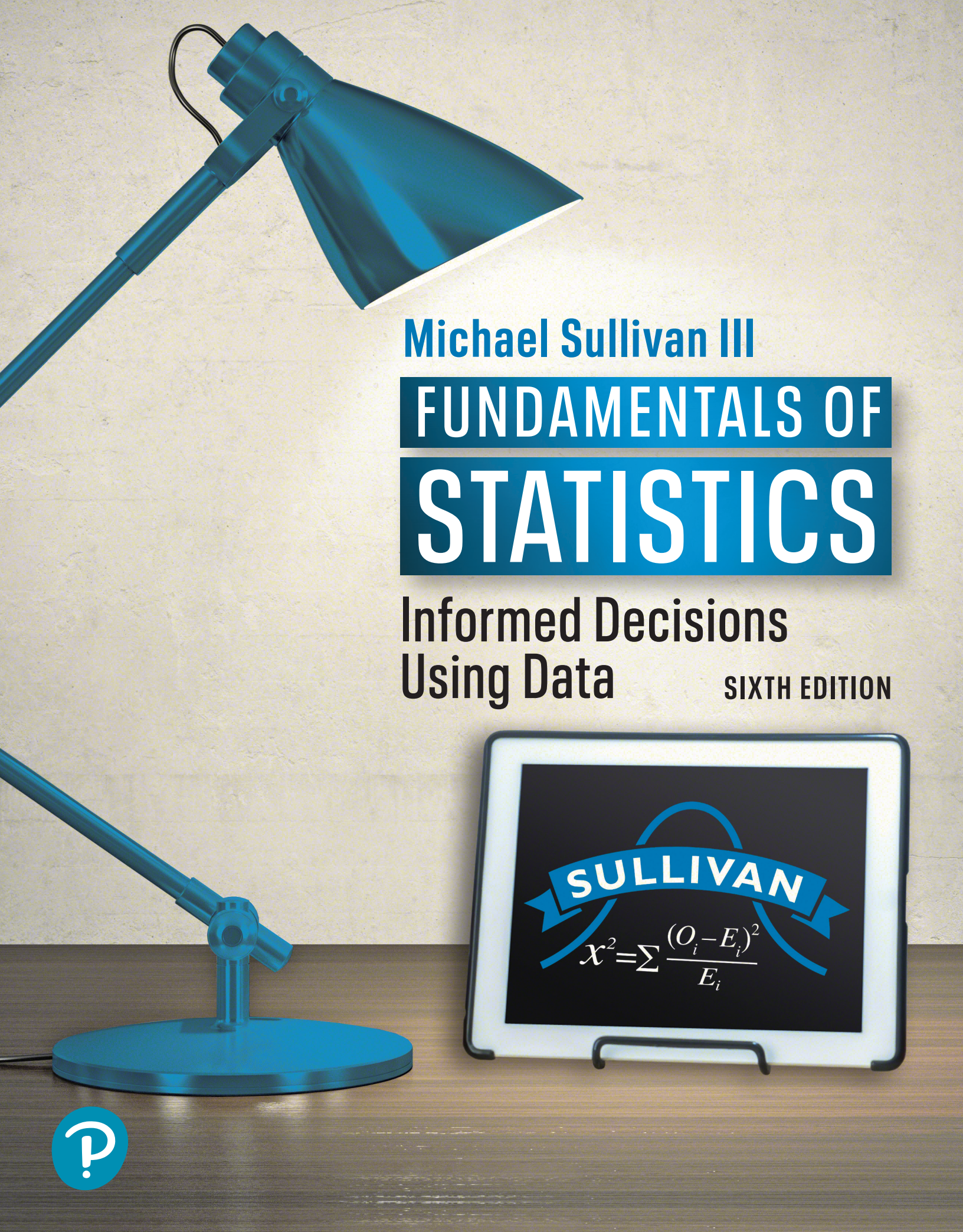

Michael Sullivan III
FUNDAMENTALS OF
STATISTICS
Informed Decisions
Using Data
SIXTH EDITION
SULLIVAN
$x^2=\Sigma\frac{(O_i-E_i)^2}{E_i}$

Sullivan's Pathway to Making an Informed Decision

Begin your journey . . .

- **Making an Informed Decision** projects at the start of each chapter allow you to work with data in order to make informed decisions that impact your life.
- **Putting It Together** overviews show how material you are about to cover relates to prior material.

Preparation is key . . .

- **Preparing for This Section** lists all of the skills needed to be successful.
- **Preparing for This Section Quizzes** are available as a digital MyLab assignment or as a print quiz to help you check your mastery.
- **Each Objective** is listed at the beginning of the section and then repeated in the text for easy reference.

Look at the model then practice, practice, practice . . .

- **Step-by-Step Annotated Examples** illustrate new concepts and methods in 3 steps:
  1. Problem
  2. Approach
  3. Solution
- **Examples** point to **Now Work Exercises** so you can solve similar exercises on your own.

Exercise Sets . . .

- **Putting It Together** exercises use skills you've acquired in various chapters. (*See facing page*)
- **You Explain It!** exercises ask you to provide an interpretation of statistical results.
- **Threaded Tornado Problems** allow you to analyze a single data set throughout the entire semester. (*See facing page*)
- **Retain Your Knowledge** exercises help you to maintain the skills you have acquired earlier in the course.

Check where you've been and test your mastery . . .

- **Putting It Together Sections** require you to decide which technique to use. (*See facing page*)
- **End-of-Chapter Objectives** are listed with page references for easy review.
- **Chapter Tests** provide an opportunity to test your knowledge.

Apply yourself . . .

- **In-Class Activities** in the Student Activity Workbook allow you to experience statistics in a fun and exciting way by experiencing the process firsthand.
- **Making an Informed Decision** projects require you to use data and statistical techniques learned in the chapter to make important life decisions.
- **End-of-Chapter Case Studies** tie statistical concepts together within an interesting application.

Sullivan's Guide to Putting It Together

Putting It Together Sections	Objective	Page(s)
5.7 Putting It Together: Which Method Do I Use?	1 Determine the appropriate probability rule to use	287–289
	2 Determine the appropriate counting technique to use	289–290
9.3 Putting It Together: Which Method Do I Use?	1 Determine the appropriate confidence interval to construct	425–426
10.4 Putting It Together: Which Method Do I Use?	1 Determine the appropriate hypothesis test to perform (one sample)	470
11.4 Putting It Together: Which Method Do I Use?	1 Determine the appropriate hypothesis test to perform (two samples)	513

Putting It Together Exercises	Skills Utilized	Section(s) Covered	Page(s)
1.2.26 Passive Smoke	Variables, observational studies, designed experiments	1.1, 1.2	23
1.4.37 Comparing Sampling Methods	Simple random sampling and other sampling techniques	1.3, 1.4	38
1.4.38 Thinking about Randomness	Random sampling	1.3, 1.4	38
2.1.29 Online Homework	Variables, designed experiments, bar graphs	1.1, 1.2, 1.6, 2.1	75
2.2.42 Time Viewing a Webpage	Graphing data	2.2	91
2.2.43 Red Light Cameras	Variables, population vs. sample, histograms, dot plots	1.1, 2.2	91
2.2.44 Which Graphical Summary?	Choosing the best graphical summary	2.1, 2.2	91
2.2.45 Shark!	Graphing data	2.3	91–92
3.1.42 Shape, Mean, and Median	Discrete vs. continuous data, histograms, shape of a distribution, mean, median, mode, bias	1.1, 1.4, 2.2, 3.1	120
3.5.18 Paternal Smoking	Observational studies, designed experiments, lurking variables, mean, median, standard deviation, quartiles, boxplots	1.2, 1.6, 3.1, 3.2, 3.4, 3.5	162–163
3.5.19 Taxi Ride	Bar graphs, histograms, boxplots, range, standard deviation	2.1, 2.2, 3.2, 3.5	163
4.2.29 Housing Prices	Scatter diagrams, correlation, linear regression	4.1, 4.2	199
4.2.30 Smoking and Birth Weight	Observational study vs. designed experiment, prospective studies, scatter diagrams, linear regression, correlation vs. causation, lurking variables	1.2, 4.1, 4.2	199–200
4.3.16 Exam Scores	Building a linear model	4.1, 4.2, 4.3	205–206
4.3.17 Cigarette Smuggling	Scatter diagrams, correlation, least-squares regression	4.1, 4.2, 4.3	206
4.4.15 Sullivan Survey II	Relative frequency distributions, bar graphs, pie charts, contingency tables, conditional distributions	2.1, 4.4	216–217
5.1.52 Drug Side Effects	Variables, graphical summaries of data, experiments, probability	1.1, 1.6, 2.1, 5.1	241
5.2.44 Speeding Tickets	Contingency tables, marginal distributions, empirical probabilities	4.4, 5.1	252
5.2.45 Red Light Cameras	Variables, relative frequency distributions, bar graphs, mean, standard deviation, probability, Simpson's Paradox	1.1, 2.1, 3.1, 3.2, 4.4, 5.1, 5.2	252–253
6.1.37 Sullivan Statistics Survey I	Mean, standard deviation, probability, probability distributions	3.1, 3.2, 5.1, 6.1	311
6.2.55 A Drug Study	Types of variables, experimental design; binomial probabilities	1.1, 1.2, 1.6, 6.2	327
6.2.56 Beating the Stock Market	Expected value, binomial probabilities	6.1, 6.2	327
7.2.52 Birth Weights	Relative frequency distribution, histograms, mean and standard deviation from grouped data, normal probabilities	2.1, 2.2, 3.3, 7.2	354
7.3.13 Disney's Dinosaur Ride	Histograms, distribution shape, normal probability plots	2.2, 7.3	359
8.1.34 Bike Sharing	Histograms, mean, standard deviation, distribution shape, sampling distribution of the mean	2.2, 3.1, 3.2, 8.1	383
8.1.35 Playing Roulette	Probability distributions, mean and standard deviation of a random variable, sampling distributions	6.1, 8.1	383
9.1.47 Hand Washing	Observational studies, bias, confidence intervals	1.2, 1.5, 9.1	409–410
9.2.47 Smoking Cessation Study	Experimental design, confidence intervals	1.6, 9.1, 9.2	424
10.2.40 Lupus	Observational studies, retrospective vs. prospective studies, bar graphs, confidence intervals, hypothesis testing	1.2, 2.1, 9.1, 10.2	457
10.2.41 Naughty or Nice?	Experimental design, determining null and alternative hypotheses, binomial probabilities, interpreting *P*-values	1.6, 6.2, 10.1, 10.2	458
11.1.36 Salk Vaccine	Completely randomized design, hypothesis testing	1.6, 11.1	490

(continued)

Putting It Together Exercises	Skills Utilized	Section(s) Covered	Page(s)
11.2.19 Glide Testing	Matched pairs design, hypothesis testing	1.6, 11.2	500–501
11.3.23 Online Homework	Completely randomized design, confounding, hypothesis testing	1.6, 11.3	512
12.1.27 The V-2 Rocket in London	Mean of discrete data, expected value, Poisson probability distribution, goodness-of-fit	6.1, 6.3, 12.1	537
12.1.28 Weldon's Dice	Addition Rule for Disjoint Events, classical probability, goodness-of-fit	5.1, 5.2, 12.1	537
12.2.22 Women, Aspirin, and Heart Attacks	Population, sample, variables, observational study vs. designed experiment, experimental design, compare two proportions, chi-square test of homogeneity	1.1, 1.2, 1.6, 11.1, 12.2	552–553
12.2.23 Corequisite College Algebra	Comparing two independent means, comparing two independent proportions, chi-square test for independence	11.1, 11.3, 12.2	553
12.4.19 Predicting Intelligence	Scatter diagrams, linear correlation coefficient, least-squares regression, normal probability plots, inference on least-squares regression, confidence and prediction intervals	4.1, 4.2, 4.3. 7.3, 12.3, 12.4	574
B.3.27 Psychological Profiles	Standard deviation, sampling methods, two-sample *t*-test, Central Limit Theorem, one-way Analysis of Variance	1.4, 3.2, 8.1, 11.2, B.3	B-28

Threaded Tornado Problems

Throughout the text a single, large data set that measures various variables on all tornadoes that struck the United States in 2017 is utilized. The problems are marked with a icon. The table below shows the sections, problems, topics covered, and page for the Threaded Tornado Problems.

Section	Problem(s)	Topics	Page(s)
1.1	47, 48	Types of variables; types of data	13
2.1	25	Frequency & relative frequency distributions; bar charts; pie charts	74–75
2.2	41	Frequency & relative frequency distributions; histogram; dot plots	91
3.1	41	Mean, median, distribution shape	120
3.2	51	Range, standard deviation	138
3.4	29	Quartiles, interquartile range, outliers	155
3.5	20	Boxplots	163
4.3	15	Scatter diagrams, correlation, least-squares regression, coefficient of determination, residual analysis	205
5.1	49	Probability models; unusual events	240
8.1	33	Describe the distribution of the sample mean from a non-normal population	382–383
9.1	33	Confidence interval for a population proportion	408
9.2	37	Confidence interval for a population mean	422–423
10.2	33	Hypothesis test for a population proportion	456
10.2B	25	Hypothesis test for a population proportion	10.2AB.24
10.3	35	Hypothesis test for a population mean	468
11.1	29	Compare two population proportions (independent samples)	489
11.3	17	Compare two population means (independent samples)	511
12.3	17	Inference on least-squares regression; prediction intervals	573–574
B.3	29	One-way Analysis of Variance (ANOVA)	B-28

FUNDAMENTALS OF STATISTICS

INFORMED DECISIONS USING DATA 6E

Michael Sullivan, III

Joliet Junior College

Pearson

Content Management: Suzanna Bainbridge
Content Production: Robert Carroll, Jean Choe, Peggy McMahon
Product Management: Karen Montgomery
Product Marketing: Alicia Wilson
Rights and Permissions: Pavithra Gunasekaran, Integra

Please contact https://support.pearson.com/getsupport/s/contactsupport with any queries on this content.

Library of Congress Cataloging-in-Publication Data

Names: Sullivan, Michael, III, 1967- author.
Title: Fundamentals of statistics : informed decisions using data / Michael Sullivan, III, Joliet Junior College.
Description: 6th edition. | Hoboken : Pearson, 2022. | Includes index. | Summary: "Popular college introductory text to the subject of Statistics that follows GAISE guidelines"-- Provided by publisher.
Identifiers: LCCN 2020035728 | ISBN 9780136807346 (hardback)
Subjects: LCSH: Statistics--Textbooks.
Classification: LCC QA276.12 .S846 2022 | DDC 519.5--dc23
LC record available at https://lccn.loc.gov/2020035728

ScoutAutomatedPrintCode

ISBN 10: 0-13-680734-8
ISBN 13: 978-0-13-680734-6

To My Wife Yolanda
and My Children
Michael, Kevin, and Marissa

This page intentionally left blank

Contents

Preface to the Instructor

Capturing a Powerful and Exciting Discipline in a Textbook

Statistics is a powerful subject, and it is one of my passions. Bringing my passion for the subject together with my desire to create a text that would work for me, my students, and my school led me to write the first edition of this textbook. It continues to motivate me as I reflect on changes in students, in the statistics community, and in the world around us.

When I started writing, I used the manuscript of this text in class. My students provided valuable, insightful feedback, and I made adjustments based on their comments. In many respects, this text was written by students and for students. I also received constructive feedback from a wide range of statistics faculty, which has refined ideas in the book and in my teaching. I continue to receive valuable feedback from both faculty and students, and this text continues to evolve with the goal of providing clear, concise, and readable explanations, while challenging students to think statistically.

In writing this edition, I continue to make a special effort to abide by the Guidelines for Assessment and Instruction in Statistics Education (GAISE) for the college introductory course endorsed by the American Statistical Association (ASA). The GAISE Report gives six recommendations for the course:

1. Emphasize statistical literacy and develop statistical thinking
2. Use real data in teaching statistics
3. Stress conceptual understanding
4. Foster active learning
5. Use technology for developing conceptual understanding
6. Use assessments to improve and evaluate student learning

Changes to this edition and the hallmark features of the text reflect a strong adherence to these important GAISE guidelines.

New to This Edition

- **Over 350 New and Updated Exercises** The sixth edition makes a concerted effort to require students to write a few sentences that explain the results of their statistical analysis. To reflect this effort, the answers in the back of the text provide recommended explanations of the statistical results. Not all the exercises are computational or require statistical analysis. Many of the exercises have been written to require students to explain statistical concepts or understand pitfalls in faulty statistical analysis.
- **Over 100 New and Updated Examples** The examples continue to engage and provide clear, concise explanations for the students while following the *Problem, Approach, Solution* presentation. Problem lays out the scenario of the example, Approach provides insight into the thought process behind the methodology used to solve the problem, and Solution goes through the solution utilizing the methodology suggested in the approach.
- **Threaded Tornado Problems** Throughout the text a single, large data set that measures various variables on all tornadoes that struck the United States in 2017 is utilized. The problems are marked with a icon. The table on the front inside cover shows the sections, problems, topics covered and pages for the Threaded Tornado Problems. In addition, the author wrote corresponding MyLab problems around this data set. The problems may serve as a semester-long project for your students.
- **Updated MyLab Problems** New MyLab problems written by Michael Sullivan utilize real data that is randomly generated from a larger data set. He also wrote new applet exercises that allow students to explore statistical concepts.
- **Optional Simulation & Randomization Sections** Simulation and randomization methods are a new approach to hypothesis testing. New to this edition are optional sections on using simulation to test hypotheses for a population proportion (Section 10.2A) and population mean (Section 10.3A), and randomization methods for testing hypotheses on two independent proportions (Section 11.1A), two independent means (Section 11.3A), and the slope of the least-squares regression model (Section 12.3A).
- **Classroom Notes** Written by Heidi Lyne and Michael Sullivan, new to this edition are classroom notes, which may be used by the instructor to deliver lectures to students. Students may print these notes out and bring them to the classroom, which facilitates good note-taking and allows them to focus on the concepts. The examples and activities in the classroom notes are different from those in the text and Instructor's Resource Guide.
- **Videos** New lightboard videos featuring the author, Michael Sullivan, develop statistical concepts for students. New animated videos explain concepts or tie material learned earlier in the course with the upcoming chapter or section. And finally, new Excel video solutions for any example in which Excel may be used to obtain statistical results are available.
- **R Technology Guide** Written by Patrick Murphy (nephew of the author) and Michael Sullivan, the R Technology Guide provides a chapter-by-chapter discussion of R commands needed for each topic. The R Technology Guide may be found under Learning Tools in MyLab or at www.sullystats.com/statistics-6e/r-guidebook.
- **Learning Catalytics** Learning Catalytics allows students to use their own mobile devices in the classroom for real-time engagement. Search "SullivanStats" in Learning Catalytics to add pre-made questions written by Michael Sullivan for Sullivan's *Statistics* series.

Hallmark Features

- **Putting It Together** When students are learning statistics, they often struggle with seeing the big picture of how it all fits together. One of my goals is to help students learn not just the important concepts and methods of statistics but also how to put them together and see how the methods work together. On the inside front cover, you'll see a pathway that provides a guide for students as they navigate through the process of learning statistics. The features and chapter organization in the sixth edition reinforce this important process. There are two categories of "Putting It Together."
  - **Putting It Together Sections** appear in Chapters 5, 9, 10, and 11. The problems in these sections are meant to help students identify the correct approach to solving a problem. Many exercises in these sections mix in inferential techniques from earlier sections. Plus, there are problems that require students to identify the inferential technique that may be used to answer the research objective (but no analysis is required). For example, see Problems 20 to 25 in Section 10.4.
  - **Putting It Together Problems** appear throughout the text. The purpose of these problems is to tie concepts together and see the entire statistical process. For example, problems on hypothesis testing may require students to first identify the data collection method (such as observational study or designed experiment, the explanatory and response variables, the role of randomization, the role of control) prior to completing the data analysis.
- **Student Activity Workbook** The student activity workbook now contains an outline for a semester-long project and suggestions for how to use the StatCrunch survey tool to develop a survey that could result in a semester-long project. Plus, there are ten new activities included in the activity workbook along with suggested answers in the corresponding instructor's guide.
- **Retain Your Knowledge** These problems occur periodically at the end of section exercises and are meant to assist students in retaining skills learned earlier in the course. This way, the material is fresh for the final exam.
- **MyLab Technology Help** Online homework problems that may be analyzed using statistical packages now have an updated technology help feature. Marked with a  icon, this feature provides step-by-step instructions on how to obtain results using StatCrunch, TI-84 Plus/TI-84 Plus CE, and Excel.
- **Instructor's Resource Guide** Written by Michael Sullivan, the Instructor's Resource Guide provides an overview of the chapter. It also details points to emphasize within each section and suggestions for presenting the material. In addition, the guide provides examples that may be used in the classroom. Many new examples have been added to this edition.
- Because the use of **Real Data** piques student interest and helps show the relevance of statistics, great efforts have been made to extensively incorporate real data in the exercises and examples.
- **Step-by-Step Annotated Examples** guide a student from problem to solution in three easy-to-follow steps.
- **"Now Work"** problems follow most examples so students can practice the concepts shown.
- Multiple types of **Exercises** are used at the end of sections and chapters to test varying skills with progressive levels of difficulty. These exercises include **Vocabulary and Skill Building**, **Applying the Concepts**, and **Explaining the Concepts**.
- **Chapter Review** sections include:
  - **Chapter Summary**.
  - A list of key chapter **Vocabulary**.
  - A list of **Formulas** used in the chapter.
  - **Chapter Objectives** listed with corresponding review exercises.
  - **Review Exercises** with all answers available in the back of the book.
  - **Chapter Test** with all answers available in the back of the book. In addition, the Chapter Test problems have **video solutions** available.
- Each chapter has **Case Studies** available at www.pearsonhighered.com/sullivanstats that help students apply their knowledge and promote active learning.

Integration of Technology

This book can be used with or without technology. Should you choose to integrate technology in the course, the following resources are available for your students:

- Technology Step-by-Step guides are included in applicable sections that show how to use Minitab®, Excel®, the TI-83/84, and StatCrunch to complete statistics processes. The Technology Step-by-Step for StatCrunch was written by Michael Sullivan.
- Any problem that has 12 or more observations in the data set has a  icon indicating that data set is included on the companion website (http://www.pearsonhighered.com/sullivanstats) in various formats.
- Where applicable, exercises and examples incorporate output screens from various software including Minitab, the TI-83/84 Plus CE, Excel, and StatCrunch.
- Applets are included on the companion website and connected with certain activities from the Student Activity Workbook, allowing students to manipulate data and interact with animations.
- A technology manual is available that contains detailed tutorial instructions and worked out examples and exercises for the TI-83/84. There is also a new R Technology Manual should you choose to incorporate R into your class.

Companion Website Contents

The companion website is http://www.pearsonhighered.com/sullivanstats.

- Data Sets
- Applets
- Formula Cards and Tables in PDF format
- Additional Topics Folder including:
  - Sections 9.4, 9.5, 10.2A, 10.2B, 10.3A, 10.5, 11.1A, 11.2A, 11.3A, 11.5, 12.3A
  - Appendix A
  - Appendix B
    - Lines
    - Inference about Two Proportions (Dependent Samples)
    - Comparing Three or More Means (One-Way Analysis of Variance)
- **Case Studies** for each chapter in the text.
- A copy of the questions asked on the Sullivan Statistics Survey I and Survey II
- Consumer Reports projects that were formerly in the text
- The author has also created a website at https://www.sullystats.com. This site has chapter-by-chapter suggestions for teaching the material, links to interesting data sets, and much more.

Key Chapter Content Changes

Chapter 1 Data Collection

Section 1.2 now includes a discussion of obtaining data through web scraping and how to obtain data from the Internet. Section 1.6 expands on the discussion of the placebo effect.

Chapter 5

Section 5.1 now distinguishes the Law of Large Numbers from the nonexistent Law of Averages. There is a new Section 5.6 on simulating probability experiments. This material is very helpful in allowing students to see the role of randomness in probability experiments. It also foreshadows topics such as sampling distributions and inference.

Chapter 9

There is an expanded discussion on the normality condition for constructing confidence intervals for the population mean using Student's *t*-distribution in Section 9.2.

Chapter 10

Chapter 10 now contains optional sections on simulation methods for conducting inference. The organization of Chapter 10 allows for presenting simulation along with traditional inference, or simply presenting traditional inference. Should you decide to present only the traditional approach to inference, simply cover Section 10.2 from the text. If you decide to present hypothesis testing using simulation, skip Section 10.2 in the text and cover Sections 10.2A and 10.2B (available in MyLab or the companion website as pdfs). Section 10.3A (MyLab) presents hypothesis testing on a mean using simulation and bootstrapping. This section is optional and may be skipped without loss of continuity.

Chapter 11

Chapter 11 has new optional sections on randomization methods. Section 11.1A (available in MyLab or the companion website as a pdf) presents randomization tests for two independent proportions. If you choose to present randomization methods, we recommend presenting Section 11.1A prior to Section 11.1. Section 11.2A (MyLab) presents hypothesis tests on dependent means using bootstrapping. This section is optional and may be skipped without loss of continuity. Section 11.3A (MyLab) presents randomization tests for two independent means. We recommend covering this section prior to Section 11.3, if you choose to discuss this approach to hypothesis testing.

Chapter 12

Chapter 12 has a new optional section on randomization. Section 12.3A (available in MyLab or the companion website as a pdf) presents randomization tests for the slope of the least-squares regression model. If you choose to cover this section, do so prior to Section 12.3.

Flexible to Work with Your Syllabus

To meet the varied needs of diverse syllabi, this book has been organized to be flexible.

You will notice the "Preparing for This Section" material at the beginning of each section, which will tip you off to dependencies within the course. The two most common variations within an introductory statistics course are the treatment of regression analysis and the treatment of probability.

- **Coverage of Correlation and Regression** The text was written with the descriptive portion of bivariate data (Chapter 4) presented after the descriptive portion of univariate data (Chapter 3). Instructors who prefer to postpone the discussion of bivariate data can skip Chapter 4 and return to it before covering Sections 12.3 and 12.4.
- **Coverage of Probability** The text allows for light to extensive coverage of probability. Instructors wishing to minimize probability may cover Section 5.1 and skip the remaining sections. A mid-level treatment of probability can be accomplished by covering Sections 5.1 through 5.3. Instructors who will cover the chi-square test for independence will want to cover Sections 5.1 through 5.3. In addition, an instructor who will cover binomial probabilities will want to cover independence in Section 5.3 and combinations in Section 5.5.

Acknowledgments

Textbooks evolve into their final form through the efforts and contributions of many people. First and foremost, I would like to thank my family, whose dedication to this project was just as much as mine: my wife, Yolanda, whose words of encouragement and support were unabashed, and my children, Michael, Kevin, and Marissa, who have been supportive throughout their childhood and now into adulthood (my how time flies). I owe each of them my sincerest gratitude. I would also like to thank the entire Mathematics Department at Joliet Junior College and my colleagues who provided support, ideas, and encouragement to help me complete this project. From Pearson Education: I thank Suzanna Bainbridge, whose editorial expertise has been an invaluable asset; Peggy McMahon, who provided organizational skills that made this project go smoothly; Alicia Wilson and Demetrius Hall, for their marketing savvy and dedication to getting the word out; Vicki Dreyfus and Jean Choe, for their dedication in organizing all the media; Rose Kernan for her ability to control the production process; Dana Bettez for her editorial skill with the Instructor's Resource Guide and Student Activity Workbook; and the Pearson sales team, for their confidence and support of this book.

I also want to thank Ryan Cromar, Susan Herring, Craig Johnson, Kathleen McLaughlin, Patrick Murphy, Heidi Lyne, and Dorothy Wakefield for their help in creating supplements. A big thank-you goes to Cindy Trimble and Associates, who assisted in verifying answers for the back of the text and helped in proofreading. I would also like to acknowledge Kathleen Almy and Heather Foes for their help and expertise in developing the Student Activity Workbook. Finally, I would like to thank George Woodbury for helping me with the incredible suite of videos that accompanies the text. Many thanks to all the reviewers, whose insights and ideas form the backbone of this text. I apologize for any omissions.

CALIFORNIA *Charles Biles, Humboldt State University • Carol Curtis, Fresno City College • Jacqueline Faris, Modesto Junior College • Freida Ganter, California State University–Fresno • Jessica Kramer, Santiago Canyon College • Sherry Lohse, Napa Valley College • Craig Nance, Santiago Canyon College • Diane Van Deusen, Napa Valley College* **COLORADO** *Roxanne Byrne, University of Colorado–Denver • Monica Geist, Front Range Community College* **CONNECTICUT** *Kathleen McLaughlin, Manchester Community College • Dorothy Wakefield, University of Connecticut • Cathleen M. Zucco Teveloff, Trinity College* **DISTRICT OF COLUMBIA** *Monica Jackson, American University • Jill McGowan, Howard University* **FLORIDA** *Randall Allbritton, Daytona Beach Community College • Greg Bloxom, Pensacola State College • Anthony DePass, St. Petersburgh College Clearwater • Kelcey Ellis, University of Central Florida • Franco Fedele, University of West Florida • Laura Heath, Palm Beach Community College • Perrian Herring, Okaloosa Walton College • Marilyn Hixson, Brevard Community College • Daniel Inghram, University of Central Florida • Philip Pina, Florida Atlantic University • Mike Rosenthal, Florida International University • James Smart, Tallahassee Community College* **GEORGIA** *Virginia Parks, Georgia Perimeter College • Chandler Pike, University of Georgia • Jill Smith, University of Georgia • John Weber, Georgia Perimeter College* **HAWAII** *Eric Matsuoka at Leeward Community College • Leslie Rush, University of Hawaii* **IDAHO** *K. Shane Goodwin, Brigham Young University • Craig Johnson, Brigham Young University • Brent Timothy, Brigham Young University • Kirk Trigsted, University of Idaho* **ILLINOIS** *Grant Alexander, Joliet Junior College • Kathleen Almy, Rock Valley College • John Bialas, Joliet Junior College • Linda Blanco, Joliet Junior College • Kevin Bodden, Lewis & Clark Community College • Rebecca Goad, Joliet Junior College • Joanne Brunner, Joliet Junior College • Robert Capetta, University of Illinois at Chicago • Elena Catoiu, Joliet Junior College • Faye Dang, Joliet Junior College • Laura Egner, Joliet Junior College • Jason Eltrevoog, Joliet Junior College • Erica Egizio, Lewis University • Heather Foes, Rock Valley College • Randy Gallaher, Lewis & Clark Community College • Melissa Gaddini, Robert Morris University • Iraj Kalantari, Western Illinois University • Donna Katula, Joliet Junior College • Diane Koenig, Rock Valley College • Diane Long, College of DuPage • Heidi Lyne, Joliet Junior College • Jean McArthur, Joliet Junior College • Patricia McCarthy, Robert Morris University • David McGuire, Joliet Junior College • Angela McNulty, Joliet Junior College • James Morgan, Joliet Junior College • Andrew Neath, Southern Illinois University-Edwardsville • Linda Padilla, Joliet Junior College • David Ruffato, Joliet Junior College • Patrick Stevens, Joliet Junior College • Robert Tuskey, Joliet Junior College • Stephen Zuro, Joliet Junior College* **INDIANA** *Susitha Karunaratne, Purdue University North Central • Jason Parcon, Indiana University–Purdue University Ft. Wayne • Henry Wakhungu, Indiana University* **KANSAS** *Donna Gorton, Butler Community College • Ingrid Peterson, University of Kansas* **LOUISIANA** *Melissa Myers, University of Louisiana at Lafayette* **MARYLAND** *Nancy Chell, Anne Arundel Community College • John Climent, Cecil Community College • Rita Kolb, The Community College of Baltimore County • Jignasa Rami, Community College of Baltimore County • Mary Lou Townsend, Wor-Wic Community College* **MASSACHUSETTS** *Susan McCourt, Bristol Community College • Daniel Weiner, Boston University • Pradipta Seal, Boston University of Public Health* **MICHIGAN** *Margaret M. Balachowski, Michigan Technological University • Diane Krasnewich, Muskegon Community College • Susan Lenker, Central Michigan University • Timothy D. Stebbins, Kalamazoo Valley Community College • Sharon Stokero, Michigan Technological University • Alana Tuckey, Jackson Community College* **MINNESOTA** *Mezbhur Rahman, Minnesota State University* **MISSOURI** *Farroll Tim*

Wright, University of Missouri–Columbia **NEBRASKA** *Jane Keller, Metropolitan Community College* **NEW YORK** *Jacob Amidon, Finger Lakes Community College • Stella Aminova, Hunter College • Jennifer Bergamo, Onondaga Community College • Kathleen Cantone, Onondaga Community College • Pinyuen Chen, Syracuse University • Sandra Clarkson, Hunter College of CUNY • Rebecca Daggar, Rochester Institute of Technology • Bryan Ingham, Finger Lakes Community College • Anne M. Jowsey, Niagara County Community College • Maryann E. Justinger, Erie Community College–South Campus • Bernadette Lanciaux, Rochester Institute of Technology • Kathleen Miranda, SUNY at Old Westbury • Robert Sackett, Erie Community College–North Campus • Sean Simpson, Westchester Community College • Bill Williams, Hunter College of CUNY* **NORTH CAROLINA** *Fusan Akman, Coastal Carolina Community College • Mohammad Kazemi, University of North Carolina–Charlotte • Janet Mays, Elon University • Marilyn McCollum, North Carolina State University • Claudia McKenzie, Central Piedmont Community College • Said E. Said, East Carolina University • Karen Spike, University of North Carolina–Wilmington • Jeanette Szwec, Cape Fear Community College* **NORTH DAKOTA** *Myron Berg, Dickinson State University • Ronald Degges, North Dakota State University* **OHIO** *Richard Einsporn, The University of Akron • Michael McCraith, Cuyaghoga Community College* **OREGON** *Daniel Kim, Southern Oregon University • Jong Sung Kin, Portland State University* **PENNSYLVANIA** *Mary Brown, Harrisburg Area Community College • LeAnne Conaway, Harrisburg Area Community College* **SOUTH CAROLINA** *Diana Asmus, Greenville Technical College • Dr. William P. Fox, Francis Marion University • Cheryl Hawkins, Greenville Technical College • Rose Jenkins, Midlands Technical College • Lindsay Packer, College of Charleston • Laura Shick, Clemson University • Erwin Walker, Clemson University* **TENNESSEE** *Tim Britt, Jackson State Community College • Nancy Pevey, Pellissippi State Technical Community College • David Ray, University of Tennessee–Martin* **TEXAS** *Edith Aguirre, El Paso Community College • Ivette Chuca, El Paso Community College • Aaron Gutknecht, Tarrant County College • Jada Hill, Richland College • David Lane, Rice University • Alma F. Lopez, South Plains College • Shanna Moody, University of Texas at Arlington • Jasdeep Pannu, Lamar University* **UTAH** *Joe Gallegos, Salt Lake City Community College • Alia Maw, Salt Lake City Community College* **VIRGINIA** *Kim Jones, Virginia Commonwealth University • Vasanth Solomon, Old Dominion University* **WEST VIRGINIA** *Mike Mays, West Virginia University* **WISCONSIN** *William Applebaugh, University of Wisconsin–Eau Claire • Carolyn Chapel, Western Wisconsin Technical College • Beverly Dretzke, University of Wisconsin–Eau Claire • Jolene Hartwick, Western Wisconsin Technical College • Thomas Pomykalski, Madison Area Technical College • Walter Reid, University of Wisconsin-Eau Claire*

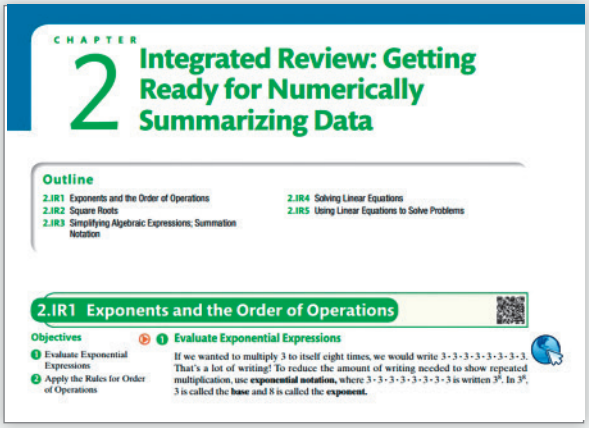
Michael Sullivan, III
Joliet Junior College

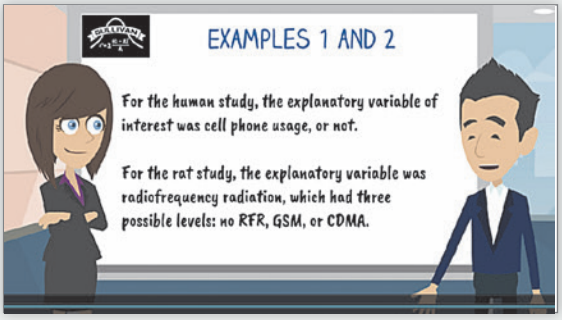
MyLab Statistics is available to accompany Pearson's market-leading text options, including Fundamentals of Statistics, 6e by Michael Sullivan (access code required).

MyLab™ is the teaching and learning platform that empowers you to reach every student. MyLab Statistics combines trusted author content—including full eText and assessment with immediate feedback—with digital tools and a flexible platform to personalize the learning experience and improve results for each student. Integrated with StatCrunch®, a web-based statistical software program, students learn the skills they need to interact with data in the real world.

Integrated Review for Corequisite Courses

This MyLab™ includes an additional eText written by the author on prerequisite skills and concepts. There are also prebuilt (and editable) MyLab quizzes that populate personalized homework assignments for gaps in skills for that chapter. These resources may be used in a corequisite course model, or simply to help underprepared students master prerequisite skills and concepts.





NEW! Videos

In addition to existing Author in the Classroom, StatTalk, and Example Videos, the following **video types** were added to this edition.

- **Innovative lightboard videos** featuring Mike Sullivan guide students towards deeper conceptual understanding of certain key topics.
- **Excel video solutions** were added to the existing suite of StatCrunch, TI-83/84 Plus, and by-hand videos for examples in the text.
- **Animation videos** remind students of where they have been and where they are going in the upcoming objectives.

pearson.com/mylab/statistics

Resources for Success

Pearson MyLab

Student Resources

Each student learns at a different pace. Personalized learning pinpoints the precise areas where each student needs practice, giving all students the support they need — when and where they need it — to be successful.

StatCrunch

StatCrunch® is powerful web-based statistical software that allows users to collect, crunch, and communicate with data. The vibrant online community offers tens of thousands of shared data sets for students and instructors to analyze, in addition to all of the data sets in the text or online homework. StatCrunch is integrated directly into MyLab Statistics or it can be purchased separately. Learn more at www.statcrunch.com.

Data Sets

All data sets from the textbook are available in MyLab Statistics. They can be analyzed in StatCrunch or downloaded for use in other statistical software programs.

Statistical Software Support

Instructors and students can copy data sets from the text and MyLab Statistics exercises directly into software such as StatCrunch or Excel®. Students can also access instructional support tools including tutorial videos, Study Cards, and manuals for a variety of statistical software programs including StatCrunch, Excel, Minitab®, JMP®, R, SPSS, and TI 83/84 calculators.

Student Solutions Manual

This manual provides detailed, worked-out solutions to all odd-numbered text exercises, as well as all solutions for the Chapter Reviews and Chapter Tests. It is available in print and can be downloaded from MyLab Statistics. (ISBN-13: 9780136969761)

Instructor Resources

Your course is unique. So whether you'd like to build your own assignments, teach multiple sections, or set prerequisites, MyLab gives you the flexibility to easily create your course to fit your needs.

Annotated Instructor's Edition

Includes answers to all text exercises, as well as teaching tips and common student errors. (ISBN-13: 9780136807315; ISBN-10:0136807313)

Instructor Solutions Manual

Contains worked-out solutions to all text exercises and case study answers. It can be downloaded from MyLab Statistics or from www.pearson.com.

PowerPoint Lecture Slides

PowerPoint Lecture Slides include key graphics and follow the sequence and philosophy of the text. They can be downloaded from MyLab Statistics or from www.pearson.com.

TestGen

TestGen® (www.pearson.com/testgen) enables instructors to build, edit, print, and administer tests using a computerized bank of questions developed to cover all the objectives of the text. TestGen is algorithmically based, allowing instructors to create multiple but equivalent versions of the same question or test with the click of a button. Instructors can also modify test bank questions or add new questions. The software and test bank are available for download from Pearson's online catalog, www.pearson.com. The questions are also assignable in MyLab Statistics.

Data Analytics / Early Alerts

Instructors have a comprehensive gradebook with enhanced reporting functionality that makes it easier to understand which students are struggling, and which topics they struggle with most. **New Early Alerts** use predictive analytics to identify struggling students as early as possible—even if their assignment scores are not a cause for concern.

Resources for Success

Pearson MyLab

Instructor Resources (continued)

Question Libraries

MyLab Statistics includes a number of question libraries that instructors can incorporate into their regular assignments. StatCrunch Projects consist of questions about large data sets in StatCrunch. The Conceptual Question Library offers 1,000 questions to help students learn concepts and how to think critically. Finally, the StatTalk Video Library includes questions associated to the video series by statistician Andrew Vickers.

Minitab and Minitab Express™

Bundling Minitab software with educational materials ensures students have access to the software they need in the classroom, around campus, and at home. And having 12-month access to Minitab and Minitab Express ensures students can use the software for the duration of their course. (ISBN-13: 9780134456409; ISBN-10: 0134456408) (access card only; not sold as stand alone.)

JMP Student Edition

An easy-to-use, streamlined version of JMP desktop statistical discovery software from SAS Institute, Inc. is available for bundling with the text. (ISBN-13: 9780134679792; ISBN-10: 0134679792)

XLSTAT™

An Excel add-in that enhances the analytical capabilities of Excel. XLSTAT is used by leading businesses and universities around the world. It is available to bundle with this text. For more information go to www.pearsonhighered.com/xlstatupdate. (ISBN-13: 9780321759320; ISBN-10: 032175932X)

Accessibility

Pearson works continuously to ensure our products are as accessible as possible to all students. We are working toward achieving WCAG 2.0 Level AA and Section 508 standards, as expressed in the Pearson Guidelines for Accessible Educational Web Media, www.pearson.com/mylab/statistics/accessibility.

Applications Index

Sports

Statistics

Temperature

Test(s)

Time

Transportation. *See also* Motor vehicle(s)

Travel

Weather

Weight(s)

Work. *See also* Business

PART 1

Getting the Information You Need

Statistics is a process—a series of steps that leads to a goal. This text is divided into four parts to help see the process of statistics.

Part 1 is focused on the first step in the process, which is to determine the research objective or question to be answered. Then information is obtained to answer the questions stated in the research objective.

Data Collection

Outline

Making an Informed Decision

It is your senior year of high school. You will have a lot of exciting experiences in the upcoming year, plus a major decision to make—which college should I attend? The choice you make may affect many aspects of your life—your career, where you live, your significant other, and so on, so you don't want to simply choose the college that everyone else picks. You need to design a questionnaire to help you make an informed decision about college. In addition, you want to know how well the college you are considering educates its students. See Making an Informed Decision on page 60.

Putting It Together

Statistics plays a major role in many aspects of our lives. It is used in sports, for example, to help a general manager decide which player might be the best fit for a team. It is used in politics to help candidates understand how the public feels about various policies. And statistics is used in medicine to help determine the effectiveness of new drugs.

Used appropriately, statistics can enhance our understanding of the world around us. Used inappropriately, it can lend support to inaccurate beliefs. Understanding statistical methods will provide you with the ability to analyze and critique studies and the opportunity to become an informed consumer of information. Understanding statistical methods will also enable you to distinguish solid analysis from bogus "facts."

To help you understand the features of this text and for hints to help you study, read the ***Pathway to Success*** on the front inside cover of the text.

1.1 Introduction to the Practice of Statistics

Objectives

1. Define statistics and statistical thinking
2. Explain the process of statistics
3. Distinguish between qualitative and quantitative variables
4. Distinguish between discrete and continuous variables
5. Determine the level of measurement of a variable

1 Define Statistics and Statistical Thinking

What is statistics? Many people say that statistics is numbers. After all, we are bombarded by numbers that supposedly represent how we feel and who we are. For example, we hear on the radio that 50% of first marriages, 67% of second marriages, and 74% of third marriages end in divorce (Forest Institute of Professional Psychology, Springfield, MO).

Another interesting consideration about the "facts" we hear or read is that two different sources can report two different results. For example, a January 12, 2019 poll by Rasmussen Reports indicated that 45% of Americans believed the country was on the right track. However, a January 16, 2019 Monmouth poll indicated that 37% of Americans believed the country was on the right track. Is it possible that the percent of Americans who believe the country is on the right track could decrease by 8% in four days, or is something else going on? Statistics helps to provide the answer.

Certainly, statistics has a lot to do with numbers, but this definition is only partially correct. Statistics is also about where the numbers come from (that is, how they were obtained) and how closely the numbers reflect reality.

Definition

Statistics is the science of collecting, organizing, summarizing, and analyzing information to draw conclusions or answer questions. In addition, statistics is about providing a measure of confidence in any conclusions.

Let's break this definition into four parts. The first part states that statistics involves the collection of information. The second refers to the organization and summarization of information. The third states that the information is analyzed to draw conclusions or answer specific questions. The fourth part states that results should be reported using some measure that represents how convinced we are that our conclusions reflect reality.

What is the information referred to in the definition? The information is **data**, which the *American Heritage Dictionary* defines as "a fact or proposition used to draw a conclusion or make a decision." Data can be numerical, as in height, or nonnumerical, as in gender. In either case, data describe characteristics of an individual.

IN OTHER WORDS

Anecdotal means that the information being conveyed is based on casual observation, not scientific research.

Analysis of data can lead to powerful results. Data can be used to offset anecdotal claims, such as the suggestion that cellular telephones cause brain cancer. After carefully collecting, summarizing, and analyzing data regarding this phenomenon, it was determined that there is no link between cell phone usage and brain cancer. See Examples 1 and 2 in Section 1.2.

Because data are powerful, they can be dangerous when misused. The misuse of data usually occurs when data are incorrectly obtained or analyzed. For example, radio or television talk shows regularly ask poll questions for which respondents must call in or use the Internet to supply their vote. Most likely, the individuals who are going to call in are those who have a strong opinion about the topic. This group is not likely to be representative of people in general, so the results of the poll are not meaningful. Whenever we look at data, we should be mindful of where the data come from.

Even when data tell us that a relation exists, we need to investigate. For example, a study showed that breast-fed children have higher IQs than those who were not breast-fed. Does this study mean that a mother who breast-feeds her child will increase the child's IQ? Not necessarily. It may be that some factor other than breast-feeding contributes to the IQ of the children. In this case, it turns out that mothers who breast-feed generally have higher IQs than those who do not. Therefore, it may be genetics that leads to the higher IQ, not breast-feeding.* This illustrates an idea in statistics known as the *lurking variable*. A good statistical study will have a way of dealing with lurking variables.

A key aspect of data is that they vary. Consider the students in your classroom. Is everyone the same height? No. Does everyone have the same color hair? No. So, within groups there is variation. Now consider yourself. Do you eat the same amount of food each day? No. Do you sleep the same number of hours each day? No. So even considering an individual there is variation. Data vary. One goal of statistics is to describe and understand the sources of variation. Variability in data may help to explain the different results obtained by the Rasmussen Reports and Monmouth polls described at the beginning of this section.

Because of this variability, the results that we obtain using data can vary. In a mathematics class, if Bob and Jane are asked to solve $3x + 5 = 11$, they will both obtain $x = 2$ as the solution when they use the correct procedures. In a statistics class, if Bob and Jane are asked to estimate the average commute time for workers in Dallas, Texas, they will likely get different answers, even though both use the correct procedure. The different answers occur because they likely surveyed different individuals, and these individuals have different commute times. Bob and Jane would get the same result if they both asked *all* commuters or the same commuters about their commutes, but how likely is this?

So, in mathematics when a problem is solved correctly, the results can be reported with 100% certainty. In statistics, when a problem is solved, the results do not have 100% certainty. In statistics, we might say that we are 95% confident that the average commute time in Dallas, Texas, is between 20 and 23 minutes. Uncertain results may seem disturbing now but will feel more comfortable as we proceed through the course.

Without certainty, how can statistics be useful? Statistics can provide an understanding of the world around us because recognizing where variability in data comes from can help us to control it. Understanding the techniques presented in this text will provide you with powerful tools that will give you the ability to analyze and critique media reports, make investment decisions, or conduct research on major purchases. This will help to make you an informed citizen, consumer of information, and critical and statistical thinker.

2 Explain the Process of Statistics

Consider the following scenario.

> You are walking down the street and notice that a person walking in front of you drops $100. Nobody seems to notice the $100 except you. Since you could keep the money without anyone knowing, would you keep the money or return it to the owner?

NOTE

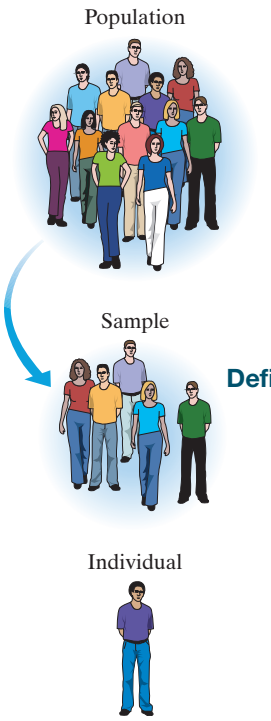
Obtaining a truthful response to a question such as this is challenging. In Section 1.5, we present some techniques for obtaining truthful responses to sensitive questions.

Suppose you wanted to use this scenario as a gauge of the morality of students at your school by determining the percent of students who would return the money. How might you do this? You could attempt to present the scenario to every student at the school, but this would be difficult or impossible if the student body is large. A second possibility is to present the scenario to 50 students and use the results to make a statement about all the students at the school.

*In fact, a study found that a gene called FADS2 is responsible for higher IQ scores in breast-fed babies. *Source:* Duke University, "Breastfeeding Boosts IQ in Infants with 'Helpful' Genetic Variant," *Science Daily* 6 November 2007.

Definitions The entire group to be studied is called the **population**. An **individual** is a person or object that is a member of the population being studied. A **sample** is a subset of the population that is being studied. See Figure 1.

Figure 1



In the $100 study presented, the population is all the students at the school. Each student is an individual. The sample is the 50 students selected to participate in the study.

Suppose 39 of the 50 students stated that they would return the money to the owner. We could present this result by saying that the percent of students in the survey who would return the money to the owner is 78%. This is an example of a *descriptive statistic* because it describes the results of the sample without making any general conclusions about the population.

Definitions A **statistic** is a numerical summary of a sample. **Descriptive statistics** consist of organizing and summarizing data. Descriptive statistics describe data through numerical summaries, tables, and graphs.

So 78% is a statistic because it is a numerical summary based on a sample. Descriptive statistics make it easier to get an overview of what the data are telling us.

If we extend the results of our sample to the population, we are performing *inferential statistics*.

Definition **Inferential statistics** uses methods that take a result from a sample, extend it to the population, and measure the reliability of the result.

The generalization contains uncertainty because a sample cannot tell us everything about a population. Therefore, inferential statistics includes a level of confidence in the results. So rather than saying that 78% of all students would return the money, we might say that we are 95% confident that between 74% and 82% of all students would return the money. Notice how this inferential statement includes a *level of confidence* (measure of reliability) in our results. It also includes a range of values to account for the variability in our results.

One goal of inferential statistics is to use statistics to estimate *parameters*.

Definition A **parameter** is a numerical summary of a population.

EXAMPLE 1 Parameter versus Statistic

(a) Suppose 48.2% of all students on your campus own a car. This value represents a parameter because it is a numerical summary of a population. Suppose a sample of 100 students is obtained, and from this sample we find that 46% own a car. This value represents a statistic because it is a numerical summary of a sample.

(b) Suppose the average salary of all employees in the City of Joliet is $78,302. This value represents a parameter because it is a numerical summary of a population. Suppose a sample of 30 employees is obtained, and from this sample we find the average salary is $75,038. This value represents a statistic because it is a numerical summary of a sample.

NW **Now Work Problem 3**

The methods of statistics follow a process.

CAUTION!
Many nonscientific studies are based on *convenience samples*, such as Internet surveys or phone-in polls. The results of any study performed using this type of sampling method are not reliable.

The Process of Statistics

1. *Identify the research objective.* A researcher must determine the question(s) he or she wants answered. The question(s) must clearly identify the population that is to be studied.
2. *Collect the data needed to answer the question(s) posed in (1).* Conducting research on an entire population is often difficult and expensive, so we typically look at a sample. This step is vital to the statistical process, because if the data are not collected correctly, the conclusions drawn are meaningless. Do not overlook the importance of appropriate data collection. We discuss this step in detail in Sections 1.2 through 1.6.
3. *Describe the data.* Descriptive statistics allow the researcher to obtain an overview of the data and can help determine the type of statistical methods the researcher should use. We discuss this step in detail in Chapters 2 through 4.
4. *Perform inference.* Apply the appropriate techniques to extend the results obtained from the sample to the population and report a level of reliability of the results. We discuss techniques for measuring reliability in Chapters 5 through 8 and inferential techniques in Chapters 9 through 15.

EXAMPLE 2 The Process of Statistics: Trust Your Neighbor

Pew Research conducted a poll and asked, "Do you trust all or most of your neighbors?" The following process allowed the researchers to conduct their study.

1. *Identify the Research Objective* The researchers wanted to determine the percentage of adult Americans who trust all or most of their neighbors. Therefore, the population was adult Americans.
2. *Collect the Data Needed to Answer the Question Posed in (1)* It is unreasonable to expect to survey the more than 200 million adult Americans to determine whether they trust all or most of their neighbors. So, the researchers surveyed a sample of 1628 adult Americans. Of those surveyed, 847 stated they trust all or most of their neighbors.
3. *Describe the Data* Of the 1628 individuals in the survey, 52% $(= 847/1628)$ stated they trust all or most of their neighbors. This is a descriptive statistic because it is a summary of the sample data.
4. *Perform Inference* Pew Research wanted to extend the results of the survey to all adult Americans. When generalizing results from a sample to a population, the results are uncertain. To account for this uncertainty, Pew reported a 2.5% *margin of error.* This means Pew feels fairly certain (in fact, 95% certain) that the percentage of *all* adult Americans who trust all or most of their neighbors is somewhere between 49.5% $(= 52\% - 2.5\%)$ and 54.5% $(= 52\% + 2.5\%)$.

NW **Now Work Problem 45**

3 Distinguish between Qualitative and Quantitative Variables

Once a research objective is stated, a list of the information we want to learn about the individuals must be created. **Variables** are the characteristics of the individuals within the population. For example, recently my son and I planted a tomato plant in our backyard. We collected information about the tomatoes harvested from the plant. The individuals we studied were the tomatoes. The variable that interested us was the weight of a tomato. My son noted that the tomatoes had different weights even though they came from the same plant. He discovered that variables such as weight may vary.

If variables did not vary, they would be constants, and statistical inference would not be necessary. Think about it this way: If each tomato had the same weight, then knowing the weight of one tomato would allow us to determine the weights of all tomatoes. However, the weights of the tomatoes vary. One goal of research is to learn the causes of the variability so that we can learn to grow plants that yield the best tomatoes.

Variables can be classified into two groups: *qualitative* or *quantitative*.

Definitions

Qualitative, or **categorical, variables** allow for classification of individuals based on some attribute or characteristic.

Quantitative variables provide numerical measures of individuals. The values of a quantitative variable can be added or subtracted and provide meaningful results.

Many examples in this text will include a suggested **approach**, or a way to look at and organize a problem so that it can be solved. The approach will be a suggested method of *attack* toward solving the problem. This does not mean that the approach given is the only way to solve the problem, because many problems have more than one approach leading to a correct solution.

EXAMPLE 3 Distinguishing between Qualitative and Quantitative Variables

Problem Determine whether the following variables are qualitative or quantitative.

(a) Gender
(b) Temperature
(c) Number of days during the past week that a college student studied
(d) Zip code

Approach Quantitative variables are numeric measures such that meaningful arithmetic operations can be performed on the values of the variable. Qualitative variables describe an attribute or characteristic of the individual that allows researchers to categorize the individual.

Solution

(a) Gender is a qualitative variable because it allows a researcher to categorize the individual as male or female. Notice that arithmetic operations cannot be performed on these attributes.
(b) Temperature is a quantitative variable because it is numeric, and operations such as addition and subtraction provide meaningful results. For example, 70°F is 10°F warmer than 60°F.
(c) Number of days during the past week that a college student studied is a quantitative variable because it is numeric, and operations such as addition and subtraction provide meaningful results.
(d) Zip code is a qualitative variable because it categorizes a location. Notice that, even though zip codes are numeric, adding or subtracting zip codes does not provide meaningful results.

NW **Now Work Problem 11**

Example 3(d) shows us that a variable may be qualitative while having numeric values. Just because the value of a variable is numeric does not mean that the variable is quantitative.

4 Distinguish between Discrete and Continuous Variables

We can further classify quantitative variables into two types: *discrete* or *continuous*.

Definitions

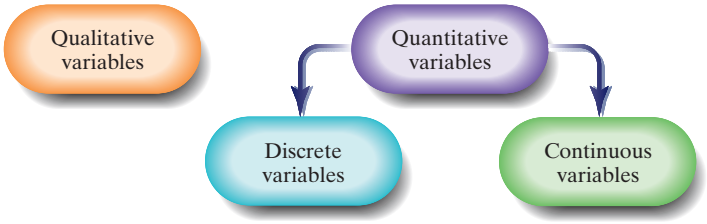
A **discrete variable** is a quantitative variable that has either a finite number of possible values or a countable number of possible values. The term *countable* means that the values result from counting, such as 0, 1, 2, 3, and so on. A discrete variable cannot take on every possible value between any two possible values.

A **continuous variable** is a quantitative variable that has an infinite number of possible values that are not countable. A continuous variable may take on every possible value between any two values.

IN OTHER WORDS
If you count to get the value of a quantitative variable, it is discrete. If you measure to get the value of a quantitative variable, it is continuous.

Figure 2 illustrates the relationship among qualitative, quantitative, discrete, and continuous variables.

Figure 2



EXAMPLE 4 Distinguishing between Discrete and Continuous Variables

Problem Determine whether the quantitative variables are discrete or continuous.

(a) The number of heads obtained after flipping a coin five times.

(b) The number of cars that arrive at a McDonald's drive-thru between 12:00 P.M. and 1:00 P.M.

(c) The distance a 2020 Toyota Prius can travel in city driving conditions with a full tank of gas.

Approach A variable is discrete if its value results from counting. A variable is continuous if its value is measured.

Solution

(a) The number of heads obtained by flipping a coin five times is a discrete variable because we can count the number of heads obtained. The possible values of this discrete variable are 0, 1, 2, 3, 4, 5.

(b) The number of cars that arrive at a McDonald's drive-thru between 12:00 P.M. and 1:00 P.M. is a discrete variable because we find its value by counting the cars. The possible values of this discrete variable are 0, 1, 2, 3, 4, and so on. Notice that this number has no upper limit.

(c) The distance traveled is a continuous variable because we measure the distance (miles, feet, inches, and so on).

NW **Now Work Problem 19**

Continuous variables are often rounded. For example, if a certain make of car gets 24 miles per gallon (mpg) of gasoline, its miles per gallon must be greater than or equal to 23.5 and less than 24.5, or $23.5 \leq \text{mpg} < 24.5$.

The type of variable (qualitative, discrete, or continuous) dictates the methods that can be used to analyze the data.

The list of observed values for a variable is **data**. Gender is a variable; the observations male and female are data. **Qualitative data** are observations corresponding to a qualitative variable. **Quantitative data** are observations corresponding to a quantitative variable. **Discrete data** are observations corresponding to a discrete variable. **Continuous data** are observations corresponding to a continuous variable.

EXAMPLE 5 Distinguishing between Variables and Data

Problem Table 1 presents a group of selected countries and information regarding these countries as of July, 2020. Identify the individuals, variables, and data in Table 1.

Table 1

Country	Government Type	Life Expectancy (years)	Population (in millions)
Australia	Federal parliamentary democracy	82.7	25.5
Canada	Constitutional monarchy	83.4	37.7
France	Republic	82.2	67.8
Morocco	Constitutional monarchy	73.3	35.6
Poland	Republic	78.3	38.3
Senegal	Presidential republic	63.2	15.7
United States	Federal republic	80.3	332.6

Source: CIA *World Factbook*

Approach An individual is an object or person for whom we wish to obtain data. The variables are the characteristics of the individuals, and the data are the specific values of the variables.

Solution The individuals in the study are the countries: Australia, Canada, and so on. The variables measured for each country are *government type*, *life expectancy*, and *population*. The variable *government type* is qualitative because it categorizes the individual. The variables *life expectancy* and *population* are quantitative.

The quantitative variable *life expectancy* is continuous because it is measured. The quantitative variable *population* is discrete because we count people. The observations are the data. For example, the data corresponding to the variable *life expectancy* are 82.7, 83.4, 82.2, 73.3, 78.3, 63.2, and 80.3. The following data correspond to the individual Poland: a republic government with residents whose life expectancy is 78.3 years and population is 38.3 million people. Republic is an instance of qualitative data that results from observing the value of the qualitative variable *government type*. The life expectancy of 78.3 years is an instance of quantitative data that results from observing the value of the quantitative variable *life expectancy*.

NW **Now Work Problem 41**

5 Determine the Level of Measurement of a Variable

Rather than classify a variable as qualitative or quantitative, we can assign a level of measurement to the variable.

Definitions

A variable is at the **nominal level of measurement** if the values of the variable name, label, or categorize. In addition, the naming scheme does not allow for the values of the variable to be arranged in a ranked or specific order.

A variable is at the **ordinal level of measurement** if it has the properties of the nominal level of measurement, however, the naming scheme allows for the values of the variable to be arranged in a ranked or specific order.

IN OTHER WORDS
The word ***nominal*** comes from the Latin word **nomen**, which means to name. When you see the word **ordinal**, think order.

(continued)