesses in a recent year. The age distribution was as follows.

Age Range	Number
Less than 1 year	7,866
1–17	103,040
18–44	79,659
45–64	121,728
65–84	80,649
85 and up	16,764
Total	409,706

a. What is the probability that an asthma patient is between 18 and 44 years old?

b. What is the probability that an asthma patient is 65 or older?

c. Using a cutoff of 0.05, is it unusual for an asthma patient to be less than 1 year old?

42. Don't smoke: The Centers for Disease Control and Prevention reported that there were 443,000 smoking-related deaths in the United States in a recent year. The numbers of deaths caused by various illnesses attributed to smoking are as follows:

Illness	Number
Lung cancer	128,900
Ischemic heart disease	126,000
Chronic obstructive pulmonary disease	92,900
Other	95,200
Total	443,000

a. What is the probability that a smoking-related death was the result of lung cancer?

b. What is the probability that a smoking-related death was the result of either ischemic heart disease or other?

Extending the Concepts

Two dice are rolled. One is red and one is blue. Each will come up with a number between 1 and 6. There are 36 equally likely outcomes for this experiment. They are ordered pairs of the form (Red die, Blue die).

43. Find a sample space: Construct a sample space for this experiment that contains the 36 equally likely outcomes.

44. Find the probability: What is the probability that the sum of the dice is 5?

45. Find the probability: What is the probability that the sum of the dice is 7?

46. The red die has been rolled: Now assume that you have rolled the red die, and it has come up 3. How many of the original 36 outcomes are now possible?

47. Find a new sample space: Construct a sample space containing the outcomes that are still possible after the red die has come up 3.

48. New information changes the probability: Given that the red die came up 3, what is the probability that the sum of the dice is 5? Is the probability the same as it was before the red die was observed?

49. New information doesn't change the probability: Given that the red die came up 3, what is the probability that the sum of the dice is 7? Is the probability the same as it was before the red die was observed?

Answers to Check Your Understanding Exercises for Section 4.1

1. 0.5

2. 0.5

3. a.

Penny	Nickel
H	H
H	T
T	H
T	T

b. 0.5

4. 0.513

SECTION 4.2 The Addition Rule and the Rule of Complements

Objectives

1. Compute probabilities by using the General Addition Rule
2. Compute probabilities by using the Addition Rule for Mutually Exclusive Events
3. Compute probabilities by using the Rule of Complements

If you go out in the evening, you might go to dinner, or to a movie, or to both dinner and a movie. In probability terminology, "go to dinner and a movie" and "go to dinner or a movie" are referred to as *compound events*, because they are composed of combinations of other events—in this case the events "go to dinner" and "go to a movie."

DEFINITION

A **compound event** is an event that is formed by combining two or more events.

In this section, we will focus on compound events of the form "A or B." We will say that the event "A or B" occurs whenever A occurs, or B occurs, or both A and B occur. We will learn how to compute probabilities of the form $P(A \text{ or } B)$.

DEFINITION

$P(A \text{ or } B) = P(A \text{ occurs or } B \text{ occurs or both occur})$

Table 4.2 presents the results of a survey in which 1000 adults were asked whether they favored a law that would provide more government support for higher education. In addition, each person was asked whether he or she voted in the last election. Those who had voted were classified as "Likely to vote," and those who had not were classified as "Not likely to vote."

Table 4.2

	Favor	Oppose	Undecided
Likely to vote	372	262	87
Not likely to vote	151	103	25

Table 4.2 is called a **contingency table**. It categorizes people with regard to two variables: whether they are likely to vote, and their opinion on the law. There are six categories, and the numbers in the table present the frequencies for each category. For example, we can see that 372 people are in the row corresponding to "Likely to vote" and the column corresponding to "Favor." Thus, 372 people were likely to vote and favored the law. Similarly, 103 people were not likely to vote and opposed the law.

EXAMPLE 4.9 Compute probabilities by using equally likely outcomes

Use Table 4.2 to answer the following questions:

a. What is the probability that a randomly selected adult is likely to vote and favors the law?

b. What is the probability that a randomly selected adult is likely to vote?

c. What is the probability that a randomly selected adult favors the law?

Solution

We think of the adults in the survey as outcomes in a sample space. Each adult is equally likely to be the one chosen. We begin by counting the total number of outcomes in the sample space:

$$372 + 262 + 87 + 151 + 103 + 25 = 1000$$

To answer part (a), we observe that there are 372 people who are likely to vote and favor the law. There are 1000 people in the survey. Therefore,

$$P(\text{Likely to vote and Favor}) = \frac{372}{1000} = 0.372$$

To answer part (b), we count the total number of outcomes corresponding to adults who are likely to vote:

$$372 + 262 + 87 = 721$$

There are 1000 people in the survey, and 721 of them are likely to vote. Therefore,

$$P(\text{Likely to vote}) = \frac{721}{1000} = 0.721$$

To answer part (c), we count the total number of outcomes corresponding to adults who favor the law:

$$372 + 151 = 523$$

There are 1000 people in the survey, and 523 of them favor the law. Therefore,

$$P(\text{Favor}) = \frac{523}{1000} = 0.523$$

Objective 1 Compute probabilities by using the General Addition Rule

The General Addition Rule

EXAMPLE 4.10

Compute a probability of the form $P(A \text{ or } B)$

Use the data in Table 4.2 to find the probability that a person is likely to vote or favors the law.

Solution

We will illustrate two approaches to this problem. In the first approach, we will use equally likely outcomes, and in the second, we will develop a method that is especially designed for probabilities of the form $P(A \text{ or } B)$.

Approach 1: To use equally likely outcomes, we reproduce Table 4.2 and circle the numbers that correspond to people who are either likely voters or who favor the law.

	Favor	Oppose	Undecided
Likely to vote	(372)	(262)	(87)
Not likely to vote	(151)	103	25

There are 1000 people altogether. The number of people who either are likely voters or favor the law is

$$372 + 262 + 87 + 151 = 872$$

Therefore,

$$P(\text{Likely to vote or Favor}) = \frac{372 + 262 + 87 + 151}{1000} = \frac{872}{1000} = 0.872$$

Approach 2: In this approach we will begin by computing the probabilities $P(\text{Likely to vote})$ and $P(\text{Favor})$ separately. We reproduce Table 4.2; this time we circle the numbers that correspond to likely voters and put rectangles around the numbers that correspond to favoring the law. Note that the number 372 has both a circle and a rectangle around it, because these 372 people are both likely to vote and favor the law.

	Favor	Oppose	Undecided
Likely to vote	[(372)]	(262)	(87)
Not likely to vote	[151]	103	25

There are $372 + 262 + 87 = 721$ likely voters and $372 + 151 = 523$ voters who favor the law. If we try to find the number of people who are likely to vote or who favor the law by adding these two numbers, we get $721 + 523 = 1244$, which is too large (there are only 1000 people in total). This happened because there are 372 people who are both likely voters and who favor the law, and these people are counted twice. We can still solve the problem by adding 721 and 523, but we must then subtract 372 to correct for the double counting.

We illustrate this reasoning, using probabilities.

$$P(\text{Likely to vote}) = \frac{721}{1000} = 0.721$$

$$P(\text{Favor}) = \frac{523}{1000} = 0.523$$

$$P(\text{Likely to vote AND Favor}) = \frac{372}{1000} = 0.372$$

$$
\begin{aligned}
P(\text{Likely to vote OR Favor}) &= P(\text{Likely to vote}) + P(\text{Favor}) \\
&\quad - P(\text{Likely to vote AND Favor}) \\
&= \frac{721}{1000} + \frac{523}{1000} - \frac{372}{1000} \\
&= \frac{872}{1000} = 0.872
\end{aligned}
$$

The method of subtracting in order to adjust for double counting is known as the General Addition Rule.

Explain It Again

The General Addition Rule: Use the General Addition Rule to compute probabilities of the form $P(A \text{ or } B)$.

The General Addition Rule

For any two events A and B,

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

EXAMPLE 4.11

Compute a probability by using the General Addition Rule

Refer to Table 4.2. Use the General Addition Rule to find the probability that a randomly selected person is not likely to vote or is undecided.

Solution

Using the General Addition Rule, we compute

$$
\begin{aligned}
&P(\text{Not likely to vote or Undecided}) \\
&\quad = P(\text{Not likely to vote}) + P(\text{Undecided}) - P(\text{Not likely to vote and Undecided})
\end{aligned}
$$

There are $151 + 103 + 25 = 279$ people not likely to vote out of a total of 1000. Therefore,

$$P(\text{Not likely to vote}) = \frac{279}{1000} = 0.279$$

There are $87 + 25 = 112$ people who are undecided out of a total of 1000. Therefore,

$$P(\text{Undecided}) = \frac{112}{1000} = 0.112$$

Finally, there are 25 people who are both not likely to vote and undecided. Therefore,

$$P(\text{Not likely to vote and Undecided}) = \frac{25}{1000} = 0.025$$

Using the General Addition Rule,

$$P(\text{Not likely to vote or Undecided}) = 0.279 + 0.112 - 0.025 = 0.366$$

Check Your Understanding

1. The following table presents numbers of U.S. workers, in thousands, categorized by type of occupation and educational level.

Type of Occupation	Non-College Graduate	College Graduate
Managers and professionals	17,564	31,103
Service	15,967	2,385
Sales and office	22,352	7,352
Construction and maintenance	12,511	1,033
Production and transportation	14,597	1,308

Source: Bureau of Labor Statistics

a. What is the probability that a randomly selected worker is a college graduate?
b. What is the probability that the occupation of a randomly selected worker is categorized either as Sales and office or as Production and transportation?
c. What is the probability that a randomly selected worker is either a college graduate or has a service occupation?

Answers are on page 178.

Objective 2 Compute probabilities by using the Addition Rule for Mutually Exclusive Events

Mutually Exclusive Events

Sometimes it is impossible for two events both to occur. For example, when a coin is tossed, it is impossible to get both a head and a tail. Two events that cannot both occur are called mutually exclusive. The term *mutually exclusive* means that when one event occurs, it excludes the other.

DEFINITION

Two events are said to be **mutually exclusive** if it is impossible for both events to occur.

Explain It Again

Meaning of mutually exclusive events: Two events are mutually exclusive if the occurrence of one makes it impossible for the other to occur.

We can use **Venn diagrams** to illustrate mutually exclusive events. In a Venn diagram, the sample space is represented by a rectangle, and events are represented by circles drawn inside the rectangle. If two circles do not overlap, the two events cannot both occur. If two circles overlap, the overlap area represents the occurrence of both events. Figures 4.3 and 4.4 illustrate the idea.

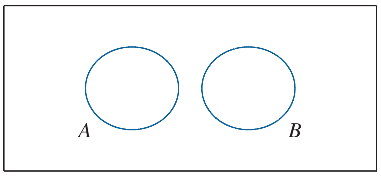


Figure 4.3 Venn diagram illustrating mutually exclusive events

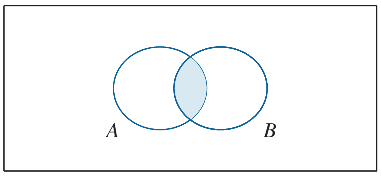


Figure 4.4 Venn diagram illustrating events that are not mutually exclusive

EXAMPLE 4.12

Determine whether two events are mutually exclusive

In each of the following, determine whether events A and B are mutually exclusive:

a. A die is rolled. Event A is that the die comes up 3, and event B is that the die comes up an even number.
b. A fair coin is tossed twice. Event A is that one of the tosses is a head, and event B is that one of the tosses is a tail.

Solution

a. These events are mutually exclusive. The die cannot both come up 3 and come up an even number.
b. These events are not mutually exclusive. If the two tosses result in HT or TH, then both events occur.

Check Your Understanding

2. A college student is chosen at random. Event A is that the student is older than 21 years, and event B is that the student is taking a statistics class. Are events A and B mutually exclusive?

3. A college student is chosen at random. Event A is that the student is an only child, and event B is that the student has a brother. Are events A and B mutually exclusive?

Answers are on page 178.

If events A and B are mutually exclusive, then $P(A \text{ and } B) = 0$. This leads to a simplification of the General Addition Rule.

The Addition Rule for Mutually Exclusive Events

If A and B are mutually exclusive events, then

$$P(A \text{ or } B) = P(A) + P(B)$$

In general, three or more events are mutually exclusive if only one of them can happen. If $A, B, C, \ldots$ are mutually exclusive, then

$$P(A \text{ or } B \text{ or } C \text{ or } \ldots) = P(A) + P(B) + P(C) + \cdots$$

EXAMPLE 4.13

Compute a probability by using the Addition Rule for Mutually Exclusive Events

In a recent Olympic Games, a total of 11,544 athletes participated. Of these, 554 represented the United States, 314 represented Canada, and 125 represented Mexico.

a. What is the probability that an Olympic athlete chosen at random represents the United States or Canada?

b. What is the probability that an Olympic athlete chosen at random represents the United States, Canada, or Mexico?

Solution

a. These events are mutually exclusive, because it is impossible to compete for both the United States and Canada. We compute P(U.S.) and P(Canada).

$$\begin{aligned} P(\text{U.S. or Canada}) &= P(\text{U.S.}) + P(\text{Canada}) \\ &= \frac{554}{11{,}544} + \frac{314}{11{,}544} \\ &= \frac{868}{11{,}544} \\ &= 0.075191 \end{aligned}$$

b. These events are mutually exclusive, because it is impossible to compete for more than one country. Therefore

$$\begin{aligned} P(\text{U.S. or Canada or Mexico}) &= P(\text{U.S.}) + P(\text{Canada}) + P(\text{Mexico}) \\ &= \frac{554}{11{,}544} + \frac{314}{11{,}544} + \frac{125}{11{,}544} \\ &= \frac{993}{11{,}544} \\ &= 0.086019 \end{aligned}$$

Check Your Understanding

4. In a statistics class of 45 students, 11 got a final grade of A, 22 got a final grade of B, and 8 got a final grade of C.

a. What is the probability that a randomly chosen student got an A or a B?
b. What is the probability that a randomly chosen student got an A, a B, or a C?

Answers are on page 178.

Objective 3 Compute probabilities by using the Rule of Complements

Complements

If there is a 60% chance of rain today, then there is a 40% chance that it will not rain. The events "Rain" and "No rain" are *complements*. The complement of an event A is the event that A does not occur.

DEFINITION

If A is any event, the **complement** of A is the event that A does not occur. The complement of A is denoted A^c.

Some complements are straightforward. For example, the complement of "the plane was on time" is "the plane was not on time." In other cases, finding the complement requires some thought. Example 4.14 illustrates this.

EXAMPLE 4.14

Find the complement of an event

Two hundred students were enrolled in a statistics class. Find the complements of the following events.

a. More than 50 of them are business majors.
b. At least 50 of them are business majors.
c. Fewer than 50 of them are business majors.
d. Exactly 50 of them are business majors.

Notation Roundup

A^c is the event that A does not occur.

Solution

a. If it is not true that more than 50 are business majors, then the number of business majors must be 50 or less than 50. The complement is that 50 or fewer of the students are business majors.

b. If it is not true that at least 50 are business majors, then the number of business majors must be less than 50. The complement is that fewer than 50 of the students are business majors.

c. If it is not true that fewer than 50 are business majors, then the number of business majors must be 50 or more than 50. Another way of saying this is that at least 50 of the students are business majors. The complement is that at least 50 of the students are business majors.

d. If it is not true that exactly 50 are business majors, then the number of business majors must not equal 50. The complement is that the number of business majors is not equal to 50.

Explain It Again

The complement occurs when the event doesn't occur: If an event does not occur, then its complement occurs. If an event occurs, then its complement does not occur.

Two important facts about complements are:

1. Either A or A^c must occur. For example, it must either rain or not rain.
2. A and A^c are mutually exclusive; they cannot both occur. For example, it is impossible for it to both rain and not rain.

In probability notation, fact 1 says that $P(A \text{ or } A^c) = 1$, and fact 2 along with the Addition Rule for Mutually Exclusive Events says that $P(A \text{ or } A^c) = P(A) + P(A^c)$. Putting them together, we get

$$P(A) + P(A^c) = 1$$

Subtracting $P(A)$ from both sides yields

$$P(A^c) = 1 - P(A)$$

This is the Rule of Complements.

The Rule of Complements

$$P(A^c) = 1 - P(A)$$

EXAMPLE 4.15 **Compute a probability by using the Rule of Complements**

According to *The Wall Street Journal*, 40% of cars sold in a recent year were small cars. What is the probability that a randomly chosen car sold in that year is not a small car?

Solution

$$P(\text{Not a small car}) = 1 - P(\text{Small car}) = 1 - 0.40 = 0.60$$

SECTION 4.2 Exercises

Exercises 1–4 are the Check Your Understanding exercises located within the section.

Understanding the Concepts

In Exercises 5–8, fill in each blank with the appropriate word or phrase.

5. The General Addition Rule states that $P(A \text{ or } B) = P(A) + P(B) -$ __________.
6. If events A and B are mutually exclusive, then $P(A \text{ and } B) =$ __________.
7. Given an event A, the event that A does not occur is called the __________ of A.
8. The Rule of Complements states that $P(A^c) =$ __________.

In Exercises 9–12, determine whether the statement is true or false. If the statement is false, rewrite it as a true statement.

9. The General Addition Rule is used for probabilities of the form $P(A \text{ or } B)$.
10. A compound event is formed by combining two or more events.
11. Two events are mutually exclusive if both events can occur.
12. If an event occurs, then its complement also occurs.

Practicing the Skills

13. If $P(A) = 0.75$, $P(B) = 0.4$, and $P(A \text{ and } B) = 0.25$, find $P(A \text{ or } B)$.
14. If $P(A) = 0.45$, $P(B) = 0.7$, and $P(A \text{ and } B) = 0.65$, find $P(A \text{ or } B)$.
15. If $P(A) = 0.2$, $P(B) = 0.5$, and A and B are mutually exclusive, find $P(A \text{ or } B)$.
16. If $P(A) = 0.7$, $P(B) = 0.1$, and A and B are mutually exclusive, find $P(A \text{ or } B)$.
17. If $P(A) = 0.3$, $P(B) = 0.4$, and $P(A \text{ or } B) = 0.7$, are A and B mutually exclusive?
18. If $P(A) = 0.5$, $P(B) = 0.4$, and $P(A \text{ or } B) = 0.8$, are A and B mutually exclusive?
19. If $P(A) = 0.35$, find $P(A^c)$.
20. If $P(B) = 0.6$, find $P(B^c)$.
21. If $P(A^c) = 0.27$, find $P(A)$.
22. If $P(B^c) = 0.64$, find $P(B)$.
23. If $P(A) = 0$, find $P(A^c)$.
24. If $P(A) = P(A^c)$, find $P(A)$.

In Exercises 25–30, determine whether events A and B are mutually exclusive.

25. A: Sophie is a member of the debate team; B: Sophie is the president of the theater club.
26. A: Jayden has a math class on Tuesdays at 2:00; B: Jayden has an English class on Tuesdays at 2:00.
27. A sample of 20 cars is selected from the inventory of a dealership. A: At least 3 of the cars in the sample are red; B: Fewer than 2 of the cars in the sample are red.
28. A sample of 75 books is selected from a library. A: At least 10 of the authors are female; B: At least 10 of the books are fiction.
29. A red die and a blue die are rolled. A: The red die comes up 2; B: The blue die comes up 3.
30. A red die and a blue die are rolled. A: The red die comes up 1; B: The total is 9.

In Exercises 31 and 32, find the complements of the events.

31. A sample of 225 internet users was selected.
    a. More than 200 of them use Google as their primary search engine.
    b. At least 200 of them use Google as their primary search engine.
    c. Fewer than 200 of them use Google as their primary search engine.
    d. Exactly 200 of them use Google as their primary search engine.
32. A sample of 700 phone batteries was selected.
    a. Exactly 24 of the batteries were defective.
    b. At least 24 of the batteries were defective.

c. More than 24 of the batteries were defective.
d. Fewer than 24 of the batteries were defective.

Working with the Concepts

33. Traffic lights: A commuter passes through two traffic lights on the way to work. Each light is either red, yellow, or green. An experiment consists of observing the colors of the two lights.
a. List the nine outcomes in the sample space.
b. Let A be the event that both colors are the same. List the outcomes in A.
c. Let B be the event that the two colors are different. List the outcomes in B.
d. Let C be the event that at least one of the lights is green. List the outcomes in C.
e. Are events A and B mutually exclusive? Explain.
f. Are events A and C mutually exclusive? Explain.

34. Dice: Two fair dice are rolled. The first die is red and the second is blue. An experiment consists of observing the numbers that come up on the dice.
a. There are 36 outcomes in the sample space. They are ordered pairs of the form (Red die, Blue die). List the 36 outcomes.
b. Let A be the event that the same number comes up on both dice. List the outcomes in A.
c. Let B be the event that the red die comes up 6. List the outcomes in B.
d. Let C be the event that one die comes up 6 and the other comes up 1. List the outcomes in C.
e. Are events A and B mutually exclusive? Explain.
f. Are events A and C mutually exclusive? Explain.

35. Car repairs: Let E be the event that a new car requires engine work under warranty, and let T be the event that the car requires transmission work under warranty. Suppose that $P(E) = 0.10$, $P(T) = 0.02$, and $P(E \text{ and } T) = 0.01$.
a. Find the probability that the car needs work on either the engine, the transmission, or both.
b. Find the probability that the car needs no work on the engine.

36. Sick computers: Let V be the event that a computer contains a virus, and let W be the event that a computer contains a worm. Suppose $P(V) = 0.15$, $P(W) = 0.05$, and $P(V \text{ and } W) = 0.03$.
a. Find the probability that the computer contains either a virus or a worm or both.
b. Find the probability that the computer does not contain a virus.

37. Computer purchases: Out of 800 large purchases made at a computer retailer, 336 were tablets, 398 were laptop computers, and 66 were desktop computers. As part of an audit, one purchase record is sampled at random.
a. What is the probability that it is a tablet?
b. What is the probability that it is not a desktop computer?

38. Visit your local library: On a recent Saturday, a total of 1200 people visited a local library. Of these people, 248 were under age 10, 472 were aged 10–18, 175 were aged 19–30, and the rest were more than 30 years old. One person is sampled at random.
a. What is the probability that the person is less than 19 years old?
b. What is the probability that the person is more than 30 years old?

Shutterstock/GaudiLab

39. How are your grades? In a recent semester at a local university, 500 students enrolled in both Statistics I and Psychology I. Of these students, 82 got an A in statistics, 73 got an A in psychology, and 42 got an A in both statistics and psychology.
a. Find the probability that a randomly chosen student got an A in statistics or psychology or both.
b. Find the probability that a randomly chosen student did not get an A in psychology.

40. Statistics grades: In a statistics class of 30 students, there were 13 men and 17 women. Two of the men and three of the women received an A in the course. A student is chosen at random from the class.
a. Find the probability that the student is a woman.
b. Find the probability that the student received an A.
c. Find the probability that the student is a woman or received an A.
d. Find the probability that the student did not receive an A.

41. Weight and cholesterol: The National Health Examination Survey reported that in a sample of 13,535 adults, 6443 had high cholesterol (total cholesterol above 200 mg/dL), 8841 were overweight (body mass index above 25), and 4629 were both overweight and had high cholesterol. A person is chosen at random from this study.
a. Find the probability that the person is overweight.
b. Find the probability that the person has high cholesterol.
c. Find the probability that the person does not have high cholesterol.
d. Find the probability that the person is overweight or has high cholesterol.

42. Paving stones: Two hundred paving stones were examined for cracks, and 15 were found to be cracked. The same 200 stones were examined for discoloration, and 27 were found to be discolored. A total of 4 stones were both cracked and discolored. One of the 200 stones is selected at random.
a. Find the probability that it is cracked.
b. Find the probability that it is discolored.
c. Find the probability that it is not cracked.
d. Find the probability that it is cracked or discolored.

43. Sick children: There are 25 students in Mrs. Bush's sixth-grade class. On a cold winter day in February, many of the students had runny noses and sore throats. After examining each student, the school nurse constructed the following table.

	Sore Throat	No Sore Throat
Runny Nose	6	12
No Runny Nose	4	3

a. Find the probability that a randomly selected student has a runny nose.
b. Find the probability that a randomly selected student has a sore throat.
c. Find the probability that a randomly selected student has a runny nose or a sore throat.
d. Find the probability that a randomly selected student has neither a runny nose nor a sore throat.

44. Flawed parts: On a certain day, a foundry manufactured 500 cast aluminum parts. Some of these had major flaws, some had minor flaws, and some had both major and minor flaws. The following table presents the results.

	Minor Flaw	No Minor Flaw
Major Flaw	20	35
No Major Flaw	75	370

a. Find the probability that a randomly chosen part has a major flaw.
b. Find the probability that a randomly chosen part has a minor flaw.
c. Find the probability that a randomly chosen part has a flaw (major or minor).
d. Find the probability that a randomly chosen part has no major flaw.
e. Find the probability that a randomly chosen part has no flaw.

45. Senators: The following table displays the 100 senators of the 116th U.S. Congress on January 3, 2019, classified by political party affiliation and gender.

	Male	Female	Total
Democrat	28	17	45
Republican	45	8	53
Independent	2	0	2
Total	75	25	100

A senator is selected at random from this group. Compute the following probabilities.
a. The senator is a male Democrat.
b. The senator is a Republican or a female.
c. The senator is a Republican.
d. The senator is not a Republican.
e. The senator is a Democrat.
f. The senator is an Independent.
g. The senator is a Democrat or an Independent.

46. Graffiti: The following table presents the number of reports of graffiti in each of New York's five boroughs over a one-year period. These reports were classified as being open, closed, or pending.

Borough	Open Reports	Closed Reports	Pending Reports	Total
Bronx	1,121	1,622	80	2,823
Brooklyn	1,170	2,706	48	3,924
Manhattan	744	3,380	25	4,149
Queens	1,353	2,043	25	3,421
Staten Island	83	118	0	201
Total	4,471	9,869	178	14,518

Source: NYC OpenData

A graffiti report is selected at random. Compute the following probabilities.
a. The report is open and comes from Brooklyn.
b. The report is closed or comes from Queens.
c. The report comes from Manhattan.
d. The report does not come from Manhattan.
e. The report is pending.
f. The report is from the Bronx or Staten Island.

47. Add probabilities? In a certain community, 28% of the houses have fireplaces and 51% have garages. Is the probability that a house has either a fireplace or a garage equal to $0.51 + 0.28 = 0.79$? Explain why or why not.

48. Add probabilities? According to the National Health Statistics Reports, 16% of American women have one child, and 21% have two children. Is the probability that a woman has either one or two children equal to $0.16 + 0.21 = 0.37$? Explain why or why not.

Extending the Concepts

49. Mutual exclusivity is not transitive: Give an example of three events A, B, and C, such that A and B are mutually exclusive, B and C are mutually exclusive, but A and C are not mutually exclusive.

50. Complements: Let A and B be events. Express $(A \text{ and } B)^c$ in terms of A^c and B^c.

Answers to Check Your Understanding Exercises for Section 4.2

1. a. 0.342 **b.** 0.361 **c.** 0.469
2. No
3. Yes
4. a. 0.733 **b.** 0.911

SECTION 4.3 Conditional Probability and the Multiplication Rule

Objectives

1. Compute conditional probabilities
2. Compute probabilities by using the General Multiplication Rule
3. Compute probabilities by using the Multiplication Rule for Independent Events
4. Compute the probability that an event occurs at least once

Objective 1 Compute conditional probabilities

Conditional Probability

Approximately 15% of adult men in the United States are more than six feet tall. Therefore, if a man is selected at random, the probability that he is more than six feet tall is 0.15. Now assume that you learn that the selected man is a professional basketball player. With this extra information, the probability that the man is more than six feet tall becomes much greater than 0.15. A probability that is computed with the knowledge of additional information is called a *conditional probability*; a probability computed without such knowledge is called an *unconditional probability*. As this example shows, the conditional probability of an event can be much different than the unconditional probability.

EXAMPLE 4.16

Compute an unconditional probability

Joe, Sam, Eliza, and Maria have been elected to the executive committee of their college's student government. They must choose a chairperson and a secretary. They decide to write each name on a piece of paper and draw two names at random. The first name drawn will be the chairperson, and the second name drawn will be the secretary. What is the probability that Joe is the secretary?

Table 4.3 is a sample space for this experiment. The first name in each pair is the chairperson, and the second name is the secretary.

Table 4.3 Twelve Equally Likely Outcomes

(Joe, Sam)	(Sam, Joe)	(Eliza, Joe)	(Maria, Joe)
(Joe, Eliza)	(Sam, Eliza)	(Eliza, Sam)	(Maria, Sam)
(Joe, Maria)	(Sam, Maria)	(Eliza, Maria)	(Maria, Eliza)

There are 12 equally likely outcomes. Three of them, (Sam, Joe), (Eliza, Joe), and (Maria, Joe), correspond to Joe's being secretary. Therefore, $P(\text{Joe is secretary}) = 3/12 = 1/4$.

EXAMPLE 4.17

Compute a conditional probability

Suppose that Eliza is the first name selected, so she is chairperson. Now what is the probability that Joe is secretary?

Solution

We'll answer this question with intuition first, then show the reasoning. Since Eliza was chosen to be chairperson, she won't be the secretary. That leaves Joe, Sam, and Maria. Each of these three is equally likely to be chosen. Therefore, the probability that Joe is chosen as secretary is 1/3. Note that this probability differs from the probability of 1/4 calculated in Example 4.16.

Now let's look at the reasoning behind this answer. The original sample space, shown in Table 4.3, had 12 outcomes. Once we know that Eliza is chairperson, we know that only three of those outcomes are now possible. Table 4.4 highlights these three outcomes from the original sample space.

Table 4.4

(Joe, Sam)	(Sam, Joe)	(Eliza, Joe)	(Maria, Joe)
(Joe, Eliza)	(Sam, Eliza)	(Eliza, Sam)	(Maria, Sam)
(Joe, Maria)	(Sam, Maria)	(Eliza, Maria)	(Maria, Eliza)

Of the three possible outcomes, only one, (Eliza, Joe), has Joe as secretary. Therefore, given that Eliza is chairperson, the probability that Joe is secretary is $1/3$.

Example 4.17 asked us to compute the probability of an event (that Joe is secretary) after giving us information about another event (that Eliza is chairperson). A probability like this is called a *conditional probability*. The notation for this conditional probability is

$$P(\text{Joe is secretary} \mid \text{Eliza is chairperson})$$

We read this as "the conditional probability that Joe is secretary, given that Eliza is chairperson." It denotes the probability that Joe is secretary, under the assumption that Eliza is chairperson.

Notation Roundup

$P(B \mid A)$ is the probability that event B occurs given that event A has occurred.

DEFINITION

The **conditional probability** of an event B, given an event A, is denoted $P(B \mid A)$.

$P(B \mid A)$ is the probability that B occurs, under the assumption that A occurs.

We read $P(B \mid A)$ as "the probability of B, given A."

The General Method for computing conditional probabilities

In Example 4.17, we computed

$$P(\text{Joe is secretary} \mid \text{Eliza is chairperson}) = \frac{1}{3}$$

Let's take a closer look at the answer of 1/3. The denominator is the number of outcomes that were left in the sample space after it was known that Eliza was chairperson. That is,

$$\text{Number of outcomes where Eliza is chairperson} = 3$$

The numerator is 1, and this corresponds to the one outcome in which Eliza is chairperson and Joe is secretary. That is,

$$\text{Number of outcomes where Eliza is chairperson and Joe is secretary} = 1$$

Therefore, we see that

$$P(\text{Joe is secretary} \mid \text{Eliza is chairperson}) = \frac{\text{Number of outcomes where Eliza is chairperson and Joe is secretary}}{\text{Number of outcomes where Eliza is chairperson}}$$

We can obtain another useful method by recalling that there were 12 outcomes in the original sample space. It follows that

$$P(\text{Eliza is chairperson}) = \frac{3}{12}$$

and

$$P(\text{Eliza is chairperson and Joe is secretary}) = \frac{1}{12}$$

We now see that

$$P(\text{Joe is secretary} \mid \text{Eliza is chairperson}) = \frac{P(\text{Eliza is chairperson and Joe is secretary})}{P(\text{Eliza is chairperson})}$$

This example illustrates the General Method for computing conditional probabilities, which we now state.

The General Method for Computing Conditional Probabilities

The probability of B given A is

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$$

Note that we cannot compute $P(B \mid A)$ if $P(A) = 0$.

When the outcomes in the sample space are equally likely, then

$$P(B \mid A) = \frac{\text{Number of outcomes corresponding to } (A \text{ and } B)}{\text{Number of outcomes corresponding to } A}$$

EXAMPLE 4.18

Use the General Method to compute a conditional probability

Table 4.5 presents the number of U.S. men and women (in millions) 25 years old and older who have attained various levels of education in a recent year.

Table 4.5 Number of Men and Women with Various Levels of Education (in millions)

	Not a high school graduate	High school graduate	Some college, no degree	Associate's degree	Bachelor's degree	Advanced degree
Men	14.0	29.6	15.6	7.2	17.5	10.1
Women	13.7	31.9	17.5	9.6	19.2	9.1

Source: U.S. Census Bureau

A person is selected at random.

a. What is the probability that the person is a man?

b. What is the probability that the person is a man with a bachelor's degree?

c. What is the probability that the person has a bachelor's degree, given that the person is a man?

Solution

a. Each person in the study is an outcome in the sample space. We first compute the total number of people in the study. We'll do this by computing the total number of men, then the total number of women.

$$\text{Total number of men} = 14.0 + 29.6 + 15.6 + 7.2 + 17.5 + 10.1 = 94.0$$

$$\text{Total number of women} = 13.7 + 31.9 + 17.5 + 9.6 + 19.2 + 9.1 = 101.0$$

There are 94.0 million men and 101.0 million women. The total number of people is 94.0 + 101.0 = 195.0 million. We can now compute the probability that a randomly chosen person is a man.

$$P(\text{Man}) = \frac{94.0}{195.0} = 0.4821$$

b. The number of men with bachelor's degrees is found in Table 4.5 to be 17.5 million. The total number of people is 195.0 million. Therefore

$$P(\text{Man with a Bachelor's degree}) = \frac{17.5}{195.0} = 0.08974$$

c. We use the General Method for computing a conditional probability.

$$P(\text{Bachelor's degree}|\text{Man}) = \frac{P(\text{Man with a Bachelor's degree})}{P(\text{Man})} = \frac{17.5/195.0}{94.0/195.0} = 0.1862$$

Check Your Understanding

1. A person is selected at random from the population in Table 4.5.

a. What is the probability that the person is a woman who is a high school graduate?

b. What is the probability that the person is a high school graduate?

c. What is the probability that the person is a woman, given that the person is a high school graduate?

Answers are on page 191.

Objective 2 Compute probabilities by using the General Multiplication Rule

The General Multiplication Rule

The General Method for computing conditional probabilities provides a way to compute probabilities for events of the form "A and B." If we multiply both sides of the equation by $P(A)$, we obtain the General Multiplication Rule.

Explain It Again

The General Multiplication Rule: Use the General Multiplication Rule to compute probabilities of the form $P(A \text{ and } B)$.

The General Multiplication Rule

$$P(A \text{ and } B) = P(A)P(B \mid A)$$

or, equivalently,

$$P(A \text{ and } B) = P(B)P(A \mid B)$$

EXAMPLE 4.19

Use the General Multiplication Rule to compute a probability

Among those who apply for a particular job, the probability of being granted an interview is 0.1. Among those interviewed, the probability of being offered a job is 0.25. Find the probability that an applicant is offered a job.

Solution

Being offered a job involves two events. First, a person must be interviewed; then, given that the person has been interviewed, the person must be offered a job. Using the General Multiplication Rule, we obtain

$$\begin{aligned} P(\text{Offered a job}) &= P(\text{Interviewed})P(\text{Offered a job} \mid \text{Interviewed}) \\ &= (0.1)(0.25) \\ &= 0.025 \end{aligned}$$

Check Your Understanding

2. In a certain city, 70% of high school students graduate. Of those who graduate, 40% attend college. Find the probability that a randomly selected high school student will attend college.

Answer is on page 191.

Objective 3 Compute probabilities by using the Multiplication Rule for Independent Events

Independence

In some cases, the occurrence of one event has no effect on the probability that another event occurs. For example, if a coin is tossed twice, the occurrence of a head on the first toss does not make it any more or less likely that a head will come up on the second toss. Example 4.20 illustrates this fact.

EXAMPLE 4.20

CAUTION

Do not confuse independent events with mutually exclusive events. Two events are independent if the occurrence of one does not affect the probability of the occurrence of the other. Two events are mutually exclusive if the occurrence of one makes it impossible for the other to occur.

Coin tossing probabilities

A fair coin is tossed twice.

a. What is the probability that the second toss is a head?

b. What is the probability that the second toss is a head given that the first toss is a head?

c. Are the answers to parts (a) and (b) different? Does the probability that the second toss is a head change if the first toss is a head?

Solution

a. There are four equally likely outcomes for the two tosses. The sample space is {HH, HT, TH, TT}. Of these, there are two outcomes where the second toss is a head. Therefore, $P(\text{Second toss is H}) = 2/4 = 1/2$.

b. We use the General Method for computing conditional probabilities.

$$P(\text{Second toss is H} \mid \text{First toss is H}) = \frac{\text{Number of outcomes where first toss is H and second is H}}{\text{Number of outcomes where first toss is H}} = \frac{1}{2}$$

c. The two answers are the same. The probability that the second toss is a head does not change if the first toss is a head. In other words,

$$P(\text{Second toss is H} \mid \text{First toss is H}) = P(\text{Second toss is H})$$

In the case of two coin tosses, the outcome of the first toss does not affect the second toss. Events with this property are said to be *independent*.

DEFINITION

Two events are **independent** if the occurrence of one does not affect the probability that the other event occurs.

If two events are not independent, we say they are **dependent**.

In many situations, we can determine whether events are independent just by understanding the circumstances surrounding the events. Example 4.21 illustrates this.

EXAMPLE 4.21

Determine whether events are independent

Determine whether the following pairs of events are independent:

a. A college student is chosen at random. The events are "being a freshman" and "being less than 20 years old."

b. A college student is chosen at random. The events are "born on a Sunday" and "taking a statistics class."

Solution

a. These events are not independent. If the student is a freshman, the probability that the student is less than 20 years old is greater than for a student who is not a freshman.

b. These events are independent. If a student was born on a Sunday, this has no effect on the probability that the student takes a statistics class.

When two events, A and B, are independent, then $P(B \mid A) = P(B)$, because knowing that A occurred does not affect the probability that B occurs. This leads to a simplified version of the Multiplication Rule.

Explain It Again

The Multiplication Rule for Independent Events: Use the Multiplication Rule for Independent Events to compute probabilities of the form $P(A \text{ and } B)$ when A and B are independent.

The Multiplication Rule for Independent Events

If A and B are independent events, then

$$P(A \text{ and } B) = P(A)P(B)$$

This rule can be extended to the case where there are more than two independent events. If A, B, C, ... are independent events, then

$$P(A \text{ and } B \text{ and } C \text{ and } \ldots) = P(A)P(B)P(C)\cdots$$

EXAMPLE 4.22 **Using the Multiplication Rule for Independent Events**

According to recent figures from the U.S. Census Bureau, the percentage of people under the age of 18 was 23.5% in New York City, 25.8% in Chicago, and 26.0% in Los Angeles. If one person is selected from each city, what is the probability that all of them are under 18? Is this an unusual event?

Solution

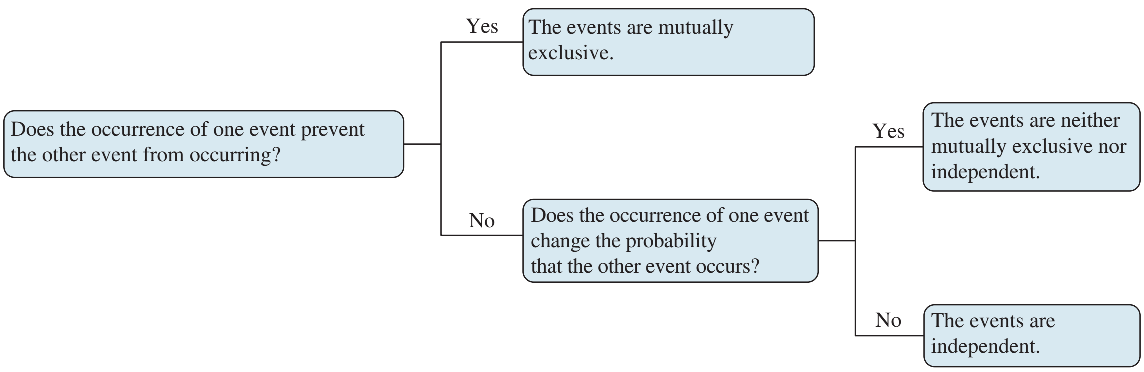
There are three events: person from New York is under 18, person from Chicago is under 18, and person from Los Angeles is under 18. These three events are independent, because the identity of the person chosen from one city does not affect who is chosen in the other cities. We therefore use the Multiplication Rule for Independent Events. Let N denote the event that the person from New York is under 18, and let C and L denote the corresponding events for Chicago and Los Angeles, respectively.

$$P(N \text{ and } C \text{ and } L) = P(N) \cdot P(C) \cdot P(L) = 0.235 \cdot 0.258 \cdot 0.260 = 0.0158$$

The probability is 0.0158. This is an unusual event, if we apply the most commonly used cutoff point of 0.05.

Distinguishing mutually exclusive from independent

Although the mutually exclusive property and the independence property are quite different, in practice it can be difficult to distinguish them. The following diagram can help you to determine whether two events are mutually exclusive, independent, or neither.



Check Your Understanding

3. Two dice are rolled. Each comes up with a number between 1 and 6. Let A be the event that the number on the first die is even, and let B be the event that the number on the second die is 6.

a. Explain why events A and B are independent.

b. Find $P(A)$, $P(B)$, and $P(A \text{ and } B)$.

Answers are on page 191.

Sampling with and without replacement

When we sample two items from a population, we can proceed in either of two ways. We can replace the first item drawn before sampling the second; this is known as **sampling with replacement**. When sampling with replacement, it is possible to draw the same item more than once. The other option is to leave the first item out when sampling the second one;

this is known as **sampling without replacement**. When sampling without replacement, it is impossible to sample an item more than once.

When sampling with replacement, each draw is made from the entire population, so the probability of drawing a particular item on the second draw does not depend on the first draw. In other words, when sampling with replacement, the draws are independent. When sampling without replacement, the draws are not independent. Examples 4.23 and 4.24 illustrate this idea.

EXAMPLE 4.23

Sampling without replacement

A box contains two cards marked with a "0" and two cards marked with a "1" as shown in the following illustration. Two cards will be sampled without replacement from this population.

$$\boxed{\boxed{0}\quad \boxed{0}\quad \boxed{1}\quad \boxed{1}}$$

a. What is the probability of drawing a $\boxed{1}$ on the second draw given that the first draw is a $\boxed{0}$?

b. What is the probability of drawing a $\boxed{1}$ on the second draw given that the first draw is a $\boxed{1}$?

c. Are the first and second draws independent?

Solution

a. If the first draw is a $\boxed{0}$, then the second draw will be made from the population $\boxed{\quad \boxed{0}\quad \boxed{1}\quad \boxed{1}}$. There are three equally likely outcomes, and two of them are $\boxed{1}$. The probability of drawing a $\boxed{1}$ is $2/3$.

b. If the first draw is a $\boxed{1}$, then the second draw will be made from the population $\boxed{\boxed{0}\quad \boxed{0}\quad \quad \boxed{1}}$. There are three equally likely outcomes, and one of them is $\boxed{1}$. The probability of drawing a $\boxed{1}$ is $1/3$.

c. The first and second draws are not independent. The probability of drawing a $\boxed{1}$ on the second draw depends on the outcome of the first draw.

EXAMPLE 4.24

Sampling with replacement

Two items will be sampled with replacement from the population in Example 4.23. Does the probability of drawing a $\boxed{1}$ on the second draw depend on the outcome of the first draw? Are the first and second draws independent?

Solution

Since the sampling is with replacement, then no matter what the first draw is, the second draw will be made from the entire population $\boxed{\boxed{0}\quad \boxed{0}\quad \boxed{1}\quad \boxed{1}}$. Therefore, the probability of drawing a $\boxed{1}$ on the second draw is $2/4 = 0.5$ no matter what the first draw is. Since the probability on the second draw does not depend on the outcome of the first draw, the first and second draws are independent.

The population in Examples 4.23 and 4.24 was very small—only four items. When the population is large, the draws will be nearly independent even when sampled without replacement, as illustrated in Example 4.25.

EXAMPLE 4.25

Sampling without replacement from a large population

A box contains 1000 cards marked with a "0" and 1000 cards marked with a "1," as shown in the following illustration. Two cards will be sampled without replacement from this population.

$$\boxed{1000\ \boxed{0}\text{'s}\qquad 1000\ \boxed{1}\text{'s}}$$

a. What is the probability of drawing a $\boxed{1}$ on the second draw given that the first draw is a $\boxed{0}$?

b. What is the probability of drawing a $\boxed{1}$ on the second draw given that the first draw is a $\boxed{1}$?

c. Are the first and second draws independent? Are they approximately independent?

Solution

a. If the first draw is a $\boxed{0}$, then the second draw will be made from the population $\boxed{999\ \boxed{0}\text{'s} \quad 1000\ \boxed{1}\text{'s}}$. There are 1999 equally likely outcomes, and 1000 of them are $\boxed{1}$. The probability of drawing a $\boxed{1}$ is $1000/1999 = 0.50025$.

b. If the first draw is a $\boxed{1}$, then the second draw will be made from the population $\boxed{1000\ \boxed{0}\text{'s} \quad 999\ \boxed{1}\text{'s}}$. There are 1999 equally likely outcomes, and 999 of them are $\boxed{1}$. The probability of drawing a $\boxed{1}$ is $999/1999 = 0.49975$.

c. The probability of drawing a $\boxed{1}$ on the second draw depends slightly on the outcome of the first draw, so the draws are not independent. However, because the difference in the probabilities is so small (0.50025 versus 0.49975), the draws are approximately independent. In practice, it would be appropriate to treat the two draws as independent.

Example 4.25 shows that when the sample size is small compared to the population size, then items sampled without replacement may be treated as independent. A rule of thumb is that the items may be treated as independent so long as the sample comprises less than 5% of the population.

Explain It Again

Replacement doesn't matter when the population is large: When the sample size is less than 5% of the population, it doesn't matter whether the sampling is done with or without replacement. In either case, we will treat the sampled items as independent.

SUMMARY

- When sampling with replacement, the sampled items are independent.
- When sampling without replacement, if the sample size is less than 5% of the population, the sampled items may be treated as independent.
- When sampling without replacement, if the sample size is more than 5% of the population, the sampled items cannot be treated as independent.

Check Your Understanding

4. A pollster plans to sample 1500 voters from a city in which there are 1 million voters. Can the sampled voters be treated as independent? Explain.

5. Five hundred students attend a college basketball game. Fifty of them are chosen at random to receive a free T-shirt. Can the sampled students be treated as independent? Explain.

Answers are on page 191.

Objective 4 Compute the probability that an event occurs at least once

Solving "at least once" problems by using complements

Sometimes we need to find the probability that an event occurs **at least once** in several independent trials. We can calculate such probabilities by finding the probability of the complement and subtracting from 1. Examples 4.26 and 4.27 illustrate the method.

EXAMPLE 4.26

Find the probability that an event occurs at least once

A fair coin is tossed five times. What is the probability that it comes up heads at least once?

Explain It Again

Solving "at least once" problems: To compute the probability that an event occurs at least once, find the probability that it does not occur at all, and subtract from 1.

Solution

The tosses of a coin are independent, since the outcome of a toss is not affected by the outcomes of other tosses. The complement of coming up heads at least once is coming up tails all five times. We use the Rule of Complements to compute the probability.

$$\begin{aligned}
&P(\text{Comes up heads at least once})\\
&\quad = 1 - P(\text{Does not come up heads at all})\\
&\quad = 1 - P(\text{Comes up tails all five times})\\
&\quad = 1 - P(\text{First toss is T and Second toss is T and} \ldots \text{and Fifth toss is T})\\
&\quad = 1 - P(\text{First toss is T})P(\text{Second toss is T})\cdots P(\text{Fifth toss is T})\\
&\quad = 1 - \left(\frac{1}{2}\right)^5\\
&\quad = \frac{31}{32}
\end{aligned}$$

EXAMPLE 4.27

Find the probability that an event occurs at least once

Items are inspected for flaws by three inspectors. If a flaw is present, each inspector will detect it with probability 0.8. The inspectors work independently. If an item has a flaw, what is the probability that at least one inspector detects it?

Solution

The complement of the event that at least one of the inspectors detects the flaw is that none of the inspectors detects the flaw. We use the Rule of Complements to compute the probability.

We begin by computing the probability that an inspector fails to detect a flaw.

$$P(\text{Inspector fails to detect a flaw}) = 1 - P(\text{Inspector detects flaw}) = 1 - 0.8 = 0.2$$

$$\begin{aligned}
&P(\text{At least one inspector detects the flaw})\\
&\quad = 1 - P(\text{None of the inspectors detects the flaw})\\
&\quad = 1 - P(\text{All three inspectors fail to detect the flaw})\\
&\quad = 1 - P(\text{First fails and second fails and third fails})\\
&\quad = 1 - P(\text{First fails})\,P(\text{Second fails})\,P(\text{Third fails})\\
&\quad = 1 - (0.2)^3\\
&\quad = 0.992
\end{aligned}$$

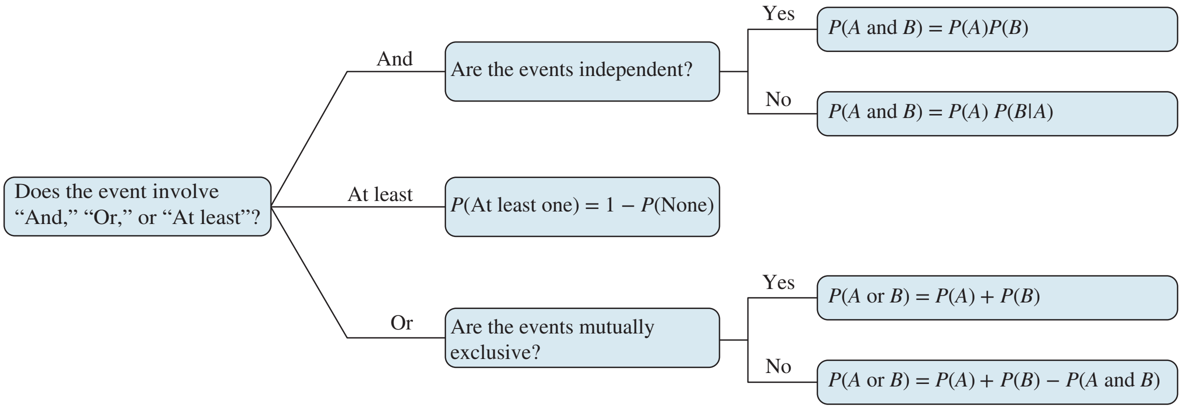
Check Your Understanding

6. An office has three smoke detectors. In case of fire, each detector has probability 0.9 of detecting it. If a fire occurs, what is the probability that at least one detector detects it?

Answer is on page 191.

Determining Which Method to Use

We have studied several methods for finding probabilities of events of the form $P(A \text{ and } B)$, $P(A \text{ or } B)$, and $P(\text{At least one})$. The following diagram can help you to determine the correct method to use for calculating these probabilities.



SECTION 4.3 Exercises

Exercises 1–6 are the Check Your Understanding exercises located within the section.

Understanding the Concepts

In Exercises 7–10, fill in each blank with the appropriate word or phrase.

7. A probability that is computed with the knowledge of additional information is called a __________ probability.

8. The General Multiplication Rule states that $P(A \text{ and } B) =$ __________.

9. When sampling without replacement, if the sample size is less than __________ % of the population, the sampled items may be treated as independent.

10. Two events are __________ if the occurrence of one does not affect the probability that the other event occurs.

In Exercises 11–14, determine whether the statement is true or false. If the statement is false, rewrite it as a true statement.

11. $P(B \mid A)$ represents the probability that A occurs under the assumption that B occurs.

12. If A and B are independent events, then $P(A \text{ and } B) = P(A)P(B)$.

13. When sampling without replacement, it is possible to draw the same item from the population more than once.

14. When sampling with replacement, the sampled items are independent.

Practicing the Skills

15. Let A and B be events with $P(A) = 0.4$, $P(B) = 0.7$, and $P(B \mid A) = 0.3$. Find $P(A \text{ and } B)$.

16. Let A and B be events with $P(A) = 0.6$, $P(B) = 0.4$, and $P(B \mid A) = 0.4$. Find $P(A \text{ and } B)$.

17. Let A and B be events with $P(A) = 0.2$ and $P(B) = 0.9$. Assume that A and B are independent. Find $P(A \text{ and } B)$.

18. Let A and B be events with $P(A) = 0.5$ and $P(B) = 0.7$. Assume that A and B are independent. Find $P(A \text{ and } B)$.

19. Let A and B be events with $P(A) = 0.8$, $P(B) = 0.1$, and $P(B \mid A) = 0.2$. Find $P(A \text{ and } B)$.

20. Let A and B be events with $P(A) = 0.3$, $P(B) = 0.5$, and $P(B \mid A) = 0.7$. Find $P(A \text{ and } B)$.

21. Let A, B, and C be independent events with $P(A) = 0.7$, $P(B) = 0.8$, and $P(C) = 0.5$. Find $P(A \text{ and } B \text{ and } C)$.

22. Let A, B, and C be independent events with $P(A) = 0.4$, $P(B) = 0.9$, and $P(C) = 0.7$. Find $P(A \text{ and } B \text{ and } C)$.

23. A fair coin is tossed four times. What is the probability that all four tosses are heads?

24. A fair coin is tossed four times. What is the probability that the sequence of tosses is HTHT?

25. A fair die is rolled three times. What is the probability that the sequence of rolls is 1, 2, 3?

26. A fair die is rolled three times. What is the probability that all three rolls are 6?

In Exercises 27–30, assume that a student is chosen at random from a class. Determine whether the events *A* and *B* are independent, mutually exclusive, or neither.

27. A: The student is a freshman.
B: The student is a sophomore.

28. A: The student is on the basketball team.
B: The student is more than six feet tall.

29. A: The student is a woman.
B: The student belongs to a sorority.

30. A: The student is a woman.
B: The student belongs to a fraternity.

31. Let A and B be events with $P(A) = 0.25$, $P(B) = 0.4$, and $P(A \text{ and } B) = 0.1$.
a. Are A and B independent? Explain.
b. Compute $P(A \text{ or } B)$.
c. Are A and B mutually exclusive? Explain.

32. Let A and B be events with $P(A) = 0.6$, $P(B) = 0.9$, and $P(A \text{ and } B) = 0.5$.
a. Are A and B independent? Explain.
b. Compute $P(A \text{ or } B)$.
c. Are A and B mutually exclusive? Explain.

33. Let A and B be events with $P(A) = 0.4$, $P(B) = 0.5$, and $P(A \text{ or } B) = 0.6$.
a. Compute $P(A \text{ and } B)$.

b. Are A and B mutually exclusive? Explain.
c. Are A and B independent? Explain.

34. Let A and B be events with $P(A) = 0.5$, $P(B) = 0.3$, and $P(A \text{ or } B) = 0.8$.
a. Compute $P(A \text{ and } B)$.
b. Are A and B mutually exclusive? Explain.
c. Are A and B independent? Explain.

35. A fair die is rolled three times. What is the probability that it comes up 6 at least once?

36. An unfair coin has probability 0.4 of landing heads. The coin is tossed four times. What is the probability that it lands heads at least once?

Working with the Concepts

37. Job interview: Seven people, named Anna, Bob, Chandra, Darnell, Emma, Francisco, and Gina, will be interviewed for a job. The interviewer will choose two at random to interview on the first day. What is the probability that Anna is interviewed first and Darnell is interviewed second?

38. Shuffle: Charles has six songs on a playlist. Each song is by a different artist. The artists are Drake, Post Malone, BTS, Ed Sheeran, Taylor Swift, and Cardi B. He programs his player to play the songs in a random order, without repetition. What is the probability that the first song is by Drake and the second song is by Cardi B?

39. Let's eat: A fast-food restaurant chain has 600 outlets in the United States. The following table categorizes them by city population size and location and presents the number of restaurants in each category. A restaurant is to be chosen at random from the 600 to test market a new menu.

Population of city	Region NE	SE	SW	NW
Under 50,000	30	35	15	5
50,000–500,000	60	90	70	30
Over 500,000	150	25	30	60

a. Given that the restaurant is located in a city with a population over 500,000, what is the probability that it is in the Northeast?
b. Given that the restaurant is located in the Southeast, what is the probability that it is in a city with a population under 50,000?
c. Given that the restaurant is located in the Southwest, what is the probability that it is in a city with a population of 500,000 or less?
d. Given that the restaurant is located in a city with a population of 500,000 or less, what is the probability that it is in the Southwest?
e. Given that the restaurant is located in the South (either SE or SW), what is the probability that it is in a city with a population of 50,000 or more?

40. Senators: The following table displays the 100 senators of the 116th U.S. Congress on January 3, 2019, classified by political party affiliation and gender.

	Male	Female	Total
Democrat	28	17	45
Republican	45	8	53
Independent	2	0	2
Total	75	25	100

A senator is selected at random from this group. Compute the following probabilities.

a. What is the probability that the senator is a woman?
b. What is the probability that the senator is a Republican?
c. What is the probability that the senator is a Republican and a woman?
d. Given that the senator is a woman, what is the probability that she is a Republican?
e. Given that the senator is a Republican, what is the probability that the senator is a woman?

41. Genetics: A geneticist is studying two genes. Each gene can be either dominant or recessive. A sample of 100 individuals is categorized as follows.

Gene 1	Gene 2 Dominant	Recessive
Dominant	56	24
Recessive	14	6

a. What is the probability that in a randomly sampled individual, gene 1 is dominant?
b. What is the probability that in a randomly sampled individual, gene 2 is dominant?
c. Given that gene 1 is dominant, what is the probability that gene 2 is dominant?
d. Two genes are said to be in linkage equilibrium if the event that gene 1 is dominant is independent of the event that gene 2 is dominant. Are these genes in linkage equilibrium?

42. Quality control: A population of 600 semiconductor wafers contains wafers from three lots. The wafers are categorized by lot and by whether they conform to a thickness specification, with the results shown in the following table. A wafer is chosen at random from the population.

Lot	Conforming	Nonconforming
A	88	12
B	165	35
C	260	40

a. What is the probability that a wafer is from Lot A?
b. What is the probability that a wafer is conforming?
c. What is the probability that a wafer is from Lot A and is conforming?
d. Given that the wafer is from Lot A, what is the probability that it is conforming?
e. Given that the wafer is conforming, what is the probability that it is from Lot A?
f. Let E_1 be the event that the wafer comes from Lot A, and let E_2 be the event that the wafer is conforming. Are E_1 and E_2 independent? Explain.

43. Stay in school: In a recent school year in the state of Washington, there were 326,000 high school students. Of these, 167,000 were male. A total of 18,100 students dropped out, and of these, 10,300 were male. A student is chosen at random.

a. What is the probability that the student is male?
b. What is the probability that the student dropped out?
c. What is the probability that the student is male and dropped out?
d. Given that the student is male, what is the probability that he dropped out?
e. Given that the student dropped out, what is the probability that the student is male?

44. **Management:** The Bureau of Labor Statistics reported that 64.5 million women and 74.6 million men were employed. Of the women, 25.8 million had management jobs, and of the men, 25.0 million had management jobs. An employed person is chosen at random.
   a. What is the probability that the person is a female?
   b. What is the probability that the person has a management job?
   c. What is the probability that the person is female and has a management job?
   d. Given that the person is female, what is the probability that she has a management job?
   e. Given that the person has a management job, what is the probability that the person is female?

45. **Asthma:** An article in the journal *Risk Analysis* reported that 5.6% of a population has asthma and that on any given day, 2.7% of asthmatics suffer an asthma attack. A person is chosen at random from this population. What is the probability that this person has an asthma attack on that day?

46. **Smoking and blood pressure:** According to a recent National Health Examination Survey, the probability that a randomly chosen adult is a smoker is 0.24, and given that a person is a smoker, the probability that the person has high blood pressure (systolic blood pressure above 130 mmHg) is 0.25. What is the probability that a randomly chosen adult is a smoker with high blood pressure?

47. **GED:** In a certain high school, the probability that a student drops out is 0.05, and the probability that a dropout gets a high-school equivalency diploma (GED) is 0.25. What is the probability that a randomly selected student gets a GED?

48. **Working for a living:** The Bureau of Labor Statistics reported that the probability that a randomly chosen employed adult worked in a service occupation was 0.17, and given that a person was in a service occupation, the probability that the person was a woman was 0.57. What is the probability that a randomly chosen employed person was a woman in a service occupation?

49. **New car:** At a certain car dealership, the probability that a customer purchases an SUV is 0.20. Given that a customer purchases an SUV, the probability that it is black is 0.25. What is the probability that a customer purchases a black SUV?

50. **Do you know Squidward?** According to a survey by Nickelodeon TV, 88% of children under 13 in Germany recognized a picture of the cartoon character SpongeBob SquarePants. Assume that among those children, 72% also recognized SpongeBob's cranky neighbor Squidward Tentacles. What is the probability that a German child recognized both SpongeBob and Squidward?

51. **Target practice:** Laura and Felipe each fire one shot at a target. Laura has probability 0.5 of hitting the target, and Felipe has probability 0.3. The shots are independent.
   a. Find the probability that both of them hit the target.
   b. Given that Laura hits the target, the probability is 0.1 that Felipe's shot hits the target closer to the bull's-eye than Laura's. Find the probability that Laura hits the target and that Felipe's shot is closer to the bull's-eye than Laura's shot is.

52. **Bowling:** Sarah and Thomas are going bowling. The probability that Sarah scores more than 175 is 0.4, and the probability that Thomas scores more than 175 is 0.2. Their scores are independent.
   a. Find the probability that both score more than 175.
   b. Given that Thomas scores more than 175, the probability that Sarah scores higher than Thomas is 0.3. Find the probability that Thomas scores more than 175 and Sarah scores higher than Thomas.

Rim Light/PhotoLink/Getty Images

53. **Defective components:** A lot of 10 components contains 3 that are defective. Two components are drawn at random and tested. Let *A* be the event that the first component drawn is defective, and let *B* be the event that the second component drawn is defective.
   a. Find $P(A)$.
   b. Find $P(B \mid A)$.
   c. Find $P(A \text{ and } B)$.
   d. Are *A* and *B* independent? Explain.

54. **More defective components:** A lot of 1000 components contains 300 that are defective. Two components are drawn at random and tested. Let *A* be the event that the first component drawn is defective, and let *B* be the event that the second component drawn is defective.
   a. Find $P(A)$.
   b. Find $P(B \mid A)$.
   c. Find $P(A \text{ and } B)$.
   d. Are *A* and *B* independent? Is it reasonable to treat *A* and *B* as though they were independent? Explain.

55. **Multiply probabilities?** In a recent year, 21% of all vehicles in operation were pickup trucks. If someone owns two vehicles, is the probability that they are both pickup trucks equal to $0.21 \times 0.21 = 0.0441$? Explain why or why not.

56. **Multiply probabilities?** A traffic light at an intersection near Jamal's house is red 50% of the time, green 40% of the time, and yellow 10% of the time. Jamal encounters this light in the morning on his way to work and again in the evening on his way home. Is the probability that the light is green both times equal to $0.4 \times 0.4 = 0.16$? Explain why or why not.

57. **Lottery:** Every day, Jorge buys a lottery ticket. Each ticket has probability 0.2 of winning a prize. After seven days, what is the probability that Jorge has won at least one prize?

58. **Car warranty:** The probability that a certain make of car will need repairs in the first six months is 0.3. A dealer sells five such cars. What is the probability that at least one of them will require repairs in the first six months?

59. **Tic-tac-toe:** In the game of tic-tac-toe, if all moves are performed randomly the probability that the game will end in a draw is 0.127. Suppose 10 random games of tic-tac-toe are played. What is the probability that at least one of them will end in a draw?

60. Enter your PIN: The technology consulting company DataGenetics suggests that 17.8% of all four-digit personal identification numbers, or PIN codes, have a repeating digits format such as 2525. Assuming this to be true, if the PIN codes of six people are selected at random, what is the probability that at least one of them will have repeating digits?

Extending the Concepts

Exercises 61–64 refer to the following situation:

A medical test is available to determine whether a patient has a certain disease. To determine the accuracy of the test, a total of 10,100 people are tested. Only 100 of these people have the disease, while the other 10,000 are disease free. Of the disease-free people, 9800 get a negative result, and 200 get a positive result. The 100 people with the disease all get positive results.

61. Find the probability: Find the probability that the test gives the correct result for a person who does not have the disease.

62. Find the probability: Find the probability that the test gives the correct result for a person who has the disease.

63. Find the probability: Given that a person gets a positive result, what is the probability that the person actually has the disease?

64. Why are medical tests repeated? For many medical tests, if the result comes back positive, the test is repeated. Why do you think this is done?

65. Mutually exclusive and independent? Let A and B be events. Assume that neither A nor B can occur; in other words, $P(A) = 0$ and $P(B) = 0$. Are A and B independent? Are A and B mutually exclusive? Explain.

66. Still mutually exclusive and independent? Let A and B be events. Now assume that $P(A) = 0$ but $P(B) > 0$. Are A and B always independent? Are A and B always mutually exclusive? Explain.

67. Mutually exclusive and independent again? Let A and B be events. Now assume that $P(A) > 0$ and $P(B) > 0$. Is it possible for A and B to be both independent and mutually exclusive? Explain.

Answers to Check Your Understanding Exercises for Section 4.3

1. a. 0.164 **b.** 0.315 **c.** 0.519

2. 0.28

3. a. The outcome on one die does not influence the outcome on the other die.

b. $P(A) = 1/2$; $P(B) = 1/6$; $P(A \text{ and } B) = 1/12$

4. Yes, because the sample is less than 5% of the population.

5. No, because the sample is more than 5% of the population.

6. 0.999

SECTION 4.4 Counting

Objectives

1. Count the number of ways a sequence of operations can be performed
2. Count the number of permutations
3. Count the number of combinations

When computing probabilities, it is sometimes necessary to count the number of outcomes in a sample space without being able to list them all. In this section, we will describe several methods for doing this.

Objective 1 Count the number of ways a sequence of operations can be performed

The Fundamental Principle of Counting

The basic rule, which we will call the **Fundamental Principle of Counting**, is presented by means of the following example:

EXAMPLE 4.28

Using the Fundamental Principle of Counting

A certain make of automobile is available in any of three colors—red, blue, or green—and comes with either a large or small engine. In how many ways can a buyer choose a car?

Solution

There are 3 choices of color and 2 choices of engine. A complete list is shown in the tree diagram in Figure 4.5 (page 192), and in the form of a table in Table 4.6 (page 192). The total number of choices is $3 \cdot 2 = 6$.

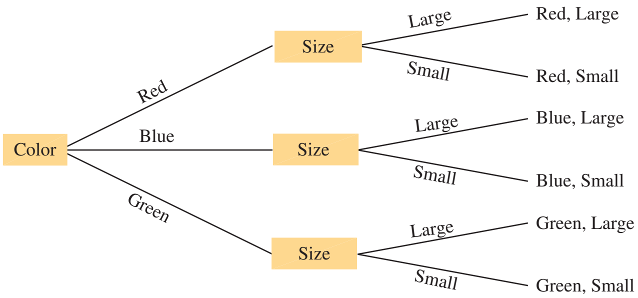


Figure 4.5 Tree diagram illustrating the six choices of color and engine size

Table 4.6 Six Outcomes for the Color and Engine Size

	Large	Small
Red	Red, Large	Red, Small
Blue	Blue, Large	Blue, Small
Green	Green, Large	Green, Small

Stockbyte/Getty Images

To generalize Example 4.28, if there are m choices of color and n choices of engine, the total number of choices is mn. This leads to the Fundamental Principle of Counting.

The Fundamental Principle of Counting

If an operation can be performed in m ways, and a second operation can be performed in n ways, then the total number of ways to perform the sequence of two operations is mn.

If a sequence of several operations is to be performed, the number of ways to perform the sequence is found by multiplying together the numbers of ways to perform each of the operations.

EXAMPLE 4.29

Using the Fundamental Principle of Counting

License plates in a certain state contain three letters followed by three digits. How many different license plates can be made?

Solution

There are six operations in all: choosing three letters and choosing three digits. There are 26 ways to choose each letter and 10 ways to choose each digit. The total number of license plates is therefore

$$26 \cdot 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 = 17{,}576{,}000$$

Check Your Understanding

1. When ordering a certain type of computer, there are three choices of hard drive, four choices for the amount of memory, two choices of video card, and three choices of monitor. In how many ways can a computer be ordered?
2. A quiz consists of three true–false questions and two multiple-choice questions with five choices each. How many different sets of answers are there?

Answers are on page 199.

Objective 2 Count the number of permutations

Permutations

The word *permutation* is another word for *ordering*. When we count the number of permutations, we are counting the number of different ways that a group of items can be ordered.

EXAMPLE 4.30

Counting the number of permutations

Five runners run a race. One of them will finish first, another will finish second, and so on. In how many different orders can they finish?

Solution

We use the Fundamental Principle of Counting. There are five possible choices for the first-place finisher. Once the first-place finisher has been determined, there are four remaining choices for the second-place finisher. Then there are three possible choices for the third-place finisher, two choices for the fourth-place finisher, and only one choice for the fifth-place finisher. The total number of orders of five individuals is

$$\text{Number of orders} = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$$

We say that there are 120 permutations of five individuals.

In Example 4.30, we computed a number of permutations by using the Fundamental Principle of Counting. We can generalize this method, but first we need some notation.

DEFINITION

For any positive integer n, the number $n!$ is pronounced "n factorial" and is equal to the product of all the integers from n down to 1.

$$n! = n(n-1)\cdots(2)(1)$$

By definition, $0! = 1$.

In Example 4.30, we found that the number of permutations of five objects is 5!. This idea holds in general.

> The number of permutations of n objects is $n!$.

Sometimes we want to count the number of permutations of a part of a group. Example 4.31 illustrates the idea.

EXAMPLE 4.31 Counting the number of permutations

Ten runners enter a race. The first-place finisher will win a gold medal, the second-place finisher will win a silver medal, and the third-place finisher will win a bronze medal. In how many different ways can the medals be awarded?

Solution

We use the Fundamental Principle of Counting. There are 10 possible choices for the gold-medal winner. Once the gold-medal winner is determined, there are nine remaining choices for the silver medal. Finally, there are eight choices for the bronze medal. The total number of ways the medals can be awarded is

$$10 \cdot 9 \cdot 8 = 720$$

In Example 4.31, three runners were chosen from a group of 10, then ordered as first, second, and third. This is referred to as a *permutation* of three items chosen from 10.

DEFINITION

A **permutation** of r items chosen from n items is an ordering of the r items. It is obtained by choosing r items from a group of n items, then choosing an order for the r items.

Notation: The number of permutations of r items chosen from n is denoted ${}_nP_r$.

In Example 4.31, we computed ${}_nP_r$ by using the Fundamental Principle of Counting. We can generalize this method by using factorial notation.

> The number of permutations of r objects chosen from n is
>
> $${}_nP_r = n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!}$$

EXAMPLE 4.32

Counting the number of permutations

Five lifeguards are available for duty one Saturday afternoon. There are three lifeguard stations. In how many ways can three lifeguards be chosen and ordered among the stations?

Solution

We are choosing three items from a group of five and ordering them. The number of ways to do this is

$$_5P_3 = \frac{5!}{(5-3)!} = \frac{5!}{2!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1} = 5 \cdot 4 \cdot 3 = 60$$

In some situations, computing the value of $_nP_r$ enables us to determine the number of outcomes in a sample space, and thereby compute a probability. Example 4.33 illustrates the idea.

EXAMPLE 4.33

Using counting to compute a probability

Refer to Example 4.32. The five lifeguards are named Abby, Bruce, Christopher, Donna, and Esmeralda. Of the three lifeguard stations, one is located at the north end of the beach, one in the middle of the beach, and one at the south end. The lifeguard assignments are made at random. What is the probability that Bruce is assigned to the north station, Donna is assigned to the middle station, and Abby is assigned to the south station?

Solution

The outcomes in the sample space consist of all the choices of three lifeguards chosen from five and ordered. From Example 4.32, we know that there are 60 such outcomes. Only one of the outcomes has Bruce, Donna, Abby, in that order. Thus, the probability is $\frac{1}{60}$.

EXAMPLE 4.34

Using counting to compute a probability

Refer to Example 4.33. What is the probability that Bruce is assigned to the north station, Abby is assigned to the south station, and either Donna or Esmeralda is assigned to the middle station?

Solution

As in Example 4.33, the sample space consists of the 60 permutations of three lifeguards chosen from five. Two of these permutations satisfy the stated conditions: Bruce, Donna, Abby; and Bruce, Esmeralda, Abby. So the probability is $\frac{2}{60} = \frac{1}{30}$.

Check Your Understanding

3. A committee of eight people must choose a president, a vice president, and a secretary. In how many ways can this be done?

4. Refer to Exercise 3. Two of the committee members are Ellen and Jose. Assume the assignments are made at random.

a. What is the probability that Jose is president and Ellen is vice president?

b. What is the probability that either Ellen or Jose is president and the other is vice president?

Answers are on page 199.

Objective 3 Count the number of combinations

Combinations

In some cases, when choosing a set of objects from a larger set, we don't care about the ordering of the chosen objects; we care only which objects are chosen. For example, we may not care which lifeguard occupies which station; we might care only which three lifeguards are chosen. Each distinct group of objects that can be selected, without regard to order, is called a **combination**. We will now show how to determine the number of combinations of r objects chosen from a set of n objects. We will illustrate the reasoning with the result of Example 4.32. In that example, we showed that there are 60 permutations of 3 objects chosen from 5. Denoting the objects A, B, C, D, E, Table 4.7 presents a list of all 60 permutations.

Table 4.7 The 60 Permutations of 3 Objects Chosen from 5

ABC	ABD	ABE	ACD	ACE	ADE	BCD	BCE	BDE	CDE
ACB	ADB	AEB	ADC	AEC	AED	BDC	BEC	BED	CED
BAC	BAD	BAE	CAD	CAE	DAE	CBD	CBE	DBE	DCE
BCA	BDA	BEA	CDA	CEA	DEA	CDB	CEB	DEB	DEC
CAB	DAB	EAB	DAC	EAC	EAD	DBC	EBC	EBD	ECD
CBA	DBA	EBA	DCA	ECA	EDA	DCB	ECB	EDB	EDC

Explain It Again

When to use combinations: Use combinations when the order of the chosen objects doesn't matter. Use permutations when the order does matter.

The 60 permutations in Table 4.7 are arranged in 10 columns of 6 permutations each. Within each column, the three objects are the same, and the column contains the 6 different permutations of those three objects. Therefore, each column represents a distinct combination of 3 objects chosen from 5, and there are 10 such combinations. Table 4.7 thus shows that the number of combinations of 3 objects chosen from 5 can be found by dividing the number of permutations of 3 objects chosen from 5, which is $\frac{5!}{(5-3)!}$, by the number of permutations of 3 objects, which is 3!. In summary:

$$\text{The number of combinations of 3 objects chosen from 5 is } \frac{5!}{3!(5-3)!}$$

The number of combinations of r objects chosen from n is often denoted by the symbol ${}_nC_r$. The reasoning above can be generalized to derive an expression for ${}_nC_r$.

The number of combinations of r objects chosen from a group of n objects is

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

EXAMPLE 4.35

Counting the number of combinations

Thirty people attend a certain event, and 5 will be chosen at random to receive prizes. The prizes are all the same, so the order in which the people are chosen does not matter. How many different groups of 5 people can be chosen?

Solution

Since the order of the 5 chosen people does not matter, we need to compute the number of combinations of 5 chosen from 30. This is

$$ {}_{30}C_5 = \frac{30!}{5!(30-5)!} = \frac{30\cdot 29\cdot 28\cdot 27\cdot 26}{5\cdot 4\cdot 3\cdot 2\cdot 1} = 142{,}506 $$

EXAMPLE 4.36

Using counting to compute a probability

Refer to Example 4.35. Of the 30 people in attendance, 12 are men and 18 are women.

a. What is the probability that all the prize winners are men?

b. What is the probability that at least one prize winner is a woman?

Solution

a. The number of outcomes in the sample space is the number of combinations of 5 chosen from 30. We computed this in Example 4.35 to be ${}_{30}C_5 = 142{,}506$. The number of outcomes in which every prize winner is a man is the number of combinations of 5 men chosen from 12 men. This is

$$ {}_{12}C_5 = \frac{12!}{5!(12-5)!} = \frac{12 \cdot 11 \cdot 10 \cdot 9 \cdot 8}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 792 $$

The probability that all prize winners are men is

$$ P(\text{All men}) = \frac{792}{142{,}506} = 0.0056 $$

b. This asks for the probability of at least one woman. We therefore find the probability of the complement; that is, we find the probability that none of the prize winners are women. The probability that none of the prize winners are women is the same as the probability that all of the prize winners are men. In part (a), we computed $P(\text{All men}) = 0.0056$. Therefore,

$$ P(\text{At least one woman}) = 1 - P(\text{All men}) = 1 - 0.0056 = 0.9944 $$

EXAMPLE 4.37 Using counting to compute a probability

A box of lightbulbs contains eight good lightbulbs and two burned-out bulbs. Four bulbs will be selected at random to put into a new lamp. What is the probability that all four bulbs are good?

Solution

The order in which the bulbs are chosen does not matter; all that matters is whether a burned-out bulb is chosen. Therefore, the outcomes in the sample space consist of all the combinations of four bulbs that can be chosen from 10. This number is

$$ {}_{10}C_4 = \frac{10!}{4!(10-4)!} = \frac{3{,}628{,}800}{24 \cdot 720} = 210 $$

To select four good bulbs, we must choose the four bulbs from the eight good bulbs. The number of outcomes that correspond to selecting four good bulbs is therefore the number of combinations of four bulbs that can be chosen from eight. This number is

$$ {}_{8}C_4 = \frac{8!}{4!(8-4)!} = \frac{40{,}320}{24 \cdot 24} = 70 $$

The probability that four good bulbs are selected is therefore

$$ P(\text{Four good bulbs are selected}) = \frac{70}{210} = \frac{1}{3} $$

Check Your Understanding

5. Eight college students have applied for internships at a local firm. Three of them will be selected for interviews. In how many ways can this be done?

6. Refer to Exercise 5. Four of the eight students are from Middle Georgia State University. What is the probability that all three of the interviewed students are from Middle Georgia State University?

Answers are on page 199.

USING TECHNOLOGY

TI-84 PLUS

Evaluating a factorial

Step 1. To evaluate $n!$, enter n on the home screen.

Step 2. Press **MATH**, scroll to the **PRB** menu, and select **4: !**

Step 3. Press **ENTER**.

Permutations and combinations

Step 1. To evaluate $_nP_r$ or $_nC_r$, enter n on the home screen.

Step 2. Press **MATH** and scroll to the **PRB** menu.

- For permutations, select **2: nPr** and press **ENTER** (Figure A).
- For combinations, select **3: nCr** and press **ENTER**.

Step 3. Enter the value for r and press **ENTER**.

The results of $_{12}P_3$ and $_{12}C_3$ are shown in Figure B.

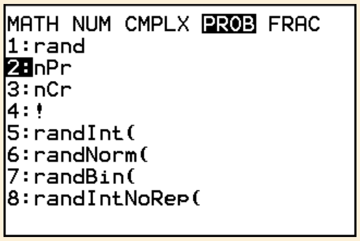


Figure A

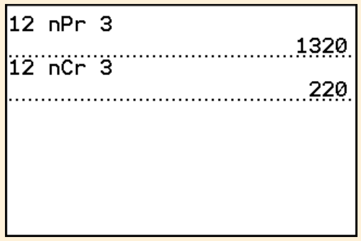


Figure B

EXCEL

Evaluating a factorial

Step 1. To evaluate $n!$, click on a cell in the worksheet and type **=FACT(n)** and press **ENTER**. For example, to compute **12!**, type **=FACT(12)** and press **ENTER**.

Permutations

Step 1. To evaluate $_nP_r$, click on a cell in the worksheet and type **=PERMUT(n,r)**. Press **ENTER**.

Combinations

Step 1. To evaluate $_nC_r$, click on a cell in the worksheet and type **=COMBIN(n,r)**. Press **ENTER**.

SECTION 4.4 Exercises

Exercises 1–6 are the Check Your Understanding exercises located within the section.

Understanding the Concepts

In Exercises 7 and 8, fill in the blank with the appropriate word or phrase:

7. If an operation can be performed in m ways, and a second operation can be performed in n ways, then the total number of ways to perform the sequence of two operations is ________.

8. The number of permutations of six objects is ________.

In Exercises 9 and 10, determine whether the statement is true or false. If the statement is false, rewrite it as a true statement.

9. In a permutation, order is not important.

10. In a combination, order is not important.

Practicing the Skills

In Exercises 11–16, evaluate the factorial.

11. 9! **12.** 5! **13.** 0!

14. 12! **15.** 1! **16.** 3!

In Exercises 17–22, evaluate the permutation.

17. ${}_7P_3$ **18.** ${}_8P_1$ **19.** ${}_{35}P_2$

20. ${}_5P_4$ **21.** ${}_{20}P_0$ **22.** ${}_{45}P_5$

In Exercises 23–28, evaluate the combination.

23. ${}_9C_5$ **24.** ${}_7C_1$ **25.** ${}_{25}C_3$

26. ${}_{10}C_9$ **27.** ${}_{12}C_0$ **28.** ${}_{50}C_{50}$

Working with the Concepts

29. Pizza time: A local pizza parlor is offering a half-price deal on any pizza with one topping. There are eight toppings from which to choose. In addition, there are three different choices for the size of the pizza, and two choices for the type of crust. In how many ways can a pizza be ordered?

30. Books: Josephine has six chemistry books, three history books, and eight statistics books. She wants to choose one book of each type to study. In how many ways can she choose the three books?

31. Playing the horses: In horse racing, one can make a trifecta bet by specifying which horse will come in first, which will come in second, and which will come in third, in the correct order. One can make a box trifecta bet by specifying which three horses will come in first, second, and third, without specifying the order.

a. In an eight-horse field, how many different ways can one make a trifecta bet?

b. In an eight-horse field, how many different ways can one make a box trifecta bet?

32. Ice cream: A certain ice cream parlor offers 15 flavors of ice cream. You want an ice cream cone with three scoops of ice cream, all different flavors.

a. In how many ways can you choose a cone if it matters which flavor is on the top, which is in the middle, and which is on the bottom?

b. In how many ways can you choose a cone if the order of the flavors doesn't matter?

Alex Cao/Getty Images

33. License plates: In a certain state, license plates consist of four digits from 0 to 9 followed by three letters. Assume the numbers and letters are chosen at random. Replicates are allowed.

a. How many different license plates can be formed?

b. How many different license plates have the letters S-A-M in that order?

c. If your name is Sam, what is the probability that your name is on your license plate?

34. Committee: The Student Council at a certain school has 10 members. Four members will form an executive committee consisting of a president, a vice president, a secretary, and a treasurer.

a. In how many ways can these four positions be filled?

b. In how many ways can four people be chosen for the executive committee if it does not matter who gets which position?

c. Four of the people on Student Council are Zachary, Yolanda, Xavier, and Walter. What is the probability that Zachary is president, Yolanda is vice president, Xavier is secretary, and Walter is treasurer?

d. What is the probability that Zachary, Yolanda, Xavier, and Walter are the four committee members?

35. Day and night shifts: A company has hired 12 new employees and must assign 8 to the day shift and 4 to the night shift.

a. In how many ways can the assignment be made?

b. Assume that the 12 employees consist of six men and six women and that the assignments to day and night shift are made at random. What is the probability that all four of the night-shift employees are men?

c. What is the probability that at least one of the night-shift employees is a woman?

36. Keep your password safe: A computer password consists of eight characters. Replications are allowed.

a. How many different passwords are possible if each character may be any lowercase letter or digit?

b. How many different passwords are possible if each character may be any lowercase letter?

c. How many different passwords are possible if each character may be any lowercase letter or digit and at least one character must be a digit?

d. A computer is generating passwords. The computer generates eight characters at random, and each is equally likely to be any of the 26 letters or 10 digits. Replications are allowed. What is the probability that the password will contain all letters?

e. A computer system requires that passwords contain at least one digit. If eight characters are generated at random, what is the probability that they will form a valid password?

37. It's in your genes: Human genetic material (DNA) is made up of sequences of the molecules adenosine (A), guanine (G), cytosine (C), and thymine (T), which are called *bases*. A *codon* is a sequence of three bases. Replicates are allowed, so AAA, CGC, and so forth are codons. Codons are important because each codon causes a different protein to be created.

a. How many different codons are there?

b. How many different codons are there in which all three bases are different?

c. The bases A and G are called *purines*, while C and T are called *pyrimidines*. How many different codons are there in which the first base is a purine and the second and third are pyrimidines?

d. What is the probability that all three bases are different?

e. What is the probability that the first base is a purine and the second and third are pyrimidines?

38. Choosing officers: A committee consists of 10 women and eight men. Three committee members will be chosen as officers.

a. How many different choices are possible?
b. How many different choices are possible if all the officers are to be women?
c. How many different choices are possible if all the officers are to be men?
d. What is the probability that all the officers are women?
e. What is the probability that at least one officer is a man?

39. Texas hold 'em: In the game of Texas hold 'em, a player is dealt two cards (called hole cards) from a standard deck of 52 playing cards. The order in which the cards are dealt does not matter.
a. How many different combinations of hole cards are possible?
b. The best hand consists of two aces. There are four aces in the deck. How many combinations are there in which both cards are aces?
c. What is the probability that a hand consists of two aces?

40. Blackjack: In single-deck casino blackjack, the dealer is dealt two cards from a standard deck of 52. The first card is dealt face down, and the second card is dealt face up.
a. How many dealer hands are possible if it matters which card is face down and which is face up?
b. How many dealer hands are possible if it doesn't matter which card is face down and which is face up?
c. Of the 52 cards in the deck, four are aces and 16 others (kings, queens, jacks, and tens) are worth 10 points each. The dealer has a blackjack if one card is an ace and the other is worth 10 points; it doesn't matter which card is face up and which card is face down. How many different blackjack hands are there?
d. What is the probability that a hand is a blackjack?

41. Lottery: In the Georgia Fantasy 5 Lottery, balls are numbered from 1 to 42. Five balls are drawn. To win the jackpot, you must mark five numbers from 1 to 42 on a ticket, and your numbers must match the numbers on the five balls. The order does not matter. What is the probability that you win?

42. Lottery: In the Colorado Lottery Lotto game, balls are numbered from 1 to 42. Six balls are drawn. To win the jackpot, you must mark six numbers from 1 to 42 on a ticket, and your numbers must match the numbers on the six balls. The order does not matter. What is the probability that you win?

Extending the Concepts

43. Sentence completion: Let A and B be events. Consider the following sentence:

If A and B are ___(i)___, then to find ___(ii)___, ___(iii)___ $P(A)$ and $P(B)$.

Each blank in the sentence can be filled in with either of two choices, as follows:
(i) independent, mutually exclusive
(ii) $P(A \text{ and } B)$, $P(A \text{ or } B)$
(iii) multiply, add
a. In how many ways can the sentence be completed?
b. If choices are made at random for each of the blanks, what is the probability that the sentence is true?

Answers to Check Your Understanding Exercises for Section 4.4

1. 72	**4. a.** 1/56 **b.** 1/28
2. 200	**5.** 56
3. 336	**6.** 1/14

Chapter 4 Summary

Section 4.1: A probability experiment is an experiment that can result in any one of a number of outcomes. The collection of all possible outcomes is a sample space. Sampling from a population is a common type of probability experiment. The population is the sample space, and the individuals in the population are the outcomes. An event is a collection of outcomes from a sample space. The probability of an event is the proportion of times the event occurs in the long run, as the experiment is repeated over and over again. A probability model specifies a probability for every event.

An unusual event is one whose probability is small. There is no hard-and-fast rule about how small a probability has to be for an event to be unusual, but 0.05 is the most commonly used value. The Empirical Method allows us to approximate the probability of an event by repeating a probability experiment many times and computing the proportion of times the event occurs.

Section 4.2: A compound event is an event that is formed by combining two or more events. An example of a compound event is one of the form "A or B." The General Addition Rule is used to compute probabilities of the form $P(A \text{ or } B)$. Two events are mutually exclusive if it is impossible for both events to occur. When two events are mutually exclusive, the Addition Rule for Mutually Exclusive Events can be used to find $P(A \text{ or } B)$. The complement of an event A, denoted A^c, is the event that A does not occur. The Rule of Complements states that $P(A^c)$ is found by subtracting $P(A)$ from 1.

Section 4.3: A conditional probability is a probability that is computed with the knowledge of additional information. Conditional probabilities can be computed with the General Method for computing conditional probabilities. Probabilities of the form $P(A \text{ and } B)$ can be computed with the General Multiplication Rule. If A and B are independent, then $P(A \text{ and } B)$ can be computed with the Multiplication Rule for Independent Events. Two events are independent if the occurrence of one does not affect the probability that the other occurs. When sampling from a population, sampled individuals are independent if the sampling is done with replacement, or if the sample size is less than 5% of the population.

Section 4.4: The Fundamental Principle of Counting states that the total number of ways to perform a sequence of operations is found by multiplying together the numbers of ways of performing each operation. We can compute the number of permutations and combinations of r items chosen from a group of n items. The number of ways that a group of r items can be chosen without regard to order is the number of combinations. The number of ways that a group of r items can be chosen and ordered is the number of permutations. Some sample spaces consist of the permutations or combinations of r items chosen from a group of n items. When working with these sample spaces, we can use the counting rules to compute probabilities.

Vocabulary and Notation

Important Formulas

General Addition Rule:
$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

Multiplication Rule for Independent Events:
$P(A \text{ and } B) = P(A)P(B)$

Addition Rule for Mutually Exclusive Events:
$P(A \text{ or } B) = P(A) + P(B)$

Rule of Complements:
$P(A^c) = 1 - P(A)$

General Method for Computing Conditional Probability:
$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$$

General Multiplication Rule:
$P(A \text{ and } B) = P(A)P(B \mid A) = P(B)P(A \mid B)$

Permutation of *r* items chosen from *n*:
$${}_nP_r = \frac{n!}{(n-r)!}$$

Combination of *r* items chosen from *n*:
$${}_nC_r = \frac{n!}{r!(n-r)!}$$

Chapter Quiz

1. Fill in the blank: The probability that a fair coin lands heads is 0.5. Therefore, we can be sure that if we toss a coin repeatedly, the proportion of times it lands heads will ________.
i. approach 0.5
ii. be equal to 0.5
iii. be greater than 0.5
iv. be less than 0.5

2. A pollster will draw a simple random sample of voters from a large city to ask whether they support the construction of a new light rail line. Assume that there are one million voters in the city, and that 560,000 of them support this proposition. One voter is sampled at random.
a. Identify the sample space.
b. What is the probability that the sampled voter supports the light rail line?

3. State each of the following rules:
a. General Addition Rule
b. Addition Rule for Mutually Exclusive Events
c. Rule of Complements
d. General Multiplication Rule
e. Multiplication Rule for Independent Events

4. The following table presents the results of a survey in which 400 college students were asked whether they listen to music while studying.

	Listen	Do Not Listen
Male	121	78
Female	147	54

a. Find the probability that a randomly selected student does not listen to music while studying.
b. Find the probability that a randomly selected student listens to music or is male.

5. Which of the following pairs of events are mutually exclusive?
i. A: A randomly chosen student is 18 years old. B: The same student is 20 years old.
ii. A: A randomly chosen student owns a red car. B: The same student owns a blue car.

6. In a group of 100 teenagers, 61 received their driver's license on their first attempt on the driver's certification exam and 18 received their driver's license on their second attempt. What is the probability that a randomly selected teenager received their driver's license on their first or second attempt?

7. A certain neighborhood has 100 households. Forty-eight households have a dog as a pet. Of these, 32 also have a cat. Given that a household has a dog, what is the probability that it also has a cat?

8. The owner of a bookstore has determined that 80% of people who enter the store will buy a book. Of those who buy a book, 60% will pay with a credit card. Find the probability that a randomly selected person entering the store will buy a book and pay for it using a credit card.

9. A jar contains 4 red marbles, 3 blue marbles, and 5 green marbles. Two marbles are drawn from the jar one at a time without replacement. What is the probability that the second marble is red, given that the first was blue?

10. A student is chosen at random. Which of the following pairs of events are independent?
i. A: The student was born on a Monday. B: The student's mother was born on a Monday.
ii. A: The student is above average in height. B: The student's mother is above average in height.

11. Individual plays on a slot machine are independent. The probability of winning on any play is 0.38. What is the probability of winning 3 plays in a row?

12. Refer to Exercise 11. Suppose that the slot machine is played 5 times in a row. What is the probability of winning at least once?

13. The Roman alphabet (the one used to write English) consists of five vowels (a, e, i, o, u), along with 21 consonants (we are considering y to be a consonant). Gregory needs to make up a computer password containing seven characters. He wants the first six characters to alternate—consonant, vowel, consonant, vowel, consonant, vowel—with repetitions allowed. Then he wants to use a digit for the seventh character.
a. How many different passwords can he make up?
b. If he makes up a password at random, what is the probability that his password is banana7?

14. A caterer offers 24 different types of dessert. In how many ways can 5 of them be chosen for a banquet if the order doesn't matter?

15. In a standard game of pool, there are 15 balls labeled 1 through 15.
a. In how many ways can the 15 balls be ordered?
b. In how many ways can 3 of the 15 balls be chosen and ordered?

Review Exercises

1. Colored dice: A six-sided die has one face painted red, two faces painted white, and three faces painted blue. Each face is equally likely to turn up when the die is rolled.
a. Construct a sample space for the experiment of rolling this die.
b. Find the probability that a blue face turns up.

2. How are your grades? There were 30 students in last semester's statistics class. Of these, 6 received a grade of A, and 12 received a grade of B. What is the probability that a randomly chosen student received a grade of A or B?

3. Statistics, anyone? Let S be the event that a randomly selected college student has taken a statistics course, and let C be the event that the same student has taken a chemistry course. Suppose $P(S) = 0.4$, $P(C) = 0.3$, and $P(S \text{ and } C) = 0.2$.
a. Find the probability that a student has taken statistics or chemistry.
b. Find the probability that a student has taken statistics given that the student has taken chemistry.

4. Blood types: Human blood may contain either or both of two antigens, A and B. Blood that contains only the A antigen is called type A, blood that contains only the B antigen is called type B, blood that contains both antigens is called type AB, and blood that contains neither antigen is called type O. A certain blood bank has blood from a total of 1200 donors. Of these, 570 have type O blood, 440 have type A, 125 have type B, and 65 have type AB.
a. What is the probability that a randomly chosen blood donor is type O?
b. A recipient with type A blood may safely receive blood from a donor whose blood does not contain the B antigen. What is the probability that a randomly chosen blood donor may donate to a recipient with type A blood?

5. Start a business: Suppose that start-up companies in the area of biotechnology have probability 0.2 of becoming profitable, and that those in the area of information technology have probability 0.15 of becoming profitable. A venture capitalist invests in one firm of each type. Assume the companies function independently.
a. What is the probability that both companies become profitable?
b. What is the probability that at least one of the two companies becomes profitable?

6. **Stop that car:** A drag racer has two parachutes, a main and a backup, that are designed to bring the vehicle to a stop at the end of a run. Suppose that the main chute deploys with probability 0.99, and that if the main fails to deploy, the backup deploys with probability 0.98.
   a. What is the probability that one of the two parachutes deploys?
   b. What is the probability that the backup parachute deploys?
7. **Defective parts:** A process manufactures microcircuits that are used in computers. Twelve percent of the circuits are defective. Assume that three circuits are installed in a computer. Denote a defective circuit by "D" and a good circuit by "G."
   a. List all eight items in the sample space.
   b. What is the probability that all three circuits are good?
   c. The computer will function so long as either two or three of the circuits are good. What is the probability that a computer will function?
   d. If we use a cutoff of 0.05, would it be unusual for all three circuits to be defective?
8. **Music to my ears:** Jeri is listening to the songs on a new CD in random order. She will listen to two different songs and will buy the CD if she likes both of them. Assume there are 10 songs on the CD and that she would like five of them.
   a. What is the probability that she likes the first song?
   b. What is the probability that she likes the second song, given that she liked the first song?
   c. What is the probability that she buys the CD?
9. **Female business majors:** At a certain university, the probability that a randomly chosen student is female is 0.55, the probability that the student is a business major is 0.20, and the probability that the student is female and a business major is 0.15.
   a. What is the probability that the student is female or a business major?
   b. What is the probability that the student is female given that the student is a business major?
   c. What is the probability that the student is a business major given that the student is female?
   d. Are the events "female" and "business major" independent? Explain.
   e. Are the events "female" and "business major" mutually exclusive? Explain.
10. **Heart attack:** The following table presents the number of hospitalizations for myocardial infarction (heart attack) for men and women in various age groups.

Age	Male	Female	Total
18–44	26,828	9,265	36,093
45–64	166,340	68,666	235,006
65–84	155,707	124,289	279,996
85 and up	35,524	57,785	93,309
Total	384,399	260,005	644,404

Source: Agency for Healthcare Research and Quality

   a. What is the probability that a randomly chosen patient is a woman?
   b. What is the probability that a randomly chosen patient is aged 45–64?
   c. What is the probability that a randomly chosen patient is a woman and aged 45–64?
   d. What is the probability that a randomly chosen patient is a woman or aged 45–64?
   e. What is the probability that a randomly chosen patient is a woman given that the patient is aged 45–64?
   f. What is the probability that a randomly chosen patient is aged 45–64 given that the patient is a woman?
11. **Rainy weekend:** Sally is planning to go away for the weekend this coming Saturday and Sunday. At the place she will be going, the probability of rain on any given day is 0.10. Sally says that the probability that it rains on both days is 0.01. She reasons as follows:

$$\begin{aligned} P(\text{Rain Saturday and Rain Sunday}) &= P(\text{Rain Saturday})P(\text{Rain Sunday}) \\ &= (0.1)(0.1) \\ &= 0.01 \end{aligned}$$

   a. What assumption is being made in this calculation?
   b. Explain why this assumption is probably not justified in the present case.
   c. Is the probability of 0.01 likely to be too high or too low? Explain.
12. **Required courses:** A college student must take courses in English, history, mathematics, biology, and physical education. She decides to choose three of these courses to take in her freshman year. In how many ways can this choice be made?
13. **Required courses:** Refer to Exercise 12. Assume the student chooses three courses at random. What is the probability that she chooses English, mathematics, and biology?
14. **Bookshelf:** Luis has six books: a novel, a biography, a dictionary, a self-help book, a statistics textbook, and a comic book.
   a. Luis's bookshelf has room for only three of the books. In how many ways can Luis choose and order three books?
   b. In how many ways may the books be chosen and ordered if he does not choose the comic book?
15. **Bookshelf:** Refer to Exercise 14. Luis chooses three books at random.
   a. What is the probability that the books on his shelf are statistics textbook, dictionary, and comic book, in that order?
   b. What is the probability that the statistics textbook, dictionary, and comic book are the three books chosen, in any order?

Write About It

1. Explain how you could use the law of large numbers to show that a coin is unfair by tossing it many times.
2. When it comes to betting, the chance of winning or losing may be expressed as odds. If there are n equally likely outcomes and m of them result in a win, then the odds of winning are m:$(n - m)$, read "m to $n - m$." For example, suppose that a player rolls a die and wins if the number of dots appearing is either 1 or 2. Since there are two winning outcomes out of six equally likely outcomes, the odds of winning are 2:4.

   Suppose that a pair of dice is rolled and the player wins if it comes up "doubles," that is, if the same number of dots appears on each die. What are the odds of winning?
3. If the odds of an event occurring are 5:8, what is the probability that the event will occur?
4. Explain why the General Addition Rule $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$ may be used even when A and B are mutually exclusive events.
5. Sometimes events are in the form "at least" a given number. For example, if a coin is tossed five times, an event could be getting at least two heads. What would be the complement of the event of getting at least two heads?
6. In practice, one must decide whether to treat two events as independent based on an understanding of the process that creates them. For example, in a manufacturing process that produces electronic circuit boards for calculators, assume that the probability that a board is defective is 0.01. You arrive at the manufacturing plant and sample the next two boards that come off the assembly line. Let A be the event that the first board is defective, and let B be the event that the second board is defective. Describe circumstances under which A and B would not be independent.
7. Describe circumstances under which you would use a permutation.
8. Describe circumstances under which you would use a combination.

In-Class Activities

1. **Law of large numbers:** Each student tosses a coin three times. Compute the proportion of heads among all the students in the class. Repeat a few times, and compute the cumulative proportion of heads each time. Observe how the proportion approaches $1/2$, in accordance with the law of large numbers.
2. **Stop after one head:** Each student tosses a coin until a head appears, then stops. Count the total number of tosses. What proportion of tosses were heads? Repeat a few times, and compute the cumulative proportion of heads each time. Does the proportion approach $1/2$?
3. **Monty Hall Problem:** This is based on the old television program *Let's Make a Deal*, hosted by Monty Hall. There are three doors. Behind one of them is a grand prize, and nothing is behind the other two. You select a door. The host then opens one of the doors you didn't select that has nothing behind it. You are offered the opportunity to switch your selection to the other unopened door. Should you switch or stick with your original selection?

   This game can be simulated with playing cards. Use three cards, one of which is an ace, which represents the grand prize. Divide students into pairs. One student puts the three cards face down in front of the other student, who then guesses where the ace is. The first student, who knows where the ace is, turns up one of the cards that is not an ace. The other student can decide whether or not to switch. Compute the proportion of times students win by switching and by not switching. Should you switch?

Case Study: How Likely Are You To Live To Age 100?

The following table is a *life table*, reproduced from the chapter introduction. With an understanding of some basic concepts of probability, one can use the life table to compute the probability that a person of a given age will still be alive a given number of years from now. Life insurance companies use this information to determine how much to charge for life insurance policies.

United States Life Table, Total Population

Age Interval	Proportion Surviving	Age Interval	Proportion Surviving
0–10	0.99123	50–60	0.94010
10–20	0.99613	60–70	0.86958
20–30	0.99050	70–80	0.70938
30–40	0.98703	80–90	0.42164
40–50	0.97150	90–100	0.12248

Source: Centers for Disease Control and Prevention

The column labeled “Proportion Surviving” gives the proportion of people alive at the beginning of an age interval who will still be alive at the end of the age interval. For example, among those currently age 20, the proportion who will still be alive at age 30 is 0.99050, or 99.050%. We will begin by computing the probability that a person lives to any of the ages 10, 20, ... , 100.

The first number in the column is the probability that a person lives to age 10. So

$$P(\text{Alive at age } 10) = 0.99123$$

The key to using the life table is to realize that the rest of the numbers in the “Proportion Surviving” column are conditional probabilities. They are probabilities that a person is alive at the end of the age interval, given that they were alive at the beginning of the age interval. For example, the row labeled “20–30” contains the conditional probability that someone alive at age 20 will be alive at age 30:

$$P(\text{Alive at age } 30 \mid \text{Alive at age } 20) = 0.99050$$

In Exercises 1–5, compute the probability that a person lives to a given age.

1. From the table, find the conditional probability $P(\text{Alive at age } 20 \mid \text{Alive at age } 10)$.
2. Use the result from Exercise 1 along with the result $P(\text{Alive at age } 10) = 0.99123$ to compute $P(\text{Alive at age } 20)$.
3. Use the result from Exercise 2 along with the appropriate number from the table to compute $P(\text{Alive at age } 30)$.
4. Use the result from Exercise 3 along with the appropriate number from the table to compute $P(\text{Alive at age } 40)$.
5. Compute the probability that a person is alive at ages 50, 60, 70, 80, 90, and 100.

In Exercises 1–5, we computed the probability that a newborn lives to a given age. Now let’s compute the probability that a person aged x lives to age y. We’ll illustrate this with an example to compute the probability that a person aged 20 lives to age 100. This is the conditional probability that a person lives to age 100, given that the person has lived to age 20.

We want to compute the conditional probability

$$P(\text{Alive at age } 100 \mid \text{Alive at age } 20)$$

Using the definition of conditional probability, we have

$$P(\text{Alive at age } 100 \mid \text{Alive at age } 20) = \frac{P(\text{Alive at age } 100 \text{ and Alive at age } 20)}{P(\text{Alive at age } 20)}$$

You computed $P(\text{Alive at age } 20)$ in Exercise 2. Now we need to compute $P(\text{Alive at age } 100 \text{ and Alive at age } 20)$. The key is to realize that anyone who is alive at age 100 was also alive at age 20. Therefore,

$$P(\text{Alive at age } 100 \text{ and Alive at age } 20) = P(\text{Alive at age } 100)$$

Therefore,

$$P(\text{Alive at age } 100 \mid \text{Alive at age } 20) = \frac{P(\text{Alive at age } 100)}{P(\text{Alive at age } 20)}$$

In general, for $y > x$,

$$P(\text{Alive at age } y \mid \text{Alive at age } x) = \frac{P(\text{Alive at age } y)}{P(\text{Alive at age } x)}$$

6. Find the probability that a person aged 20 is still alive at age 100.
7. Find the probability that a person aged 50 is still alive at age 70.
8. Which is more probable, that a person aged 20 is still alive at age 50, or that a person aged 50 is still alive at age 60?

9. A life insurance company sells term insurance policies. These policies pay $100,000 if the policyholder dies before age 70, but pay nothing if a person is still alive at age 70. If a person buys a policy at age 40, what is the probability that the insurance company does not have to pay?