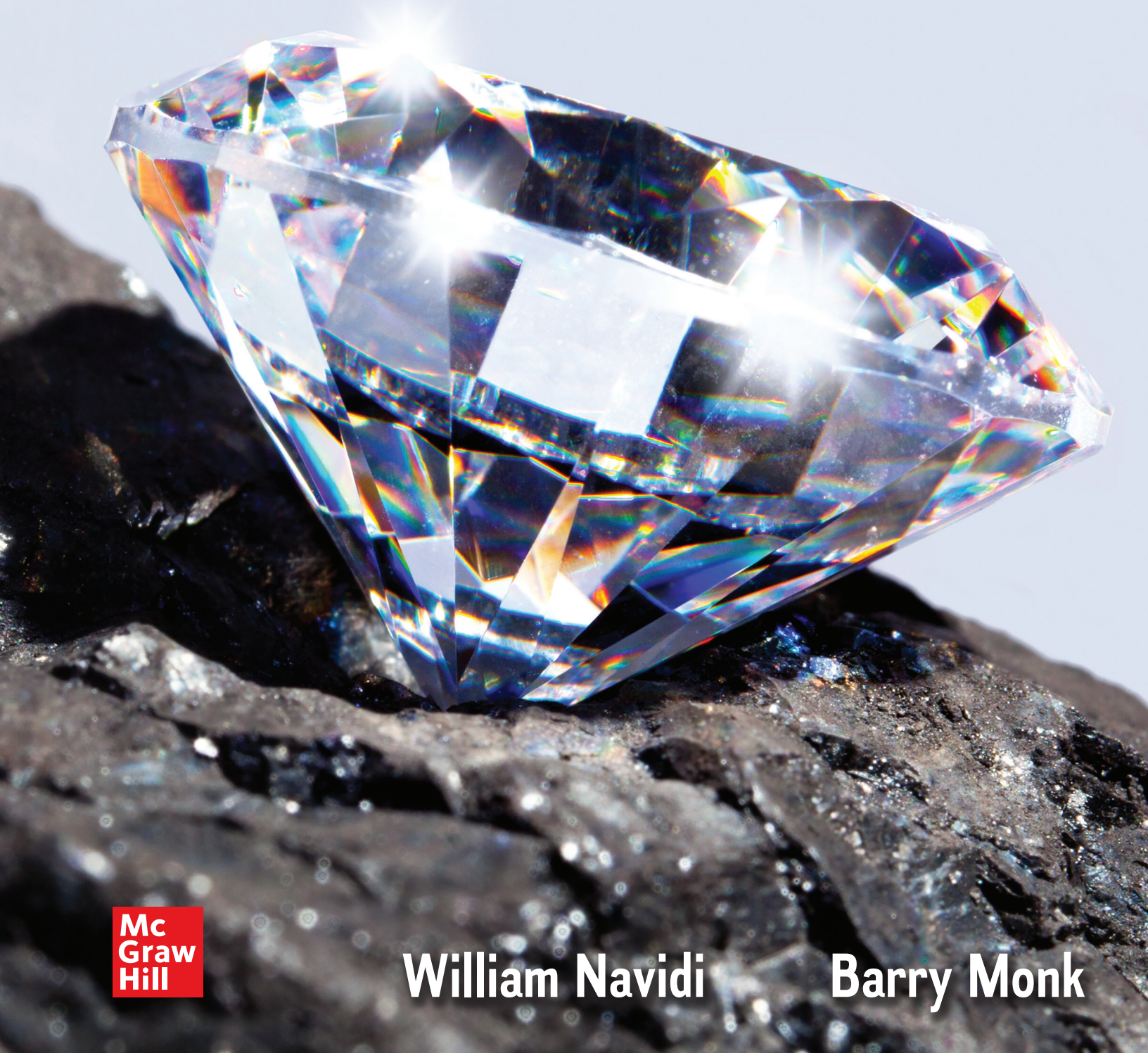
Elementary STATISTICS

Fourth Edition



William Navidi

Barry Monk

Mc Graw Hill

Elementary STATISTICS

Fourth Edition

William Navidi
Colorado School of Mines

Barry Monk
Middle Georgia State University

ELEMENTARY STATISTICS, FOURTH EDITION

Published by McGraw Hill LLC, 1325 Avenue of the Americas, New York, NY 10121. Copyright ©2022 by McGraw Hill LLC. All rights reserved. Printed in the United States of America. Previous editions ©2019, 2016, and 2013. No part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written consent of McGraw Hill LLC, including, but not limited to, in any network or other electronic storage or transmission, or broadcast for distance learning.

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

This book is printed on acid-free paper.

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

ISBN 978-1-260-72787-6 (bound edition)
MHID 1-260-72787-4 (bound edition)

ISBN 978-1-264-13640-7 (loose-leaf edition)
MHID 1-264-13640-4 (loose-leaf edition)

ISBN 978-1-264-13635-3 (instructor's edition)
MHID 1-264-13635-8 (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: ©*RTimages/Shutterstock*
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: Elementary statistics / William Navidi, Colorado School of Mines, Barry Monk, Middle Georgia State University
Description: Fourth edition. | Dubuque : McGraw-Hill Education, [2022] | Includes index.
Identifiers: LCCN 2020024274 (print) | LCCN 2020024275 (ebook) | ISBN 9781260727876 (hardcover) | ISBN 9781264136407 (spiral bound) | ISBN 9781264136353 (hardcover) | ISBN 9781264136384 (ebook) | ISBN 9781264136414 (ebook other)
Subjects: LCSH: Mathematical statistics—Textbooks.
Classification: LCC QA276.12 .N385 2022 (print) | LCC QA276.12 (ebook) | DDC 519.5—dc23
LC record available at https://lccn.loc.gov/2020024274
LC ebook record available at https://lccn.loc.gov/2020024275

The Internet addresses listed in the text were accurate at the time of publication. The inclusion of a website does not indicate an endorsement by the authors or McGraw Hill LLC, and McGraw Hill LLC does not guarantee the accuracy of the information presented at these sites.

mheducation.com/highered

To Catherine, Sarah, and Thomas

—William Navidi

To Shaun

—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 at 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

CHAPTER 1 Basic Ideas 1

CHAPTER 2 Graphical Summaries of Data 37

CHAPTER 3 Numerical Summaries of Data 97

CHAPTER 4 Summarizing Bivariate Data 165

CHAPTER 5 Probability 213

CHAPTER 6 Discrete Probability Distributions 263

CHAPTER 7 The Normal Distribution 303

CHAPTER 8 Confidence Intervals 371

CHAPTER 9 Hypothesis Testing 431

CHAPTER 10 Two-Sample Confidence Intervals 505

CHAPTER 11 Two-Sample Hypothesis Tests 539

CHAPTER 12 Tests with Qualitative Data 587

CHAPTER 13 Inference in Linear Models 613

CHAPTER 14 Analysis of Variance 661

CHAPTER 15 Nonparametric Statistics 697

Mc Graw Hill
501-1111

Contents

Preface xv
Acknowledgments xix
Index of Applications xxvii

CHAPTER 1 Basic Ideas 1

Tatiana Grozetskaya/Shutterstock

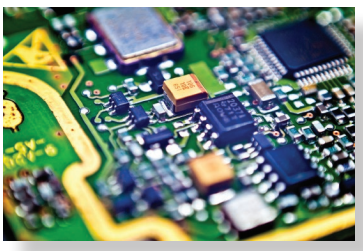
1.1 Sampling 2
1.2 Types of Data 12
1.3 Design of Experiments 19
1.4 Bias in Studies 27
Chapter 1 Summary 31
Vocabulary and Notation 31
Chapter Quiz 32
Review Exercises 32

CHAPTER 2 Graphical Summaries of Data 37

Mike Flippo/123RF

2.1 Graphical Summaries for Qualitative Data 38
2.2 Frequency Distributions and Their Graphs 51
2.3 More Graphs for Quantitative Data 71
2.4 Graphs Can Be Misleading 83
Chapter 2 Summary 90
Vocabulary and Notation 90
Chapter Quiz 91
Review Exercises 92

CHAPTER 3 Numerical Summaries of Data 97



adistock/123RF

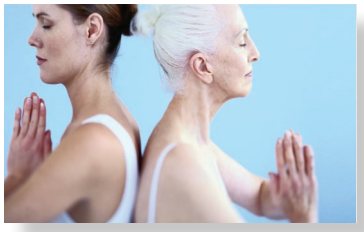
3.1 Measures of Center 98
3.2 Measures of Spread 118
3.3 Measures of Position 139
Chapter 3 Summary 158
Vocabulary and Notation 158
Chapter Quiz 159
Review Exercises 160

CHAPTER 4 Summarizing Bivariate Data 165

Stockbyte/PunchStock

4.1 Correlation 166
4.2 The Least-Squares Regression Line 178
4.3 Features and Limitations of the Least-Squares Regression Line 190
Chapter 4 Summary 204
Vocabulary and Notation 204
Chapter Quiz 205
Review Exercises 206

CHAPTER 5 Probability 213


Ryan McVay/Getty Images

5.1 Basic Concepts of Probability 214
5.2 The Addition Rule and the Rule of Complements 226
5.3 Conditional Probability and the Multiplication Rule 235
5.4 Counting 248
Chapter 5 Summary 257
Vocabulary and Notation 257
Chapter Quiz 258
Review Exercises 259

CHAPTER 6 Discrete Probability Distributions 263

Andriy Popov/123RF

6.1 Random Variables 264
6.2 The Binomial Distribution 278
6.3 The Poisson Distribution 290
Chapter 6 Summary 296
Vocabulary and Notation 297
Chapter Quiz 297
Review Exercises 298

CHAPTER 7 The Normal Distribution 303

Steve Allen/Getty Images

7.1 The Standard Normal Curve 304
7.2 Applications of the Normal Distribution 321
7.3 Sampling Distributions and the Central Limit Theorem 334
7.4 The Central Limit Theorem for Proportions 343
7.5 The Normal Approximation to the Binomial Distribution 349
7.6 Assessing Normality 356
Chapter 7 Summary 366
Vocabulary and Notation 367
Chapter Quiz 367
Review Exercises 368

CHAPTER 8 Confidence Intervals 371

Schroptschop/E+/Getty Images

8.1 Confidence Intervals for a Population Mean, Standard Deviation Known 372

8.2 Confidence Intervals for a Population Mean, Standard Deviation Unknown 385

8.3 Confidence Intervals for a Population Proportion 405

8.4 Confidence Intervals for a Standard Deviation 417

8.5 Determining Which Method to Use 422

Chapter 8 Summary 425

Vocabulary and Notation 426

Chapter Quiz 426

Review Exercises 427

CHAPTER 9 Hypothesis Testing 431

Murat Taner/Getty Images

9.1 Basic Principles of Hypothesis Testing 432

9.2 Hypothesis Tests for a Population Mean, Standard Deviation Known 439

9.3 Hypothesis Tests for a Population Mean, Standard Deviation Unknown 458

9.4 Hypothesis Tests for Proportions 478

9.5 Hypothesis Tests for a Standard Deviation 488

9.6 Determining Which Method to Use 492

9.7 Power 494

Chapter 9 Summary 499

Vocabulary and Notation 499

Chapter Quiz 500

Review Exercises 500

CHAPTER 10 Two-Sample Confidence Intervals 505

Ridofranz/Getty Images

10.1 Confidence Intervals for the Difference Between Two Means: Independent Samples 506

10.2 Confidence Intervals for the Difference Between Two Means: Paired Samples 517

10.3 Confidence Intervals for the Difference Between Two Proportions 525

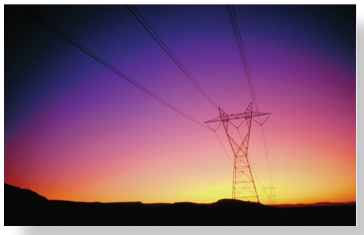
Chapter 10 Summary 533

Vocabulary and Notation 533

Chapter Quiz 534

Review Exercises 535

CHAPTER 11 Two-Sample Hypothesis Tests 539



Malcolm Fife/age fotostock

11.1 Hypothesis Tests for the Difference Between Two Means: Independent Samples 540
11.2 Hypothesis Tests for the Difference Between Two Means: Paired Samples 551
11.3 Hypothesis Tests for the Difference Between Two Proportions 560
11.4 Hypothesis Tests for Two Population Standard Deviations 570
11.5 The Multiple Testing Problem 577
Chapter 11 Summary 581
Vocabulary and Notation 581
Chapter Quiz 581
Review Exercises 582

CHAPTER 12 Tests with Qualitative Data 587

Mark Scott/Getty Images

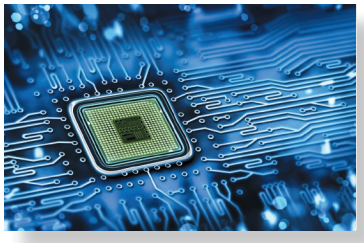
12.1 Testing Goodness of Fit 588
12.2 Tests for Independence and Homogeneity 597
Chapter 12 Summary 607
Vocabulary and Notation 607
Chapter Quiz 607
Review Exercises 608

CHAPTER 13 Inference in Linear Models 613

Vincent Ting/Getty Images

13.1 Inference on the Slope of the Regression Line 614
13.2 Inference About the Response 630
13.3 Multiple Regression 636
Chapter 13 Summary 652
Vocabulary and Notation 653
Chapter Quiz 653
Review Exercises 654

CHAPTER 14 Analysis of Variance 661



crstrbrt/123RF

14.1 One-Way Analysis of Variance 662
14.2 Two-Way Analysis of Variance 678
Chapter 14 Summary 691
Vocabulary and Notation 691
Chapter Quiz 692
Review Exercises 693

CHAPTER 15 Nonparametric Statistics 697

CUHRIG/Getty Images

15.1 The Sign Test 698
15.2 The Rank-Sum Test 703
15.3 The Signed-Rank Test 709
Chapter 15 Summary 714
Vocabulary and Notation 714
Chapter Quiz 715
Review Exercises 715

Appendix A Tables A-1
Appendix B TI-84 PLUS Stat Wizards B-1
Answers to Odd-Numbered Exercises (Student edition only) SA-1
Answers to Selected Exercises (Instructor's edition only) IA-1
Index I-1
Tables A.2 & A.3 T-1
Important Formulas IF-1

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 presented 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. The material on two-sample inference is divided into two chapters—Chapter 10 on two-sample confidence intervals, and Chapter 11 on two-sample hypothesis tests. This gives instructors the option of covering all the material on confidence intervals before starting hypothesis testing, by covering Chapter 10 immediately after Chapter 8.

We have placed the material on descriptive statistics for bivariate data immediately following descriptive statistics for univariate data. Those who wish to cover bivariate description and inference together may postpone Chapter 4 until sometime before covering Chapter 13.

Instructors differ widely in their preferences regarding the depth of coverage of probability. A light treatment of the subject may be obtained by covering Section 5.1 and skipping the rest of the chapter. More depth can be obtained by covering Sections 5.2 and 5.3. Section 5.4 on counting can be included for an even more comprehensive treatment.

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 fourth edition of the book is intended to extend the strengths of the third. 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 have 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 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.
- Material on hypothesis tests for the population correlation has been added.
- 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. 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 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 ***Elementary 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 Community 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 L. Anderson, *Northwest Vista College*
Peter Arvanites, *Rockland Community College*
Diana Asmus, *Greenville Technical College*
John Avioli, *Christopher Newport University*
Heather A. Barker, *Elon 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*
Gabi Booth, *Daytona 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*
Dr. Angela Everett, *Chattanooga State 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*
Marnie Francisco, *Foothill 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 A. Hendricks, *Perimeter Georgia State University, College*
Jada Hill, *Richland College*
Mary Hill, *College of DuPage*
William Huepenbecker, *Bowling Green State University-Firelands*
Jenny Hughes, *Columbia Basin College*
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*
Dr. Ryan H. Kasha, Ph.D., *Valencia College*
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*
Alex Kolesnik, *Ventura College*
JoAnn Kump, *West Chester University*
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*
Habid Maagoul, *Northern Essex Community 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*
Angie Matthews, *Broward College*
Catherine Matos, *Clayton State University*
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 Kyei 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*
O. Dale Saylor, *University of South Carolina Upstate*
Vicki Schell, *Pensacola State College*
Angela Schirck, *Broward 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*
Aileen Solomon, *Trident Technical College*
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*
Erwin Walker, *Clemson University*
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*

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

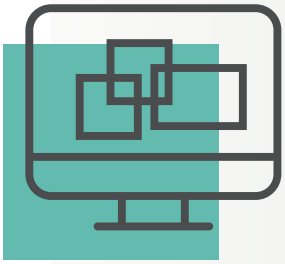
ALEKS® creates an optimized path with an ongoing cycle of learning and assessment, celebrating students' small wins along the way with positive real-time feedback. Rooted in research and analytics, ALEKS® improves student outcomes by fostering better preparation, increased motivation and knowledge retention.

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

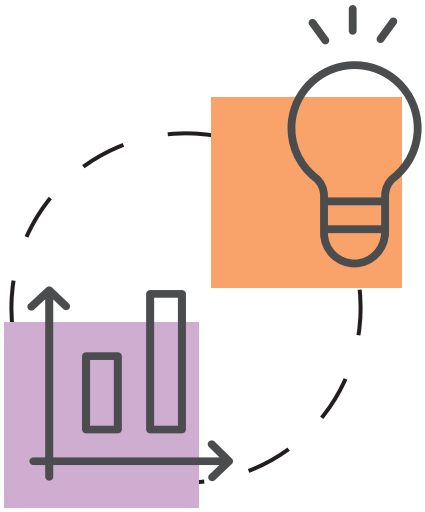
ALEKS® enables you to structure your course regardless of your instruction style and format. From a traditional classroom, to various corequisite models, to an online prep course before the start of the term, ALEKS® can supplement your instruction or play a lead role in delivering the content.

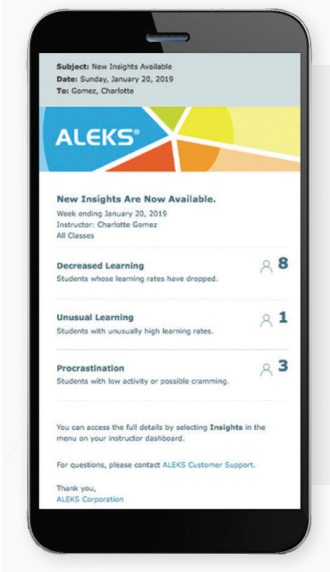
Teaching online? Respondus integration with ALEKS® supports secure testing with LockDown Browser and Respondus Monitor for Tests, Quizzes, and Scheduled Knowledge Checks.

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

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.

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.

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.

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.

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.

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.

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, co-written by 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

Agricultural/gardening/farming applications. *See also* Forestry applications

absorption rate of pesticides into skin, 675
acid levels of soil, 693
bean plants, 576
chicken weights, 332
fertilizer, 580
fertilizer effect on maple sap production, 675, 688
fertilizer effect on orange trees, 437, 514, 694
fertilizer effect on tomato yield, 26
great pumpkins, 187–188, 200
herbicides and phosphate concentration of bean plants, 535
pesticides and fruit yields, 690
phosphorus in soil, 677
seed germination rate, 355, 532
temperature and water evaporation, 628–629, 635
wheat yields, 703

Archaeology applications

mummy's curse and life spans, 549

Automotive/motor vehicle applications

accidents at intersections, 531
accidents per year, 117, 296
air pollution caused by motor vehicles, 79–80, 523, 535, 568, 558–559
battery lifetimes, 368
BMW prices, 68, 79
brake wear and mileage, 522
car repairs, 233–234, 247, 355
car speeds and stopping distance, 202–203
car speeds and street noise level, 176, 628, 634
car speeds on highways, 333, 494
coating for gears, 580
delivery truck load weights, 369
don't drink and drive, 11, 114
drag-racing parachutes, 259
emissions at sea level and high altitude, 156–157, 535, 568
emissions testing, 288, 647–648
fuel efficiency, 32, 133, 524
gas mileage, 33, 96, 133, 206, 341, 402, 707–708
gas prices, 112
hybrid cars, 92
license plate choices, 255
new car choice probabilities, 247
new car sales, 49–50
new car satisfaction, 349
new truck sales, 50
parking on campus, 11
phones and driving, 33
pickup truck ownership, 247
pillars giving head protection in cars, 677
risky drivers, 224
tire lifetimes, 332, 498
tires and fuel economy, 33, 524
tires and number of miles traveled, 558, 694–695
tires with low air pressure, 275
tire tread depth, 716–717
truck pollution, 523, 558–559
truck weight and gas mileage, 206, 654
used car ages, 404
years between accidents, 516, 550
yellow light and driver behavior, 569

Aviation applications

airline ticketing policy, 355
flight delays, 288

Behavioral study applications. *See also* Psychological applications

dog weights, 438
don't drink and drive, 11, 114
nicotine patch to quit smoking, 486
weight loss during counseling, 474

Beverage applications

breweries in selected states, 112, 134
caffeine in coffee, 437–438
calories in milkshakes, 384
carbohydrates in espresso beverages, 155
coffee prices, 708
daily coffee consumption, 347
don't drink and drive, 11, 114
energy drinks, 383–384, 515
moisture content of coffee beans, 498
strength of aluminum cans, 370
volume of beverage in can, 333, 474, 491, 717
volume of beverage in large-sized sodas, 676

Biology/life science applications. *See also* Genetics/gender applications

age of babies beginning to talk, 401
bird brain weight *versus* body weight, 189
blood typing system for humans, 46, 259, 288
butterfly wingspan *versus* lifespan, 177, 188, 627–628, 634
Down Syndrome births, 296
forearm length and height, 206
gestational age and baby weights, 207
height and age, 455
heights of boys and men, 116, 138, 333, 524, 559
heights of college students, 66, 117
heights of fathers and sons, 627, 634
heights of women, 110, 116, 138, 333
random sampling of animal groups, 11
rats running mazes, 66, 475
shrimp habitat and water pollution, 535
weights/lengths of fish, 154, 331–332, 365, 437
weights of animals, 457, 476
weights of girls and women, 115
weights of newborn babies, 137, 207, 421–422, 549, 716
weights of older babies, 289, 331, 355, 383, 401, 474, 491, 515

Business applications

advertisement views, 498
advertising expenditures, 113–114, 161–162
age discrimination, 605
bond prices, 136
cable subscriptions, 87
CEO salaries, 402
college graduates working at home, 414
company revenue, 111
confidence in banks, 486
corporate profits, 160, 162
credit card balances, 457
credit card payments, 425
credit scores, 203, 474, 651–652
day and night shifts, 255
days per year spent traveling, 438
Dow Jones Industrial Average, 87, 136
economy in the future, 18, 416, 533, 597
extending banking hours, 30
financial literacy, 713
finding job after job loss, 299
flextime in company to reduce sick days, 582
free desserts to increase business, 475
gas prices, 112
government employee percentages, 348
hourly wages and/or years employed, 70, 116, 177–178, 385
inflation rate and unemployment rate, 210–211
interest in economics, 91, 487
job changes, 416
job satisfaction, 224, 501
job security, 48
life expectancy, 33
life insurance, 261–262, 277
management jobs and gender, 247
marketing firm survey, 19, 498
mean and median salaries, 117
mean federal income tax, 341
mean state income tax, 368
multiple jobs held, 348
on-site day care, 12
opinions of service at a store, 11, 12
personal incomes, 115, 116, 157
phone company market shares, 92
prime interest rate, 189
random-digit dialing as sampling method, 30
restaurant customers on Mondays *versus* Tuesdays, 716
retail spending, 11–12, 79
return on investment, 136
sales commissions, 583
sales tax increase, 487
sales tax on internet purchases, 487

satisfaction with financial situation, 532, 596
saving by U.S. residents, 88
science careers, 494
spreadsheet testing, 583, 676–677
start-up probabilities, 259, 277
stock market averages and first digits of the number, 301
stock market crash, 87
stock prices, 87, 113, 135, 161, 383, 488, 535
streaming video fees, 487
stress at work, 289, 355
tax return refunds, 12
tax returns and first digits of the number, 302
tax returns for which no tax paid, 341
unemployment rate in U.S., 80, 200
U.S. imports and exports by month, 201
weekly earnings, 348
women in service occupations, 247
words typed per minute, 384
workforce increases during economic downturn, 416

Chemicals/chemistry applications

acid levels of soil, 693
adhesive bond strength, 689
ammonium in water wells, 333
bacteria in pond water, 295–296
benzene in wastewater, 514, 549, 583
calories in fat, 629
carbon monoxide concentration, 404
chemicals in well water, 693–694
cleaning solutions, 689
gypsum and acid level, 693
gypsum and electrical conductivity, 693, 694
lead in water, 415, 456
mercury in lakes, 494
mineral content of kale, 403
mineral content of spinach, 402, 422
ozone concentration, 189
paint drying time, 655
paint lifetime, 650
phosphorus in soil, 677
soap lather, 688–689
soap weights, 162
suspended particles, 299
temperature, concentration, and time in chemical reactions, 651
temperature and water evaporation, 628–629, 635

Computer applications. *See*

Electronics/computer applications

Construction/home improvement/home purchases and sales applications

air conditioner costs, 716
apartment rents, 161, 341, 502, 703, 713
asphalt-rubber concrete beam strength, 689
breaking strength of wood boards, 676
building new parking structure, 427
construction site water contamination, 415
electricity bills/prices, 138, 332
expansion of concrete over time, 628, 635
fireplaces and garages, 235
household income and energy consumed, 206, 655
house prices, 81, 113, 135, 456, 474
house size and energy expenditures, 201
house size and selling price, 202
house sizes, 477, 494
mortgage payments/rates, 177, 188, 402, 422
radon levels, 203, 332
recovering from heart attacks in smaller houses, 26
sales price *versus* rental price of homes, 202
smoke detectors in homes, 289
vacant apartment rates, 80–81

Crime

graffiti by borough, 235
illegal drug use in high school, 30
murder rates in cities, 67–68
police department effectiveness, 30
police numbers and crimes committed, 177
taxicabs and crime rates, 26
violent crime categories and rates, 596

Dental applications

fluoride and tooth decay, 33

Earth science applications

altitude and pollution from cars, 156–157, 535, 568
arctic ice sheet, 83
floodwater drainage time, 676
geyser eruption durations, 68, 651
global warming, 18
humidity and ozone levels, 629, 635
hurricanes, 568, 713–714
ice on lakes, 656–657
silver ore in rock samples, 70
wind speeds and electric power generated, 659

Education/school applications

absences from school, 610
adults with degrees, 89, 115
advanced placement tests, 298
age and education, 115, 177
age at high school graduation, 177
ages of college students, 404, 424, 457
air pollution in Chinese schools, 26
annual incomes of college graduates, 384
arithmetic teaching method, 477
attitudes toward school, 384
cafeteria food rating, 299
choosing books to study, 255
college enrollments, 81, 288, 355
college freshmen majoring in sciences, 348
college graduates working at home, 414
college tuition, 425, 474, 535
computers in the classroom, 531
confidence in educational institutions, 224
course selection, 259
distance learning effectiveness, 515
educational issues and visits to science museums, 605
education levels, 49, 401, 609
energy drinks consumed by college students, 515
exam scores, 69, 139, 156, 333
exam score *versus* hours spent studying, 188–189
favorite subjects, 688
female business majors, 259
final exam scores, 116, 156, 189, 333
final grade probabilities, 234, 259
final grades, 116
gender bias in graduate school admissions, 610–612
gender differences in attitudes and behaviors, 580
girls enrolled in elementary school, 369
grade distributions, 595–596
grade point averages, 69, 318–319, 384, 456, 649–650
graduate salaries *versus* years of experience, 189
heights of college students, 66, 117
high school dropouts, 247
high school freshmen graduating with their class, 347–348
homework not completed by college freshmen, 92
homework score, 116
homework submitted electronically, 33
hours of sleep, 403, 428
hours spent studying or reading, 188–189, 333, 404
illegal drug use in high school, 30
IQ test scores, 421, 438, 491, 515, 549, 576
male *versus* female college students, 81
math skills testing, 428, 455, 713
multiple choice questions, 30, 277, 288
number of siblings, 69
online learning courses, 401
ordering of supplies, 32
parking on campus, 11
phonics, 401
quiz scores, 93
reading skills, 11, 299, 365, 415, 569
recreation fee increase, 12
required courses, 260
satisfaction with college life, 487–488
SAT math preparation, 477, 524, 559
SAT math scores, 116, 189, 341, 383, 415, 421, 456, 491, 498
SAT score and language study, 207
SAT verbal scores, 189
saving for college, 494
school enrollment by grade, 277
sick children in class, 234
software instruction hours, 401
standardized test scores, 154
student executive committee choices, 255
student loans, 347
teacher salaries, 494
tenth-graders planning to attend college, 417
tests online *versus* on paper, 516, 550
text messages sent by teenagers, 276
true–false exam questions, 224, 287, 596
tuition and fees of private colleges, 474

vocabulary and height of elementary students, 26
weight gain or loss during freshman year, 715
working college students, 487
years of education for adults, 401

Elderly/older-person applications

prescription drug program, 11

Electronics/computer applications

ages of video gamers, 48
analyze digital music, 569
battery lifetimes, 318, 342, 368
bias in phone polls, 30
Blu-ray load times, 708
changing jobs, 416
computer chips, 160–161, 570
computer crashes, 514–515
computer data backup, 48–49
computer passwords, 256
computer purchases, 234
computer virus/worm, 11, 234
defective circuits, 259, 276
defective electronic components, 428, 532
delivery of products from e-store, 30
drone delivery of packages, 384
Facebook usage, 110, 369, 501
flickering screen on DVR, 295
freeze-up times for computers, 576
Google search engine usage, 288, 354–355
HDTV screen sizes, 402
homework submitted electronically, 33
hours spent on internet, 548
households with computers, 383
Instagram followers, 49
internet provider cost, 137
internet radio subscribers, 68–69
internet service interruptions, 414–415
interpreting calculator display, 188–189, 207–208, 384, 403–404, 416, 456–457, 475–476, 487, 502, 517, 524, 532, 535–537, 550, 559, 569, 583–584, 585, 655–656
interpreting computer output, 384, 385, 404, 416–417, 457, 476, 487, 501, 517, 524–525, 533, 536, 537, 550, 559–560, 569–570, 584–585, 656
microprocessor speed, 522
mobile (smartphone) apps, 111, 347
Netflix usage, 455
phone batteries, 161
phone features, 12
phone ownership by children, 415, 486
phones in purchasing decisions, 288
PIN codes, 248
posting photos online, 486
random-digit dialing as sampling method, 30
semiconductor wafer quality control, 246
smartphone ownership, 347, 414
smartphone prices, 383
smartphone sales, 49
social media usage, 475–476, 532
spam in email, 486
text messages, 276, 457, 476, 549
time spent playing video games, 67
top ten video games, 18
Twitter usage, 49, 295, 401, 486
video game consoles, 486
video games sales, 18, 46, 47, 48
website ratings, 32

Entertainment applications. *See also* Gambling applications

Avengers: Endgame, 205
ages of Grammy Award winners, 340–341
ages of video gamers, 48
Blu-ray load times, 708
Broadway shows, 115, 202
cable news viewing, 502
digital music sales, 49, 87, 88, 94
Dungeons & Dragons, 225
favorite newscaster, 33
favorite performer, 32
favorite television show, 111, 111–112, 135
federal support for arts and culture, 160, 175, 187
football stadium seating, 156
Halloween costumes, 50, 403
internet radio subscribers, 68–69
League of Legends game critical strikes, 415
movie running times, 80, 401, 702
pandora and spotify subscribers, 94
Pokémon characters, 156
random playing of songs, 246, 259
recreation fee increase, 12
roller coaster weight carried safely, 341
SpongeBob picture recognition, 223, 247
Squidward Tentacles picture recognition, 247
Super Bowl on television, 47, 428
target practice, 247
television sets in households, 341, 501
television viewing habits, 11, 478
tic-tac-toe, 247
time spent playing video games, 67
time spent watching television, 474, 475–476, 491, 583
top-grossing movies, 18, 80
top ten video games, 18
video game consoles, 486
video game sales, 18, 46, 47, 48

Environmental applications

air pollution and colds, 26
air pollution and respiratory health, 34–35
air pollution caused by motor vehicles, 79, 535, 568
air pollution in Chinese schools, 26
air pollution standards, 276
benzene in wastewater, 514, 549, 583
carbon monoxide concentration, 404
construction site water contamination, 415
contaminated wells, 333
environmental restoration cost, 427
fine particle air pollution, 157
floodwater drainage time, 676
fracking, 648–649
fuel tank leaks, 428
global warming, 18
hazardous waste sites, 154–155
humidity and ozone levels, 629, 635
ice on lakes, 656–657
ice sheet coverage in Arctic, 83
interest in environmental issues and science museum visits, 605
lead in water, 415, 456
mercury in lakes, 494
microorganism diversity, 717–718
ozone in atmosphere, 189, 629, 635
particulate matter air pollution, 34–35, 79, 656
PCB contamination in water, 427
power plant emission levels, 676, 694
secondhand smoke, 26
shrimp habitat and water pollution, 535
vehicle emissions at low and high altitudes, 156–157, 535, 568
wastewater treatment, 717–718
wood stove pollution, 429–430

Exercise applications. *See* Sports/exercise/fitness applications

Farming applications. *See* Agricultural/gardening/farming applications

Fitness applications. *See* Sports/exercise/fitness applications

Food applications. *See also* Beverage applications

cafeteria food rating, 299
calories in bread, 425
calories in eggs, 535
calories in fat, 629
calories in hamburgers, 110
chocolate chips in each cookie, 295
eat healthy, 115
eggs and milk prices, 175, 187
family food expenditures, 49
fast-food chain outlets, 246
food spending by U.S. residents, 90
free desserts to increase business, 475
grams in boxes of crackers, 502
hot dog eating contest, 88
ice cream flavors, 255
mineral content of kale, 403
mineral content of spinach, 402, 422
ounces of cereal in boxes, 342, 402, 422
ounces of cookies in box, 494
pizza toppings, 255
pretzels in advertising campaign, 531–532
protein and calories in fast-food products, 627, 634
restaurant food rating, 30
restaurant outlets by city population and location, 246
restaurant spending, 437
spending on lunch, 136, 137
sugar content of selected products, 501, 576
sugar in apples, 425
vegetables and colon cancer, 26
weight of bag of apples, 342

Forestry applications

breaking strength of wood boards, 676
cost of restoration of forests, 427
flaws in lumber, 295
tree heights, 332, 333, 703
trees of certain species in forest, 295
volume of lumber in trees, 206, 654–655

Gambling applications. *See also* Entertainment applications

blackjack, 256
coin tossing, 223, 224, 225, 355, 416, 438, 488
craps, 277, 299
dice rolling, 224, 225, 233, 259, 596–597
horse racing, 255
lottery profits, 277
lottery tickets sold, 299
lottery winnings, 116–117, 247, 256
Mega Millions lottery, 595, 610
roulette, 223, 224
Texas hold 'em, 256
tribal casinos, 155–156

Gardening applications. *See* Agricultural/gardening/farming applications

Genetics/gender applications

birth order and intelligence, 515, 549, 576
DNA sequences, 256, 606
dominant/recessive genes, 246
gender bias in graduate school admissions, 610–612
gender differences in attitudes and behaviors, 580
gender of business majors, 259
gender of senators, 89, 417
heights of fathers and sons, 627, 634
high school dropouts by gender, 247
male *versus* female college students, 81
management jobs and gender, 247
maternal age at time of first birth, 487
mean weights of men, 474
mutant genes and age, 201
mutation causing deafness, 299
pea plants genetic studies, 355
political party affiliation and gender, 234–235, 246
service occupation jobs, 247
sickle-cell anemia carriers, 299

Health/health-care applications. *See also* Medical/medical research/alternative medicine applications

absorption rate of pesticides into skin, 675
air pollution and colds, 26
air pollution and respiratory health, 34–35
alcohol and liver disease, 27
anterior cruciate ligament (ACL) injury, 487
beryllium disease, 605
blood pressure by age, 157
blood pressure in men/women, 66, 140, 154, 331, 523
blood sugar, 487
body mass index of adults with diabetes, 515
cancer and life expectancy, 94–95
carbon monoxide exposure and symptoms reported, 605
childhood obesity, 568
cholesterol levels, 66, 155, 341, 383, 568, 629, 635
cold medications, 26
days of coughing after acute cough illness, 69
days of pain suffered before seeking medical treatment, 276–277
deaths due to smoking, 225
diet effectiveness, 476
drug to lower blood pressure, 477
exercise and blood pressure, 25–26, 716
exercise and heart rate, 535
flu types, 113
foot ulcers and foot temperatures, 176–177, 188
headache drug testing, 12
health-care proposal opinions, 33
heart attacks by age and gender, 260
heart disease study, 88, 515
high blood pressure probabilities, 289, 348, 355
hours of relaxation per day, 276
hours of sleep, 403, 428, 476
lead in water, 415, 456
life expectancy, 33, 110
low-fat *versus* low-carb diets, 515, 548
lung function measurement, 456
mean weights of men, 474
nicotine patch to quit smoking, 486, 569
pneumonia patients under age 18, 425
prescription drug program, 11
quit smoking, 415
recovering from heart attacks in smaller houses, 26
secondhand smoke, 26
sleep apnea, 414
smoking and blood pressure, 247
soft drinks and smoking, 532
stress at work, 289, 355
systolic/diastolic blood pressure measurements, 66, 129, 177, 188, 628, 634
trust in doctors, 92
waist size of men, 428
water loss, 523
weight and cholesterol, 234
weight loss diets and programs, 12, 384, 477, 476, 514, 515, 548, 702

Home improvement. *See* Construction/home improvement/home purchases and sales applications

Home purchases and sales applications. *See* Construction/home improvement/home purchases and sales applications

Library/book applications

age and visits to library, 234
age of college library books, 456
best graphic novels, 18
books arranged on bookshelf, 260
choosing books to study, 255
favorite types of books, 115
funny sounding and interesting words, 18
library hours changes, 12
magazine subscription cost, 114
newspaper circulation, 113
reviewing articles for a journal, 136
thickness of pages in a book, 342

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

Manufacturing applications

ball bearing diameters, 333
batteries for pacemaker, 438
battery lifetimes, 342
breaking strength of bolts, 585
breaking strength of wires, 438
calibration of a scale, 438, 455–456
ceramic tiles, 349
coating for gears, 580
component design and materials, 690
component repair costs, 427, 428
defective products, 247, 276, 289, 365, 369, 428, 532, 580
delivery truck load weights, 369
drill lifetimes, 427
expansion of concrete over time, 628, 635
flaws in aluminum foil, 295
flaws in aluminum parts, 234
impurities in aluminum cans, 365
ladies' shoe sizes, 494
lightbulb lifetimes, 368
microbalance performance, 696
paint manufacture, 516
product rating, 606
quality control, 12, 246, 609, 695
rivet length, 161
solar power collectors, 675
shipping speed and quality, 690
steel rod lengths/diameters, 583
strength of concrete, 416, 522–523, 558, 689, 708
thread strength, 576
washer thicknesses, 438
weights obtained on different scales, 535, 558
wood stove pollution, 429–430

Medical/medical research/alternative medicine applications. *See also* Health/health-care applications

alcohol and liver disease, 27
anterior cruciate ligament (ACL) injury, 487
angioplasty to treat circulatory disease, 532, 568
annual earnings of doctors, 474
antifungal drug testing, 475, 523, 558
artificial hip ball wear, 675–676
asthma, 247
beryllium disease, 605
blood sugar, 487
blood typing system for humans, 46, 259, 288
breathing rates, 523
body mass index of adults with diabetes, 515
brain cancer and electrical field exposure, 586
cancer and life expectancy, 94
cholesterol lowering drugs, 383, 498, 523
choosing doctors, 486
cold medications, 26
colonoscopy effectiveness, 532
coronary bypass surgery, 288, 716
crossover trials to compare drugs, 557

days of coughing after acute cough illness, 69
disease test results, 248
dosages administered by syringe, 368
Down Syndrome births, 296
drug concentration in bloodstream, 365
drug half-life, 491
drugs to prevent heart attacks or strokes, 568
drug to lower blood pressure, 477
flu types, 113
foot ulcers and foot temperatures, 176–177, 188
headache drug testing, 12
health of former *versus* current smokers, 27
heart attacks by age and gender, 260
heart disease, 88, 515–516
hip surgery times, 401
hospital admissions by reason, 48
hospital stay costs, 202
hospital stay lengths, 476
hospitals winning safety awards, 80
hospital visits related to asthma, 225
kidney transplants, 348–349
knee surgery complications, 415
life expectancy, 94, 110
lung function measurement, 456
mean weights of men, 474
nicotine patch to quit smoking, 486, 569
outcomes of surgery in hospitals, 609
pain reliever effectiveness, 25, 402, 532
pain reliever half-life, 402, 491
pneumonia patients under age 18, 425
polio vaccine trials, 27
postoperative pain, 531
postoperative treatment for surgical patients, 477
postsurgical recovery times of patients, 535, 549, 708, 716
prescription drug program, 11
quit smoking, 415
radon and cancer, 203
recovering from heart attacks in smaller houses, 26
reducing stomach volume to cure diabetes, 486
Salk vaccine trials, 27
sickle-cell anemia carriers, 299
side effects of medical procedure, 369
sleep apnea, 414
soft drinks and smoking, 532
sugar content of pharmaceutical products, 675
surgical procedure time, 161
treadmill walking to alleviate claudication, 516, 549, 583
treatments to prevent heart failure, 537
trust in doctors, 92
vegetables and colon cancer, 26
water loss, 523
women heart disease study, 525

Miscellaneous applications

age at death, 69
ages at death for British monarchs, 93
committee officer choices, 256
confidence intervals/levels in sampling, 383–384
customer service survey, 11
customers in express checkout line, 275–276
dogs in households, 703
Empirical Rule, 137–138
false fire alarms, 596
family size, 605
first digits of numbers, 300–302
fraud complaint rates, 703
household income, 110, 157
household income and energy consumed, 206, 655
job interview probabilities, 246
languages spoken in households, 50
large samples and practical significance, 477
laundry costs, 708
marketing firm survey, 19
mean absolute deviation, 139
newspaper circulation, 113
number of people in households, 384, 476, 501
personal incomes, 115, 116, 157, 342
planet distances from sun, 111, 132
probabilities of events, 223
proportion of females in U.S. population, 50
raffle tickets, 11
refrigerator costs, 475
residents of geographic regions of U.S., 596
retail spending, 11, 79
reviewing articles for a journal, 136
shopping at the mall, 11
somebody's wrong, 477
sources of current event news, 115
spending on gifts, 115
spending on lunch, 136, 137
spreadsheet task completion time, 365
standard deviation, 138
state populations, 300–302, 403
tattoos among adults, 486
testing enough information, 477
timing of bills payments, 137, 224
U.S. population by geographic region, 47, 596
volunteer work, 415

Motor vehicle applications. *See* Automotive/motor vehicle applications

Nuclear applications

nuclear power plant construction, 30
nuclear reactors in selected countries, 155

Political applications

abortion policy, 596
ages of U.S. presidents and their wives, 201
ages of U.S. presidents at death, 93
approval ratings for Congress, 703
economic future polls, 416, 596
election reform, 30
electoral votes cast, 83
female senators, 89, 417
freshmen in House of Representatives, 93
government spending, 46–47, 160, 175–176, 187
handling of economy by U.S. Congress, 30
health-care proposal opinions, 33
heights of U.S. presidents, 403
immigration policy, 11
Literary Digest political polling, 30–31
mayoral elections, 223–224
military spending, 81
number of words in inaugural speeches, 68, 156
order of choices in political polling, 30
party affiliation and gender of senators, 234–235, 246
political debates, 488
satisfaction with presidential candidates, 416
smoking in public places, 605
voter preferences on candidates, 12, 32, 223–224, 369, 416, 488
voter preferences on selected issues, 355, 535, 583
votes for incumbant, 416
women's effectiveness at governing, 417

Psychological applications. *See also* Behavioral study applications

attitudes toward school, 384
choosing doctors, 486–487
empathy types, 549
IQ test scores, 421, 438, 491, 515, 549, 576
optimism about the future, 606–607
reaction times to visual and auditory stimuli, 628, 655
weight loss during counseling, 474

Safety applications

car accidents at intersections, 531
car accidents per year, 117, 296
car speed and stopping distance, 202–203
delivery truck load weights, 369
don't drink and drive, 11
elevator weight carried safely, 341
hospitals winning safety awards, 80
pillars giving head protection in cars, 677
roller coaster weight carried safely, 341
seat belt effectiveness, 30
seat belt use, 416
years between car accidents, 516, 550

School applications. *See* Education/school applications

Sociological applications

age and education, 115, 177
age distribution, 114, 136
ages of tennis and golf champions, 80
ages of video gamers, 48
ages when women first marry, 89
birth rates, 82, 95
birth rates after hurricane, 713–714
changing jobs, 416
gender differences in attitudes and behaviors, 580
languages spoken in households, 50
number of children, 157, 224–225, 403, 605
number of people in households, 384, 476, 501
number of siblings, 69, 132, 605
phone ownership by children, 415
poverty rates, 92
smoking in public places, 605
world population, 48

Sports/exercise/fitness applications

ages of tennis and golf champions, 80
baseball batting averages, 67
baseball pitches, 224
baseball runs scored, 500–501, 716
baseball salaries, 156
basketball three-point shots, 82
bowling scores, 116, 247
breathing rates, 523
exercise and blood pressure, 25–26, 716
exercise and heart rate, 535
exercise and weight, 713
football points by teams, 133
football stadium seating, 156
football turnover margins and wins, 206
heights of football players, 176, 187
hockey stick breaking strength, 576
margin of victory in Olympic swimming events, 110
Olympic athletes/participation by country, 66, 82
Olympic gold medal winners, 81–82
soccer goals scored, 277
strike three, 89
Super Bowl on television, 47, 428
tennis ball circumference, 161
treadmill walking to alleviate claudication, 516, 549, 583
weight-loss programs, 514
weights of football players, 113, 134, 176, 187, 424
weights of soccer players, 403

Travel applications

bus route changes, 11
car accidents at intersections, 531
car accidents per year, 117, 296
carpool occupants, 276
car speed and street noise level, 628, 634
car speeds on highways, 333, 494
commuting to work in miles/times, 70, 155, 207, 333, 342, 474, 476, 703, 716
fuel efficiency, 33, 133, 524
gas prices, 112
phones and driving, 33
potholes in street, 295
risky drivers, 224
tire lifetimes, 332, 498
tires and fuel economy, 33, 524
traffic during morning *versus* evening rush hours, 713
traffic lights, 233, 247
years between car accidents, 516, 550

Weather applications

annual precipitation, 111
daily high temperatures, 67, 79
days of rain, 299, 428
humidity and ozone levels, 629, 635
mean temperature, 79, 138, 200
monthly rainfall, 158
rain probability, 260
snowfall amounts, 82, 569
summer temperatures, 340
temperature and barometric pressure, 201
temperature warming trends, 503–504
wind speeds, 155, 425

chapter

1

Basic Ideas

Tatiana Grozetskaya/Shutterstock

Introduction

How does air pollution affect your health? Over the past several decades, scientists have become increasingly convinced that air pollution is a serious health hazard. The World Health Organization has estimated that air pollution causes 2.4 million deaths each year. The health effects of air pollution have been investigated by measuring air pollution levels and rates of disease, then using statistical methods to determine whether higher levels of pollution lead to higher rates of disease.

Many air pollution studies have been conducted in the United States. For example, the town of Libby, Montana, was the focus of a recent study of the effect of particulate matter—air pollution that consists of microscopic particles—on the respiratory health of children. As part of this study, parents were asked to fill out a questionnaire about their children's respiratory symptoms. It turned out that children exposed to higher levels of particulate pollution were more likely to exhibit symptoms of wheezing, as shown in the following table.

Level of Exposure	Percentage with Symptoms
High	8.89%
Low	4.56%

The rate of symptoms was almost twice as high among those exposed to higher levels of pollution. At first, it might seem easy to conclude that higher levels of pollution cause

symptoms of wheezing. However, drawing accurate conclusions from information like this is rarely that simple. The case study at the end of this chapter will present more complete information and will show that additional factors must be considered.

Section 1.1 Sampling

Objectives

1. Describe the investigative process of statistics
2. Construct a simple random sample
3. Determine when samples of convenience are acceptable
4. Describe stratified sampling, cluster sampling, systematic sampling, and voluntary response sampling
5. Distinguish between statistics and parameters

Objective 1 Describe the investigative process of statistics

In the months leading up to an election, polls often tell us the percentages of voters that prefer each of the candidates. How do pollsters obtain this information? The ideal poll would be one in which every registered voter were asked his or her opinion. Of course, it is impossible to conduct such an ideal poll, because it is impossible to contact every voter. Instead, pollsters contact a relatively small number of voters, usually no more than a couple of thousand, and use the information from these voters to predict the preferences of the entire group of voters.

The process of polling requires two major steps. First, the voters to be polled must be selected and interviewed. In this way the pollsters collect information. In the second step, the pollsters analyze the information to make predictions about the upcoming election. Both the collection and the analysis of the information must be done properly for the results to be reliable. The field of statistics provides appropriate methods for the collection, description, and analysis of information.

DEFINITION

Statistics is the study of procedures for collecting, describing, and drawing conclusions from information.

Polling follows a process that is typical in statistics. We formulate questions (Which candidate is most likely to win?), then collect and analyze data to address the questions. In general, statistics is an investigative process that involves the following four steps:

- Formulate questions.
- Collect data needed to answer the questions.
- Describe the data.
- Draw conclusions, using appropriate methods.

In this section, we will focus on the second step in the process—the collection of data. The polling problem is typical of a data collection problem. We want some information about a large group of individuals, but we are able to collect information on only a small part of that group. In statistical terminology, the large group is called a *population*, and the part of the group on which we collect information is called a *sample*.

EXPLAIN IT AGAIN

Why do we draw samples?
It's usually impossible to examine every member of a large population. So we select a group of a manageable size to examine. This group is called a sample.

DEFINITION

- A **population** is the entire collection of individuals about which information is sought.
- A **sample** is a subset of a population, containing the individuals that are actually observed.

Ideally, we would like our sample to represent the population as closely as possible. For example, in a political poll, we would like the proportions of voters preferring each of the candidates to be the same in the sample as in the population. Unfortunately, there are no methods that can guarantee that a sample will represent the population well. The best we can do is to use a method that makes it very likely that the sample will be similar to the population. The best sampling methods all involve some kind of random selection. The most basic, and in many cases the best, sampling method is the method of **simple random sampling**.

Simple Random Sampling

Objective 2 Construct a simple random sample

To understand the nature of a simple random sample, think of a lottery. Imagine that 10,000 lottery tickets have been sold, and that 5 winners are to be chosen. What is the fairest way to choose the winners? The fairest way is to put the 10,000 tickets in a drum, mix them thoroughly, then reach in and draw 5 tickets out one by one. These 5 winning tickets are a simple random sample from the population of 10,000 lottery tickets. Each ticket is equally likely to be one of the 5 tickets drawn. More importantly, each collection of 5 tickets that can be formed from the 10,000 is equally likely to comprise the group of 5 that is drawn.

DEFINITION

A **simple random sample** of size n is a sample chosen by a method in which each collection of n population items is equally likely to make up the sample, just as in a lottery.

Since a simple random sample is analogous to a lottery, it can often be drawn by the same method now used in many lotteries: with a computer random number generator. Suppose there are N items in the population. We number the items 1 through N. Then we generate a list of random integers between 1 and N and choose the corresponding population items to comprise the simple random sample.

Example 1.1

Choosing a simple random sample

There are 300 employees in a certain company. The Human Resources department wants to draw a simple random sample of 20 employees to fill out a questionnaire about their attitudes toward their jobs. Describe how technology can be used to draw this sample.

Solution

Step 1: Make a list of all 300 employees, and number them from 1 to 300.

Step 2: Use a random number generator on a computer or a calculator to generate 20 random numbers between 1 and 300. The employees who correspond to these numbers comprise the sample.

Example 1.2

Determining whether a sample is a simple random sample

A physical education professor wants to study the physical fitness levels of students at her university. There are 20,000 students enrolled at the university, and she wants to draw a sample of size 100 to take a physical fitness test. She obtains a list of all 20,000 students, numbered from 1 to 20,000. She uses a computer random number generator to generate 100 random integers between 1 and 20,000, then invites the 100 students corresponding to those numbers to participate in the study. Is this a simple random sample?

Solution

Yes, this is a simple random sample because any group of 100 students would have been equally likely to have been chosen.

Example 1.3

Determining whether a sample is a simple random sample

The professor in Example 1.2 now wants to draw a sample of 50 students to fill out a questionnaire about which sports they play. The professor's 10:00 A.M. class has 50 students. She uses the first 20 minutes of class to have the students fill out the questionnaire. Is this a simple random sample?

Solution

No. A simple random sample is like a lottery, in which each student in the population has an equal chance to be part of the sample. In this case, only the students in a particular class had a chance to be in the sample.

Example 1.4

In a simple random sample, all samples are equally likely

To play the Colorado Lottery Lotto game, you must select six numbers from 1 to 42. Then lottery officials draw a simple random sample of six numbers from 1 to 42. If your six numbers match the ones in the simple random sample, you win the jackpot. Sally plays the lottery and chooses the numbers 1, 2, 3, 4, 5, 6. Her friend George says that this isn't a good choice, since it is very unlikely that a random sample will turn up the first six numbers. Is he right?

Solution

No. It is true that the combination 1, 2, 3, 4, 5, 6 is unlikely, but every other combination is equally unlikely. In a simple random sample of size 6, every collection of six numbers is equally likely (or equally unlikely) to come up. So Sally has the same chance as anyone to win the jackpot.

Example 1.5

Using technology to draw a simple random sample

Use technology to draw a simple random sample of five employees from the following list.

1. Dan Aaron	11. Johnny Gaines	21. Jorge Ibarra	31. Edward Shingleton
2. Annie Bienh	12. Carlos Garcia	22. Maurice Jones	32. Michael Speciale
3. Oscar Bolivar	13. Julio Gonzalez	23. Jared Kerns	33. Andrew Steele
4. Dominique Bonnaud	14. Jacqueline Gordon	24. Kevin King	34. Neil Swain
5. Paul Campbell	15. James Graves	25. Frank Lipka	35. Sherry Thomas
6. Jeffrey Carnahan	16. Ronald Harrison	26. Carl Luther	36. Shequiea Thompson
7. Joel Chae	17. Andrew Huang	27. Laverne Mitchell	37. Barbara Tilford
8. Dustin Chen	18. Anthony Hunter	28. Zachary Quesada	38. Jermaine Tryon
9. Steven Coleman	19. Jonathan Jackson	29. Donnell Romaine	39. Lizbet Valdez
10. Richard Davis	20. Bruce Johnson	30. Gary Sanders	40. Katelyn Yu

Solution

We will use the TI-84 Plus graphing calculator. The step-by-step procedure is presented in the Using Technology section on page 9. We begin by choosing a **seed**, which is a number that the calculator uses to get the random number generator started. Display (a) shows the seed being set to 21. (The seed can be chosen in almost any way; this number was chosen by looking at the seconds display on a digital watch.) Display (b) presents the five numbers in the sample.

(a) (b)

The simple random sample consists of the employees with numbers 27, 39, 30, 35, and 17. These are Laverne Mitchell, Lizbet Valdez, Gary Sanders, Sherry Thomas, and Andrew Huang.

CAUTION
If you use a different type of calculator, a different statistical package, or a different seed, you will get a different random sample. This is perfectly all right. So long as the sample is drawn by using a correct procedure, it is a valid random sample.

Check Your Understanding

1. A pollster wants to estimate the proportion of voters in a certain town who are Democrats. He goes to a large shopping mall and approaches people to ask whether they are Democrats. Is this a simple random sample? Explain.
2. A telephone company wants to estimate the proportion of customers who are satisfied with their service. They use a computer to generate a list of random phone numbers and call those people to ask them whether they are satisfied. Is this a simple random sample? Explain.

Answers are on page 12.

Objective 3 Determine when samples of convenience are acceptable

Samples of Convenience

In some cases, it is difficult or impossible to draw a sample in a truly random way. In these cases, the best one can do is to sample items by some convenient method. A sample obtained in such a way is called a *sample of convenience*.

DEFINITION

A **sample of convenience** is a sample that is not drawn by a well-defined random method.

Example 1.6

Creatas Images/Jupiterimages

Drawing a sample of convenience

A construction engineer has just received a shipment of 1000 concrete blocks, each weighing approximately 50 pounds. The blocks have been delivered in a large pile. The engineer wishes to investigate the crushing strength of the blocks by measuring the strengths in a sample of 10 blocks. Explain why it might be difficult to draw a simple random sample of blocks. Describe how the engineer might draw a sample of convenience.

Solution

To draw a simple random sample would require removing blocks from the center and bottom of the pile, which might be quite difficult. One way to draw a sample of convenience would be to simply take 10 blocks off the top of the pile.

Problems with samples of convenience

The big problem with samples of convenience is that they may differ systematically in some way from the population. For this reason, samples of convenience should not be used, except in some situations where it is not feasible to draw a random sample. When it is necessary to

CAUTION
Don't use a sample of convenience when it is possible to draw a simple random sample.

draw a sample of convenience, it is important to think carefully about all the ways in which the sample might differ systematically from the population. If it is reasonable to believe that no important systematic difference exists, then it may be acceptable to treat the sample of convenience as if it were a simple random sample. With regard to the concrete blocks, if the engineer is confident that the blocks on the top of the pile do not differ systematically in any important way from the rest, then he can treat the sample of convenience as a simple random sample. If, however, it is possible that blocks in different parts of the pile may have been made from different batches of mix, or may have different curing times or temperatures, a sample of convenience could give misleading results.

SUMMARY

- A sample of convenience may be acceptable when it is reasonable to believe that there is no systematic difference between the sample and the population.
- A sample of convenience is not acceptable when it is possible that there is a systematic difference between the sample and the population.

Objective 4 Describe stratified sampling, cluster sampling, systematic sampling, and voluntary response sampling

Some Other Sampling Methods

Stratified sampling

In **stratified sampling**, the population is divided into groups, called **strata**, where the members of each stratum are similar in some way. Then a simple random sample is drawn from each stratum. Stratified sampling is useful when the strata differ from one another, but the individuals within a stratum tend to be alike.

Example 1.7

Drawing a stratified sample

A company has 1000 employees, of whom 800 are full-time and 200 are part-time. The company wants to survey 50 employees about their opinions regarding benefits. Attitudes toward benefits may differ considerably between full-time and part-time employees. Why might it be a good idea to draw a stratified sample? Describe how one might be drawn.

Solution

If a simple random sample is drawn from the entire population of 1000 employees, it is possible that the sample will contain only a few part-time employees, and their attitudes will not be well represented. For this reason, it might be advantageous to draw a stratified sample. To draw a stratified sample, one would use two strata. One stratum would consist of the full-time employees, and the other would consist of the part-time employees. A simple random sample would be drawn from the full-time employees, and another simple random sample would be drawn from the part-time employees. This method guarantees that both full-time and part-time employees will be well represented.

EXPLAIN IT AGAIN
Example of a cluster sample: Imagine drawing a simple random sample of households and interviewing every member of each household. This would be a cluster sample, with the households as the clusters.

Cluster sampling

In **cluster sampling**, items are drawn from the population in groups, or clusters. Cluster sampling is useful when the population is too large and spread out for simple random sampling to be feasible. Cluster sampling is used extensively by U.S. government agencies in sampling the U.S. population to measure sociological factors such as income and unemployment.

Example 1.8

Drawing a cluster sample

To estimate the unemployment rate in a county, a government agency draws a simple random sample of households in the county. Someone visits each household and asks how

many adults live in the household and how many of them are unemployed. What are the clusters? Why is this a cluster sample?

Solution

The clusters are the groups of adults in each of the households in the county. This is a cluster sample because a simple random sample of clusters is selected, and every individual in each selected cluster is part of the sample.

EXPLAIN IT AGAIN

The difference between cluster sampling and stratified sampling: In both cluster sampling and stratified sampling, the population is divided into groups. In stratified sampling, a simple random sample is chosen from each group. In cluster sampling, a random sample of groups is chosen, and every member of the chosen groups is sampled.

Systematic sampling

Imagine walking alongside a line of people and choosing every third one. That would produce a **systematic sample**. In a systematic sample, the population items are ordered. It is decided how frequently to sample items; for example, one could sample every third item, or every fifth item, or every hundredth item. Let k represent the sampling frequency. To begin the sampling, choose a starting place at random. Select the item in the starting place, along with every kth item after that.

Systematic sampling is sometimes used to sample products as they come off an assembly line, in order to check that they meet quality standards.

Example 1.9

Describe a systematic sample

Automobiles are coming off an assembly line. It is decided to draw a systematic sample for a detailed check of the steering system. The starting point will be the third car, then every fifth car after that will be sampled. Which cars will be sampled?

Solution

We start with the third car, then count by fives to determine which cars will be sampled. The sample will consist of cars numbered 3, 8, 13, 18, and so on.

Digital Vision/Punchstock

Voluntary response sampling

Voluntary response samples are often used by the media to try to engage the audience. For example, a news commentator will invite people to tweet an opinion, or a radio announcer will invite people to call the station to say what they think. How reliable are voluntary response samples? To put it simply, *voluntary response samples are never reliable*. People who go to the trouble to volunteer an opinion tend to have stronger opinions than is typical of the population. In addition, people with negative opinions are often more likely to volunteer their responses than those with positive opinions.

Figures 1.1–1.4 illustrate several valid methods of sampling.

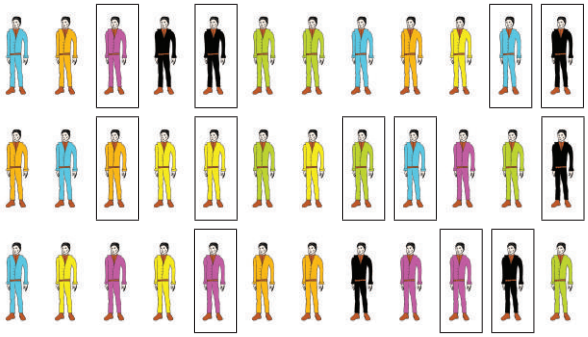


Figure 1.1 Simple random sampling

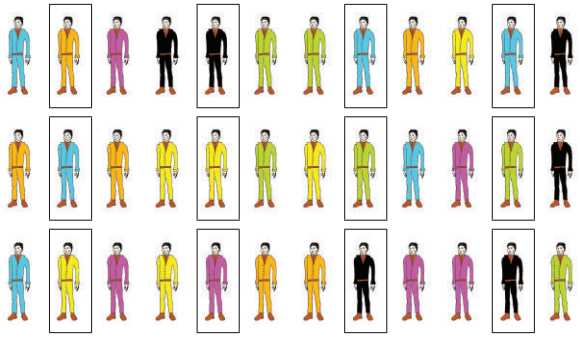


Figure 1.2 Systematic sampling

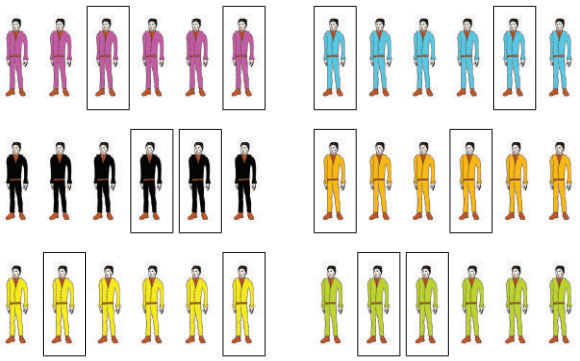


Figure 1.3 Stratified sampling

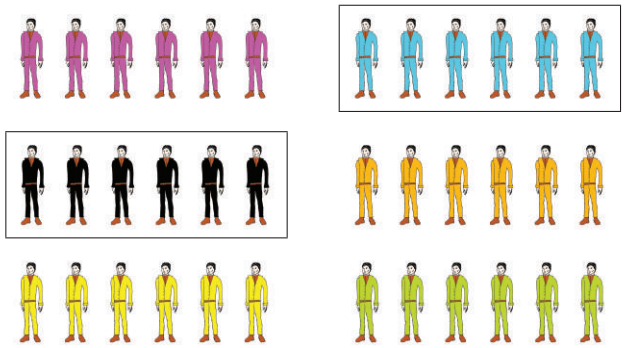


Figure 1.4 Cluster sampling

Check Your Understanding

3. A radio talk show host invites listeners to send an email to express their opinions on an upcoming election. More than 10,000 emails are received. What kind of sample is this?
4. Every 10 years, the U.S. Census Bureau attempts to count every person living in the United States. To check the accuracy of their count in a certain city, they draw a sample of census districts (roughly equivalent to a city block) and recount everyone in the sampled districts. What kind of sample is formed by the people who are recounted?
5. A public health researcher is designing a study of the effect of diet on heart disease. The researcher knows that the diets of men and women tend to differ and that men are more susceptible to heart disease. To be sure that both men and women are well represented, the study comprises a simple random sample of 100 men and another simple random sample of 100 women. What kind of sample do these 200 people represent?
6. A college basketball team held a promotion at one of its games in which every twentieth person who entered the arena won a free basketball. What kind of sample do the winners represent?

Answers are on page 12.

Simple random sampling is the most basic method

Simple random sampling is not the only valid method of random sampling. But it is the most basic, and we will focus most of our attention on this method. From now on, unless otherwise stated, the terms *sample* and *random sample* will be taken to mean *simple random sample*.

Objective 5 Distinguish between statistics and parameters

Statistics and Parameters

We often use numbers to describe, or summarize, a sample or a population. For example, suppose that a pollster draws a sample of 500 likely voters in an upcoming election, and 68% of them say that the state of the economy is the most important issue for them. The quantity "68%" describes the sample. A number that describes a sample is called a *statistic*.

DEFINITION

A **statistic** is a number that describes a sample.

EXPLAIN IT AGAIN

Statistic and parameter: An easy way to remember these terms is that "statistic" and "sample" both begin with "s," and "parameter" and "population" both begin with "p."

Now imagine that the election takes place and that one of the items on the ballot is a proposition to raise the sales tax to pay for the development of a new park downtown. Let's say that 53% of the voters vote in favor of the proposition. The quantity "53%" describes the population of voters who voted in the election. A number that describes a population is called a *parameter*.

DEFINITION

A **parameter** is a number that describes a population.

Example 1.10

Distinguishing between a statistic and a parameter

Which of the following is a statistic and which is a parameter?

a. 57% of the teachers at Central High School are female.

b. In a sample of 100 surgery patients who were given a new pain reliever, 78% of them reported significant pain relief.

Solution

a. The number 57% is a parameter, because it describes the entire population of teachers in the school.

b. The number 78% is a statistic, because it describes a sample.

Using Technology

We use Example 1.5 to illustrate the technology steps.

TI-84 PLUS

Drawing a simple random sample

Step 1. Enter any nonzero number on the HOME screen as the seed.

Step 2. Press **STO** >.

Step 3. Press **MATH**, select **PRB**, then **1: rand**, and then press **ENTER**. This enters the seed into the calculator memory. See Figure A, which uses the number 21 as the seed.

Step 4. Press **MATH**, select **PRB**, then **5: randIntNoRep**. Then enter **1, N, n**, where **N** is the population size and **n** is the desired sample size. In Example 1.5, we use **N** = 40 and **n** = 5 (Figure B).

Step 5. Press **ENTER**. The five values in the random sample for Example 1.5 are **27, 39, 30, 35, 17** (Figure C).

Figure A

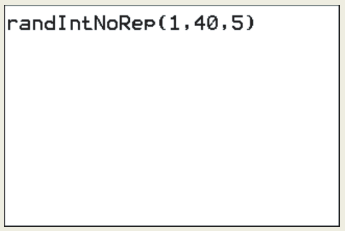


Figure B

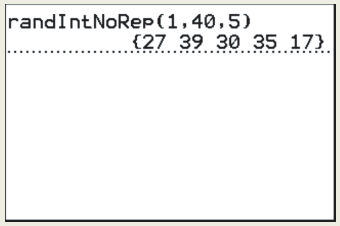


Figure C

EXCEL

Drawing a simple random sample

Step 1. In **Column A**, enter the values **1** through the population size **N**. For Example 1.5, **N** = 40.

Step 2. In **Column B**, next to each value in **Column A**, enter the command **=RAND( )**. This results (Figure D) in a randomly generated number between 0 and 1 in each cell in **Column B**.

Step 3. Select all values in **Columns A** and **B** and then click on the **Data** menu and select **Sort**.

Step 4. In the **Sort by** field, enter **Column B** and select **Smallest to Largest** in the **Order** field. Press **OK**. **Column A** now contains the random sample. Our random sample begins with **17, 12, 28, 20, 6, ...** (Figure E).

	A	B	C
1	1	0.09259	
2	2	0.93582	=rand()
3	3	0.01524	
4	4	0.74766	

Figure D

	A	B
1	17	0.9198
2	12	0.83333
3	28	0.1054
4	20	0.8635
5	6	0.82038
6	11	0.51474
7	36	0.05115
8	19	0.75697
9	39	0.82656
10	26	0.31556
11	8	0.11958
12	30	0.32979
13	35	0.59734

Figure E

MINITAB

Drawing a simple random sample

Step 1. Click **Calc**, then **Random Data**, then **Integer...**

Step 2. In the **Number of rows of data to generate** field, enter twice the desired sample size. For example, if the desired sample size is 10, enter 20. The reason for this is that some sample items may be repeated, and these will need to be deleted.

Step 3. In the **Store in column(s)** field, enter **C1**.

Step 4. Enter **1** for the **Minimum value** and the population size **N** for the **Maximum value**. We use **Maximum value** = 40 for Example 1.5. Click **OK**.

Step 5. **Column C1** of the worksheet will contain a list of randomly selected numbers between **1** and **N**. If any number appears more than once in **Column C1**, delete the replicates so that the number appears only once. For Example 1.5, our random sample begins with **16, 14, 30, 28, 17, ...** (Figure F).

↓	C1
1	16
2	14
3	30
4	28
5	17
6	13
7	4
8	8
9	6
10	15
11	35
12	5

Figure F

Section 1.1 Exercises

Exercises 1–6 are the Check Your Understanding exercises located within the section.

Understanding the Concepts

In Exercises 7–12, fill in each blank with the appropriate word or phrase.

7. The entire collection of individuals about which information is sought is called a _____.

8. A _____ is a subset of a population.

9. A _____ is a type of sample that is analogous to a lottery.

10. A sample that is not drawn by a well-defined random method is called a _____.

11. A _____ sample is one in which the population is divided into groups and a random sample of groups is drawn.

12. A _____ sample is one in which the population is divided into groups and a random sample is drawn from each group.

In Exercises 13–16, determine whether the statement is true or false. If the statement is false, rewrite it as a true statement.

13. A sample of convenience is never acceptable.

14. In a cluster sample, the population is divided into groups, and a random sample from each group is drawn.

15. Both stratified sampling and cluster sampling divide the population into groups.

16. One reason that voluntary response sampling is unreliable is that people with stronger views tend to express them more readily.

Practicing the Skills

In Exercises 17–22, determine whether the number described is a statistic or a parameter.

17. In a recent poll, 57% of the respondents supported a school bond issue.

18. The average age of the employees in a certain company is 35 years.

19. Of the students enrolled in a certain college, 80% are full-time.

20. In a survey of 500 high school students, 60% of them said that they intended to go to college.

21. The U.S. Census reports that 95% of California residents live in urban areas.

22. In a survey of parents with children under the age of six, 90% said that their child had been seen by a pediatrician within the past year.

Exercises 23–26 refer to the population of animals in the following table. The population is divided into four groups: mammals, birds, reptiles, and fish.

Mammals		Birds	
1. Aardvark	6. Lion	11. Flamingo	16. Hawk
2. Buffalo	7. Zebra	12. Swan	17. Owl
3. Elephant	8. Pig	13. Sparrow	18. Chicken
4. Squirrel	9. Dog	14. Parrot	19. Duck
5. Rabbit	10. Horse	15. Pelican	20. Turkey
Reptiles		**Fish**	
21. Gecko	26. Python	31. Catfish	36. Shark
22. Iguana	27. Turtle	32. Tuna	37. Trout
23. Chameleon	28. Tortoise	33. Cod	38. Perch
24. Rattlesnake	29. Alligator	34. Salmon	39. Guppy
25. Boa constrictor	30. Crocodile	35. Goldfish	40. Minnow

23. Simple random sample: Draw a simple random sample of eight animals from the list of 40 animals in the table.

24. Another sample: Draw a sample of eight animals by drawing a simple random sample of two animals from each group. What kind of sample is this?

25. Another sample: Draw a simple random sample of two groups of animals from the four groups, and construct a sample of 20 animals by including all the animals in the sampled groups. What kind of sample is this?

26. Another sample: Choose a random number between 1 and 5. Include the animal with that number in your sample, along with every fifth animal thereafter, to construct a sample of eight animals. What kind of sample is this?

In Exercises 27–42, identify the kind of sample that is described.

27. Parking on campus: A college faculty consists of 400 men and 250 women. The college administration wants to draw a sample of 65 faculty members to ask their opinion about a new parking fee. They draw a simple random sample of 40 men and another simple random sample of 25 women.

28. Cruising the mall: A pollster walks around a busy shopping mall and approaches people passing by to ask them how often they shop at the mall.

29. What's on TV? A pollster obtains a list of all the residential addresses in a certain town and uses a computer random number generator to choose 150 of them. The pollster visits each of the 150 households and interviews all the adults in each household about their television viewing habits.

30. Don't drink and drive: Police at a sobriety checkpoint pull over every fifth car to determine whether the driver is sober.

31. Tell us your opinion: A television newscaster invites viewers to tweet their opinions on a proposed bill on immigration policy. More than 50,000 people express their opinions in this way.

32. Reading program: The superintendent of a large school district wants to test the effectiveness of a new program designed to improve reading skills among elementary school children. There are 30 elementary schools in the district. The superintendent chooses a simple random sample of five schools and institutes the new reading program in those schools. A total of 4700 children attend these five schools.

Somos Images/Alamy

33. Customer survey: All the customers who entered a store on a particular day were given a survey to fill out concerning their opinions of the service at the store.

34. Raffle: Five hundred people attend a charity event, and each buys a raffle ticket. The 500 ticket stubs are put in a drum and thoroughly mixed, and 10 of them are drawn. The 10 people whose tickets are drawn win a prize.

35. Hospital survey: The director of a hospital pharmacy chooses at random 100 people age 60 or older from each of three surrounding counties to ask their opinions of a new prescription drug program.

36. Bus schedule: Officials at a metropolitan transit authority want to get input from people who use a certain bus route about a possible change in the schedule. They randomly select 5 buses during a certain week and poll all riders on those buses about the change.

37. How much did you spend? A retailer samples 25 receipts from the past week by numbering all the receipts, generating

25 random numbers, and sampling the receipts that correspond to these numbers.

38. **Phone features:** A phone company wants to draw a sample of 600 customers to gather opinions about potential new features on upcoming phone models. The company draws a random sample of 200 from customers with iPhones, a random sample of 100 from customers with LG phones, a random sample of 100 from customers with Samsung phones, and a random sample of 200 from customers with other phones.
39. **Computer network:** Every third day, a computer network administrator analyzes the company's network logs to check for signs of computer viruses.
40. **Apps:** An app produces a message requesting customers to click on a link to rate the app.
41. **Survey:** A nutritionist randomly chooses 100 people who have been using a certain weight loss program and asks them how much weight they have lost.
42. **Good service:** Receipts at a department store indicate a website where customers can take a survey indicating their level of satisfaction with the service they received.

Working with the Concepts

43. **You're giving me a headache:** A pharmaceutical company wants to test a new drug that is designed to provide superior relief from headaches. They want to select a sample of headache sufferers to try the drug. Do you think that it is feasible to draw a simple random sample of headache sufferers, or will it be necessary to use a sample of convenience? Explain your reasoning.
44. **Pay more for recreation?** The director of the recreation center at a large university wants to sample 100 students to ask them whether they would support an increase in their recreation fees in order to expand the hours that the center is open. Do you think that it is feasible to draw a simple random sample of students, or will it be necessary to use a sample of convenience? Explain your reasoning.
45. **Voter preferences:** A pollster wants to sample 500 voters in a town to ask them who they plan to vote for in an upcoming election. Describe a sampling method that would be appropriate in this situation. Explain your reasoning.
46. **Quality control:** Products come off an assembly line at the rate of several hundred per hour. It is desired to sample 10% of them to check whether they meet quality standards. Describe a sampling method that would be appropriate in this situation. Explain your reasoning.
47. **On-site day care:** A large company wants to sample 200 employees to ask their opinions about providing a day care center for the employees' children. They want to be sure to sample equal numbers of men and women. Describe a sampling method that would be appropriate in this situation. Explain your reasoning.
48. **The tax man cometh:** The Internal Revenue Service wants to sample 1000 tax returns that were submitted last year to determine the percentage of returns that had a refund. Describe a sampling method that would be appropriate in this situation. Explain your reasoning.

Extending the Concepts

49. **Draw a sample:** Imagine that you are asked to determine students' opinions at your school about a potential change in library hours. Describe how you could go about getting a sample of each of the following types: simple random sample, sample of convenience, voluntary response sample, stratified sample, cluster sample, systematic sample.
50. **A systematic sample is a cluster sample:** Explain how a systematic sample is actually a type of cluster sample.

Answers to Check Your Understanding Exercises for Section 1.1

1. No; this sample consists only of people in the town who visit the mall.
2. Yes; every group of n customers, where n is the sample size, is equally likely to be chosen.
3. Voluntary response sample
4. Cluster sample
5. Stratified sample
6. Systematic sample

Section 1.2 Types of Data

Objectives

1. Understand the structure of a typical data set
2. Distinguish between qualitative and quantitative variables
3. Distinguish between ordinal and nominal variables
4. Distinguish between discrete and continuous variables
5. Distinguish between ratio and interval levels of measurement

Objective 1 Understand the structure of a typical data set

Data Sets

In Section 1.1, we described various methods of collecting information by sampling. Once the information has been collected, the collection is called a **data set**. A simple example of a data set is presented in Table 1.1, which shows the major, final exam score, and grade for several students in a certain statistics class.

Comstock/Stockbyte/Getty Images

Table 1.1 Major, Final Exam Score, and Grade for Several Students

Student	Major	Exam Score	Grade
1	Psychology	92	A
2	Business	75	B
3	Communications	82	B
4	Psychology	72	C
5	Art	85	B

Table 1.1 illustrates some basic features that are found in most data sets. Information is collected on **individuals**. In this example, the individuals are students. In many cases, individuals are people; in other cases, they can be animals, plants, or things. The characteristics of the individuals about which we collect information are called **variables**. In this example, the variables are major, exam score, and grade. Finally, the values of the variables that we obtain are the **data**. So, for example, the data for individual #1 are Major = Psychology, Exam score = 92, and Grade = A.

Check Your Understanding

1. A pollster asks a group of six voters about their political affiliation (Republican, Democrat, or Independent), their age, and whether they voted in the last election. The results are shown in the following table.

Voter	Political Affiliation	Age	Voted in Last Election?
1	Republican	34	Yes
2	Democrat	56	Yes
3	Democrat	21	No
4	Independent	28	Yes
5	Republican	61	No
6	Independent	46	Yes

a. How many individuals are there?
b. Identify the variables.
c. What are the data for individual #3?

Answers are on page 19.

Objective 2 Distinguish between qualitative and quantitative variables

EXPLAIN IT AGAIN
Another way to distinguish qualitative from quantitative variables: Quantitative variables are counts or measurements, whereas qualitative variables are descriptions.

Qualitative and Quantitative Variables

Variables can be divided into two types: qualitative and quantitative. **Qualitative variables**, also called **categorical variables**, classify individuals into categories. For example, college major and gender are qualitative variables. **Quantitative variables** are numerical and tell how much or how many of something there is. Height and score on an exam are examples of quantitative variables.

SUMMARY

- Qualitative variables classify individuals into categories.
- Quantitative variables tell how much or how many of something there is.

Example 1.11

Distinguishing between qualitative and quantitative variables

Which of the following variables are qualitative and which are quantitative?

a. A person's age
b. A person's place of birth

c. The mileage (in miles per gallon) of a car
d. The color of a car

Solution

a. Age is quantitative. It tells how much time has elapsed since the person was born.
b. Place is qualitative. It includes categories such as "New York," "Atlanta," "Denver," and "Los Angeles."
c. Mileage is quantitative. It tells how many miles a car will go on a gallon of gasoline.
d. Color is qualitative. It includes categories such as "blue," "red," and "yellow."

Objective 3 Distinguish between ordinal and nominal variables

Ordinal and Nominal Variables

Qualitative variables come in two types: **ordinal variables** and **nominal variables**. An ordinal variable is one whose categories have a natural ordering. The letter grade received in a class, such as A, B, C, D, or F, is an ordinal variable. A nominal variable is one whose categories have no natural ordering. Gender is an example of a nominal variable. Figure 1.5 illustrates how qualitative variables are divided into nominal and ordinal variables.

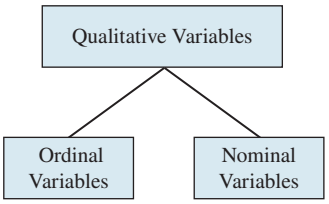


Figure 1.5 Qualitative variables come in two types: ordinal variables and nominal variables.

SUMMARY

- Ordinal variables are qualitative variables whose categories have a natural ordering.
- Nominal variables are qualitative variables whose categories have no natural ordering.

Example 1.12

Distinguishing between ordinal and nominal variables

Which of the following variables are ordinal and which are nominal?

a. State of residence
b. Gender
c. Letter grade in a statistics class (A, B, C, D, or F)
d. Size of soft drink ordered at a fast-food restaurant (small, medium, or large)

Solution

a. State of residence is nominal. There is no natural ordering to the states.
b. Gender is nominal.
c. Letter grade in a statistics class is ordinal. The order, from high to low, is A, B, C, D, F.
d. Size of soft drink is ordinal.

Objective 4 Distinguish between discrete and continuous variables

Discrete and Continuous Variables

Quantitative variables can be either *discrete* or *continuous*. **Discrete variables** are those whose possible values can be listed. Often, discrete variables result from counting something, so the possible values of the variable are 0, 1, 2, and so forth. **Continuous variables** can, in principle, take on any value within some interval. For example, height is a continuous variable because someone's height can be 68, or 68.1, or 68.1452389 inches. The possible values for height are not restricted to a list. Figure 1.6 illustrates how quantitative variables are divided into discrete and continuous variables.

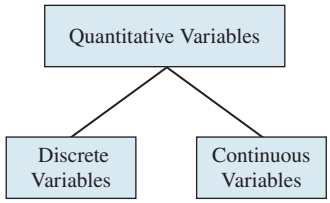


Figure 1.6 Quantitative variables come in two types: discrete variables and continuous variables.

SUMMARY

- Discrete variables are quantitative variables whose possible values can be listed. The list may be infinite—for example, the list of all whole numbers.
- Continuous variables are quantitative variables that can take on any value in some interval. The possible values of a continuous variable are not restricted to any list.

Example 1.13

Distinguishing between discrete and continuous variables

Which of the following variables are discrete, and which are continuous?

a. The age of a person at his or her last birthday
b. The height of a person
c. The number of siblings a person has
d. The distance a person commutes to work

Solution

a. Age at a person's last birthday is discrete. The possible values are 0, 1, 2, and so forth.
b. Height is continuous. A person's height is not restricted to any list of values.
c. Number of siblings is discrete. The possible values are 0, 1, 2, and so forth.
d. Distance commuted to work is continuous. It is not restricted to any list of values.

Objective 5 Distinguish between ratio and interval levels of measurement

Ratio and Interval Levels of Measurement

Quantitative variables can be categorized as having a ratio or an interval level of measurement. A variable has a **ratio level of measurement** if a value of zero indicates that none of the quantity is present, and if ratios of values of the variable are meaningful. For example, money is measured at the ratio level. An amount of \$0 represents no money. Ratios are meaningful as well: for example, \$20 is twice as much as \$10.

A variable has the **interval level of measurement** if a value of zero does not indicate that none of the quantity is present. Two commonly encountered interval variables are dates and temperature. There is no value of 0 for dates; we do not denote the beginning of time as the year 0. Similarly, when measured in °F or °C, a temperature of 0° does not indicate an absence of heat. Ratios are not meaningful for interval variables; for example, a temperature of 100° is not twice as hot as a temperature of 50°. Differences are meaningful, however; for example, the year 2030 is 10 years later than the year 2020.

SUMMARY

- A variable has the ratio level of measurement if zero represents the absence of the quantity, and ratios are meaningful.
- A variable has the interval level of measurement if zero does not represent the absence of the quantity, and ratios are not meaningful. Differences are meaningful, however.

Example 1.14

Distinguishing between interval and ratio levels of measurement

Which of the following variables are at the interval and which are at the ratio level of measurement?

1. The number of siblings you have
2. The outdoor temperature in °C

3. The year of the next presidential election
4. The price of a pair of shoes

Solution

a. Ratio. Zero siblings means the absence of siblings, and 4 siblings is twice as many as 2.
b. Interval. Temperature in °C has the interval level of measurement, because 0° does not represent the absence of heat.
c. Interval. Dates have the interval level of measurement.
d. Ratio. A price of $0 indicates that the shoes cost no money, and a $100 pair of shoes is twice as expensive as a $50 pair of shoes.

Check Your Understanding

2. Which are qualitative and which are quantitative?
- **a.** The number of patients admitted to a hospital on a given day
- **b.** The model of car last sold by a particular car dealer
- **c.** The name of your favorite song
- **d.** The seating capacity of an auditorium

3. Which are nominal and which are ordinal?
- **a.** The names of the streets in a town
- **b.** The movie ratings G, PG, PG-13, R, and NC-17
- **c.** The winners of the gold, silver, and bronze medals in an Olympic swimming competition

4. Which are discrete and which are continuous?
- **a.** The number of female members of the U.S. House of Representatives
- **b.** The amount of water used by a household during a given month
- **c.** The number of stories in an apartment building
- **d.** A person's body temperature

5. Which are at the interval and which are at the ratio level of measurement?
- **a.** The year you started school
- **b.** Your age in years
- **c.** The time that your first class starts
- **d.** The price of a loaf of bread

Answers are on page 19.

Section 1.2

Exercises

Exercises 1–5 are the Check Your Understanding exercises located within the section.

Understanding the Concepts

In Exercises 6–12, fill in each blank with the appropriate word or phrase.

6. The characteristics of individuals about which we collect information are called _____.

7. The values of variables are called _____ .

8. Variables that classify individuals into categories are called _____.

9. _____ variables are always numerical.

10. Qualitative variables can be divided into two types: _____ and _____.

11. A _____ variable is a quantitative variable whose possible values can be listed.

12. _____ variables can take on any value in some interval.

In Exercises 13–16, determine whether the statement is true or false. If the statement is false, rewrite it as a true statement.

13. Qualitative variables describe how much or how many of something there is.

14. A nominal variable is a qualitative variable with no natural ordering.

15. A discrete variable is one whose possible values can be listed.

16. A person's height is an example of a continuous variable.

Practicing the Skills

In Exercises 17–26, determine whether the data described are qualitative or quantitative.

17. Your best friend's name

18. Your best friend's age

19. The number of touchdowns in a football game

20. The title of your statistics book

21. The number of files on a computer

22. The waist size of a pair of jeans

23. The ingredients in a recipe

24. Your school colors

25. The makes of cars sold by a particular car dealer

26. The number of cars sold by a car dealer last month

In Exercises 27–34, determine whether the data described are nominal or ordinal.

27. The categories Strongly disagree, Disagree, Neutral, Agree, and Strongly agree on a survey

28. The names of the counties in a state

29. The shirt sizes of Small, Medium, Large, and X-Large

30. I got an A in statistics, a B in biology, and C's in history and English.

31. This semester, I am taking statistics, biology, history, and English.

32. I ordered a pizza with pepperoni, mushrooms, olives, and onions.

33. In the track meet, I competed in the high jump and the pole vault.

34. I finished first in the high jump and third in the pole vault.

In Exercises 35–42, determine whether the data described are discrete or continuous.

35. The amount of caffeine in a cup of Starbucks coffee

36. The distance from a student's home to his school

37. The number of steps in a stairway

38. The number of students enrolled at a college

39. The amount of charge left in a phone battery

40. The number of patients who reported that a new drug had relieved their pain

41. The number of electrical outlets in a coffee shop

42. The time it takes for a text message to be delivered

In Exercises 43–50, determine whether the data described are at the interval level or the ratio level of measurement.

43. The year of your birth

44. The sales price of a car

45. The weight in pounds of a sack of potatoes

46. The score on an SAT exam (range is 200 to 800 points)

47. The water temperature, in °C, in a swimming pool

48. The number of Snapchat friends someone has

49. A student's GPA

50. The amount of data, in gigabytes, remaining on your phone

Working with the Concepts

51. Popular Videos: Following are the ten most-viewed YouTube videos, as of January 2020:

1. Despacito—Luis Fonsi featuring Daddy Yankee
2. Shape of You—Ed Sheeran
3. Baby Shark Dance—Pinkfong Kids' Songs & Stories
4. See You Again—Wiz Khalifa featuring Charlie Puth
5. Masha and the Bear: "Recipe for Disaster"—Get Movies
6. Uptown Funk—Mark Ronson featuring Bruno Mars
7. Gangnam Style—Psy
8. Sorry—Justin Bieber
9. Sugar—Maroon 5
10. Roar—Katy Perry

Source: Wikipedia

a. Are these data nominal or ordinal?
b. Are these data qualitative or quantitative?

52. More Videos: The following table presents the number of views (in millions) for the videos in Exercise 51.

Video	Number of views (millions)
1. Despacito—Luis Fonsi featuring Daddy Yankee	6610
2. Shape of You—Ed Sheeran	4600
3. Baby Shark Dance—Pinkfong Kids' Songs & Stories	4470
4. See You Again—Wiz Khalifa featuring Charlie Puth	4390
5. Masha and the Bear: "Recipe for Disaster"—Get Movies	4220
6. Uptown Funk—Mark Ronson featuring Bruno Mars	3770
7. Gangnam Style—Psy	3500
8. Sorry—Justin Bieber	3240
9. Sugar—Maroon 5	3120
10. Roar—Katy Perry	2990

Source: Wikipedia

a. Are these data discrete or continuous?
b. Are these data qualitative or quantitative?

53. How's the economy? A poll conducted by the American Research Group asked individuals their views on how the economy will be a year from now. Respondents were given four choices: Better than today, Same as today, Worse than today, and Undecided. Are these choices nominal or ordinal?

54. Global warming: A recent Pew poll asked people between the ages of 18 and 29 how serious a problem global warming is. Of those who responded, 43% thought it was very serious, 24% thought it was somewhat serious, 15% thought it was not too serious, and 17% thought it was not a problem. Are these percentages qualitative or quantitative?

55. Graphic Novels: According to *Time* magazine, some of the best graphic novels of all time are:

Watchmen by Alan Moore and Dave Gibbons
Sandman by Neil Gaiman
Jimmy Corrigan, the Smartest Kid on Earth by Chris Ware
Maus by Art Spiegelman
The Adventures of Tintin: The Black Island by Hergé
Miracleman: The Golden Age by Neil Gaiman and Mark Buckingham
Fun Home: A Family Tragicomic by Allison Bechdel
The Greatest of Marlys by Lynda Berry

Are these data nominal or ordinal?

56. Watch your language: According to MerriamWebster Online, the top ten Funny Sounding and Interesting words are:

1. Bumfuzzle
2. Cattywampus
3. Gardyloo
4. Taradiddle
5. Billingsgate
6. Snickersnee
7. Widdershins
8. Collywobbles
9. Gubbins
10. Diphthong

Are these data nominal or ordinal?

57. Top ten video games: According to *Wikipedia*, the following are the top ten selling video games of all time.

Game Title	Initial Release Year	Developer	Copies Sold (millions)
1. Tetris	1984	Elektronorgtechnica	170.0
2. Minecraft	2011	Mojang	154.0
3. Grand Theft Auto V	2013	Rockstar North	100.0
4. Wii Sports	2008	Nintendo	82.9
5. PlayerUnknown's Battlegrounds	2017	PUBG Corporation	50.0
6. Pokemon Red/Green/Blue/Yellow	1996	Game Freak	47.5
7. Wii Fit and Wii Fit Plus	2007	Nintendo	43.8
8. Super Mario Bros.	1985	Nintendo	43.2
9. Mario Kart Wii	2008	Wii	37.1
10. Wii Sports Resort	2009	Nintendo	33.1

a. Which of the columns represent qualitative variables?
b. Which of the columns represent quantitative variables?
c. Which of the columns represent nominal variables?
d. Which of the columns represent ordinal variables?

58. At the movies: The following table provides information about the top ten grossing movies of all time.

Movie Title	Release Year	Studio	Ticket Sales (millions of $)	Running Time (minutes)
1. Avengers: Endgame	2019	Disney	2797.8	182
2. Avatar	2009	Fox	2790.4	162
3. Titanic	1997	Paramount	2194.4	194
4. Star Wars: The Force Awakens	2015	Disney	2068.2	136
5. Avengers: Infinity War	2018	Disney	2048.4	160
6. Jurassic World	2015	Univision	1670.4	124
7. The Lion King	2019	Disney	1656.9	118
8. Marvel's The Avengers	2012	Disney	1518.8	143
9. Furious 7	2015	Univision	1515.0	140
10. Frozen II	2019	Disney	1420.8	103

Source: Box Office Mojo

a. Which of the columns represent qualitative variables?
b. Which of the columns represent quantitative variables?
c. Which of the columns represent nominal variables?
d. Which of the columns represent ordinal variables?

Extending the Concepts

59. What do the numbers mean? A survey is administered by a marketing firm. Two of the people surveyed are Brenda and Jason. Three of the questions are as follows:

i. Do you favor the construction of a new shopping mall?
(1) Strongly oppose (2) Somewhat oppose (3) Neutral (4) Somewhat favor (5) Strongly favor

ii. How many cars do you own?

iii. What is your marital status?
(1) Married (2) Single (3) Divorced (4) Domestically partnered (5) Other

a. Are the responses for question (i) nominal or ordinal?

b. On question (i), Brenda answers (2) and Jason answers (4). Jason's answer (4) is greater than Brenda's answer (2). Does Jason's answer reflect more of something?

c. Jason's answer to question (i) is twice as large as Brenda's answer. Does Jason's answer reflect twice as much of something? Explain.

d. Are the responses for question (ii) qualitative or quantitative?

e. On question (ii), Brenda answers 2 and Jason answers 1. Does Brenda's answer reflect more of something? Does Brenda's answer reflect twice as much of something? Explain.

f. Are the responses for question (iii) nominal or ordinal?

g. On question (iii), Brenda answers (4) and Jason answers (2). Does Brenda's answer reflect more of something? Does Brenda's answer reflect twice as much of something? Explain.

Answers to Check Your Understanding Exercises for Section 1.2

1. a. 6 **b.** Political affiliation, Age, and Voted in last election **c.** Political affiliation = Democrat, Age = 21, Voted in last election = no

2. a. Quantitative **b.** Qualitative **c.** Qualitative **d.** Quantitative

3. a. Nominal **b.** Ordinal **c.** Ordinal

4. a. Discrete **b.** Continuous **c.** Discrete **d.** Continuous

5. a. Interval **b.** Ratio **c.** Interval **d.** Ratio

Section 1.3 Design of Experiments

Objectives

1. Distinguish between a randomized experiment and an observational study
2. Understand the advantages of randomized experiments
3. Understand how confounding can affect the results of an observational study
4. Describe various types of observational studies

Objective 1 Distinguish between a randomized experiment and an observational study

Experiments and Observational Studies

Will a new drug help prevent heart attacks? Does one type of seed produce a larger wheat crop than another? Does exercise lower blood pressure? To illustrate how scientists address questions like these, we describe how a study might be conducted to determine which of three types of seed will result in the largest wheat yield.

- Prepare three identically sized plots of land, with similar soil types.
- Plant each type of seed on a different plot, choosing the plots at random.
- Water and fertilize the plots in the same way.
- Harvest the wheat, and measure the amount grown on each plot.
- If one type of seed produces substantially more (or less) wheat than the others, then scientists will conclude that it is better (or worse) than the others.

The following terminology is used for studies like this.

DEFINITION

The **experimental units** are the individuals that are studied. These can be people, animals, plants, or things. When the experimental units are people, they are sometimes called **subjects**.

In the wheat study just described, the experimental units are the three plots of land.

DEFINITION

The **outcome**, or **response**, is what is measured on each experimental unit.

In the wheat study, the outcome is the amount of wheat produced.

DEFINITION

The **treatments** are the procedures applied to each experimental unit. There are always two or more treatments. The purpose is to determine whether the choice of treatment affects the outcome.

In the wheat study, the treatments are the three types of seed.

In general, studies fall into two categories: *randomized experiments* and *observational studies*.

DEFINITION

A **randomized experiment** is a study in which the investigator assigns the treatments to the experimental units at random.

The wheat study described above is a randomized experiment. In some situations, randomized experiments cannot be performed, because it isn't possible to randomly assign the treatments. For example, in studies to determine how smoking affects health, people cannot be assigned to smoke. Instead, people choose for themselves whether to smoke, and scientists observe differences in health outcomes between groups of smokers and nonsmokers. Studies like this are called *observational studies*.

DEFINITION

An **observational study** is one in which the assignment to treatment groups is not made by the investigator.

When possible, it is better to assign treatments at random and perform a randomized experiment. As we will see, the results of randomized experiments are generally easier to interpret than the results of observational studies.

Randomized Experiments

Objective 2 Understand the advantages of randomized experiments

Ingram Publishing/SuperStock

An article in *The New England Journal of Medicine* (359:339–354) reported the results of a study to determine whether a drug called raltegravir is effective in reducing levels of virus in patients with human immunodeficiency virus (HIV). A total of 699 patients participated in the experiment. These patients were divided into two groups. One group was given raltegravir. The other group was given a placebo. (A placebo is a harmless tablet, such as a sugar tablet, that looks like the drug but has no medical effect.) Thus there were two treatments in this experiment, raltegravir and placebo.

The experimenters had decided to give raltegravir to about two-thirds of the subjects and the placebo to the others. To determine which patients would be assigned to which group, a simple random sample consisting of 462 of the 699 patients was drawn; this sample constituted the raltegravir group. The remaining 237 patients were assigned to the placebo group.

It was decided to examine subjects after 16 weeks and measure the levels of virus in their blood. Thus the outcome for this experiment was the number of copies of virus per milliliter of blood. Patients were considered to have a successful outcome if they had fewer than 50 copies of the virus per milliliter of blood. In the raltegravir group, 62% of the subjects had a successful outcome, but only 35% of the placebo group did. The conclusion was that raltegravir was effective in lowering the concentration of virus in HIV patients. We will examine this study and determine why it was reasonable to reach this conclusion.

The raltegravir study was a randomized experiment, because the treatments were assigned to the patients at random. What are the advantages of randomized experiments? In a perfect study, the treatment groups would not differ from each other in any important way except that they receive different treatments. Then, if the outcomes differ among the groups, we may be confident that the differences in outcome must have been caused by differences in treatment. In practice, it is impossible to construct treatment groups that are exactly alike. But randomization does the next best thing. In a randomized experiment, any differences between the groups are likely to be small. In addition, the differences are due only to chance.

Because the raltegravir study was a randomized experiment, it is reasonable to conclude that the higher success rate in the raltegravir group was actually due to raltegravir.

SUMMARY

In a randomized experiment, if there are large differences in outcomes among the treatment groups, we can conclude that the differences are due to the treatments.

Example 1.15

Identifying a randomized experiment

To assess the effectiveness of a new method for teaching arithmetic to elementary school children, a simple random sample of 30 first graders was taught with the new method, and another simple random sample of 30 first graders was taught with the currently used method. At the end of eight weeks, the children were given a test to assess their knowledge. What are the treatments in this study? Explain why this is a randomized experiment.

Solution

The treatments are the two methods of teaching. This is a randomized experiment because children were assigned to the treatment groups at random.

Double-blind experiments

We have described the advantages of assigning treatments at random. It is a further advantage if the assignment can be done in such a way that neither the experimenters nor the subjects know which treatment has been assigned to which subject. Experiments like this are called *double-blind* experiments. The raltegravir experiment was a double-blind experiment, because neither the patients nor the doctors treating them knew which patients were receiving the drug and which were receiving the placebo.

DEFINITION

An experiment is **double-blind** if neither the investigators nor the subjects know who has been assigned to which treatment.

Experiments should be run double-blind whenever possible, because when investigators or subjects know which treatment is being given, they may tend to report the results differently. For example, in an experiment to test the effectiveness of a new pain reliever, patients who know they are getting the drug may report their pain levels differently than those who know they are taking a placebo. Doctors can be affected as well; a doctor's diagnosis may be influenced by a knowledge of which treatment a patient received.

In some situations, it is not possible to run a double-blind experiment. For example, in an experiment that compares a treatment that involves taking medication to a treatment that involves surgery, both patients and doctors will know who got which treatment.

Example 1.16

Determining whether an experiment is double-blind

Is the experiment described in Example 1.15 a double-blind experiment? Explain.

Solution

This experiment is not double-blind, because the teachers know whether they are using the new method or the old method.

Randomized block experiments

The type of randomized experiment we have discussed is sometimes called a **completely randomized experiment**, because there is no restriction on which subjects may be assigned which treatment. In some situations, it is desirable to restrict the randomization a bit. For example, imagine that two reading programs are to be tested in an elementary school that has children in grades 1 through 4. If children are assigned at random to the programs, it is possible that one of the programs will end up with more fourth graders while the other one will end up with more first graders. Since fourth graders tend to be better readers, this will give an advantage to the program that happens to end up with more of them. This possibility can be avoided by randomizing the students within each grade separately. In other words, we randomly assign exactly half of the students within each grade to each reading program.

This type of experiment is called a **randomized block experiment**. In the example just discussed, each grade constitutes a block. In a randomized block experiment, the subjects are divided into blocks in such a way that the subjects in each block are the same or similar with regard to a variable that is related to the outcome. Age and gender are commonly used blocking variables. Then the subjects within each block are randomly assigned a treatment.

Observational Studies

Recall that an observational study is one in which the investigators do not assign the treatments. In most observational studies, the subjects choose their own treatments. Observational studies are less reliable than randomized experiments. To see why, imagine a study that is intended to determine whether smoking increases the risk of heart attack. Imagine that a group of smokers and a group of nonsmokers are observed for several years, and during that time a higher percentage of the smoking group experiences a heart attack. Does this prove that smoking increases the risk of heart attack? No. The problem is that the smoking group will differ from the nonsmoking group in many ways other than smoking, and these other differences may be responsible for differences in the rate of heart attacks. For example, smoking is more prevalent among men than among women. Therefore, the smoking group will contain a higher percentage of men than the nonsmoking group. It is known that men have a higher risk of heart attack than women. So the higher rate of heart attacks in the smoking group could be due to the fact that there are more men in the smoking group, and not to the smoking itself.

Objective 3 Understand how confounding can affect the results of an observational study

Confounding

The preceding example illustrates the major problem with observational studies. It is difficult to tell whether a difference in the outcome is due to the treatment or to some other difference between the treatment and control groups. This is known as **confounding**. In the preceding example, gender was a *confounder*. Gender is related to smoking (men are more likely to smoke) and to heart attacks (men are more likely to have heart attacks). For this reason, it is difficult to determine whether the difference in heart attack rates is due to differences in smoking (the treatment) or differences in gender (the confounder).

EXPLAIN IT AGAIN

Another way to describe a confounder: A confounder is something other than the treatment that can cause the treatment groups to have different outcomes.

SUMMARY

A **confounder** is a variable that is related to both the treatment and the outcome. When a confounder is present, it is difficult to determine whether differences in the outcome are due to the treatment or to the confounder.