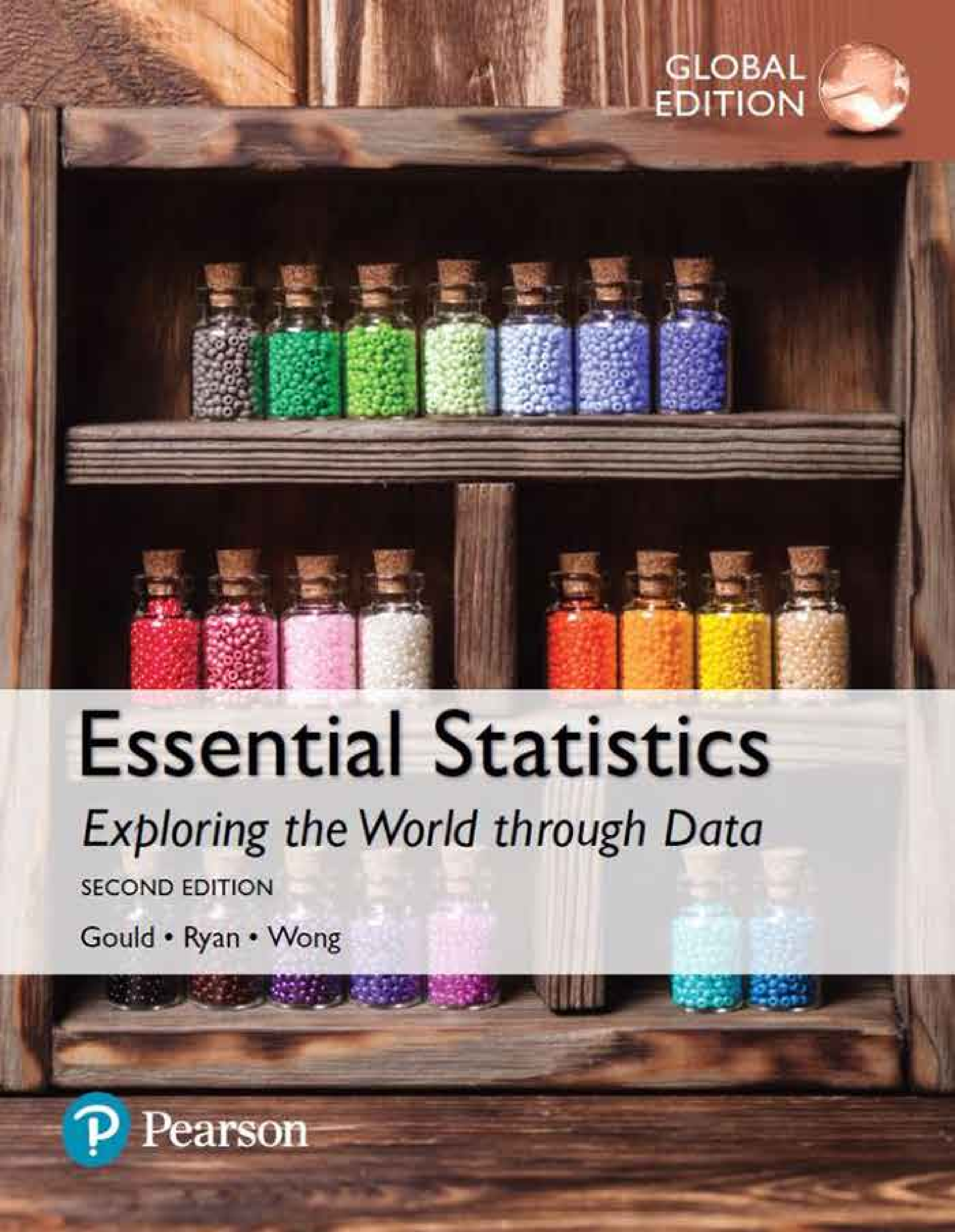


GLOBAL EDITION

Essential Statistics

Exploring the World through Data

SECOND EDITION

Gould • Ryan • Wong

Pearson

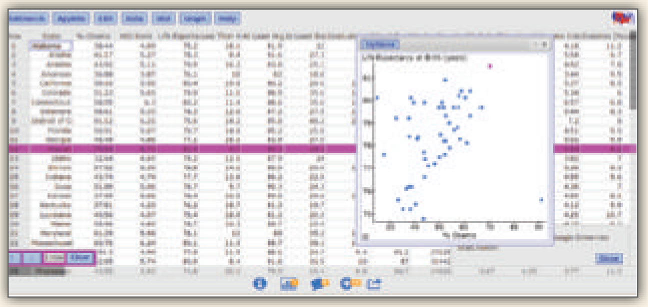
BREAK THROUGH
To improving results

Available in MyStatLab™ for your Introductory Statistics Courses

MyStatLab is the market-leading online resource for learning and teaching statistics.

Leverage the Power of StatCrunch

MyStatLab leverages the power of StatCrunch—powerful, web-based statistics software. Integrated into MyStatLab, students can easily analyze data from their exercises and etext. In addition, access to the full online community allows users to take advantage of a wide variety of resources and applications at www.statcrunch.com.

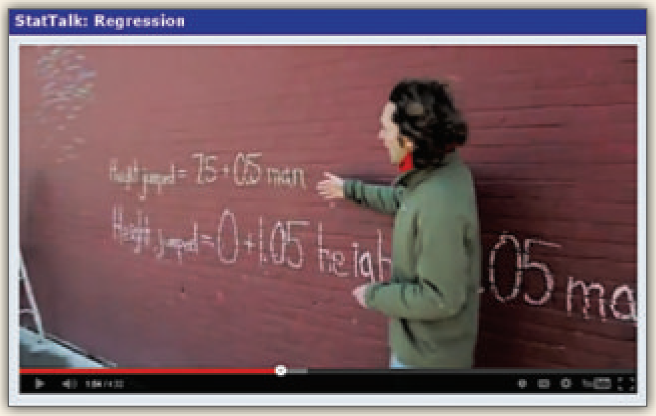


Bring Statistics to Life

Virtually flip coins, roll dice, draw cards, and interact with animations on your mobile device with the extensive menu of experiments and applets in StatCrunch. Offering a number of ways to practice resampling procedures, such as permutation tests and bootstrap confidence intervals, StatCrunch is a complete and modern solution.

Real-World Statistics

MyStatLab video resources help foster conceptual understanding. StatTalk Videos, hosted by fun-loving statistician Andrew Vickers, demonstrate important statistical concepts through interesting stories and real-life events. This series of 24 videos includes assignable questions built in MyStatLab and an instructor's guide.



www.mystatlab.com

This page intentionally left blank

Essential Statistics: Exploring the World through Data

Second Edition

Global Edition

Robert Gould

University of California, Los Angeles

Colleen Ryan

California Lutheran University

Rebecca Wong

West Valley College

Pearson

Boston Columbus Indianapolis New York San Francisco Amsterdam
Cape Town Dubai London Madrid Milan Munich Paris Montreal Toronto Delhi
Mexico City Sao Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

Editor in Chief: Deirdre Lynch
Senior Acquisitions Editor: Suzanna Bainbridge
Editorial Assistant: Justin Billing
Field Marketing Manager: Andrew Noble
Product Marketing Manager: Tiffany Bitzel
Marketing Assistant: Jennifer Myers
Program Team Lead: Karen Wernholm
Program Manager: Chere Bemelmans
Project Team Lead: Peter Silvia
Project Manager: Peggy McMahon
Senior Author Support/Technology Specialist: Joe Vetere
Manager, Rights Management, Higher Education: Gina M. Cheselka
Media Producer: Aimee Thorne
Acquisitions Editor, Global Edition: Sourabh Maheshwari
Assistant Project Editor, Global Edition: Sulagna Dasgupta
Media Production Manager, Global Edition: Vikram Kumar
Senior Manufacturing Controller, Production, Global Edition: Trudy Kimber
Associate Director of Design, USHE EMSS/HSC/EDU: Andrea Nix
Program Design Lead: Beth Paquin
Design, Full-Service Project Management, Composition, and Illustration: Cenveo® Publisher Services
Senior Project Manager, MyStatLab: Robert Carroll
QA Manager, Assessment Content: Marty Wright
Procurement Manager: Carol Melville

Pearson Education Limited
Edinburgh Gate
Harlow
Essex CM20 2JE
England

and Associated Companies throughout the world

Visit us on the World Wide Web at:
www.pearsonglobaleditions.com

Authorized adaptation from the United States edition, entitled Essential Statistics, 2nd edition, *ISBN 978-0-134-13440-6, by Robert Gould, Colleen Ryan, and Rebecca Wong, published by Pearson Education © 2017.*

Acknowledgments of third-party content appear in Appendix D, which constitutes an extension of this copyright page.

TI-84+C screenshots courtesy of Texas Instruments. Data and screenshots from StatCrunch used by permission of StatCrunch. Screenshots from Minitab courtesy of Minitab Corporation. Screenshot from SOCR used by the permission of the Statistics Online Computational Resource, UCLA. XLSTAT screenshots courtesy of Addinsoft, Inc. Used with permission. All Rights Reserved. XLSTAT is a registered trademark of Addinsoft SARL.

ISBN 10: 1-292-16122-1
ISBN 13: 978-1-292-16122-8

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

10 9 8 7 6 5 4 3 2 1

Printed and bound by Vivar in Malaysia

Dedication

To my parents and family, my friends, and my colleagues who are also friends. Without their patience and support, this would not have been possible.

—Rob

To my teachers and students, and to my family who have helped me in many different ways.

—Colleen

To my students, colleagues, family, and friends who have helped me be a better teacher and a better person.

—Rebecca

About the Authors

Robert Gould

Robert L. Gould (Ph.D., University of California, San Diego) is a leader in the statistics education community. He has served as chair of the American Statistical Association's Committee on Teacher Enhancement, has served as chair of the ASA's Statistics Education Section, and served on a panel of co-authors for the *Guidelines for Assessment in Instruction on Statistics Education (GAISE) College Report*. While serving as the associate director of professional development for CAUSE (Consortium for the Advancement of Undergraduate Statistics Education), Rob worked closely with the American Mathematical Association of Two-Year Colleges (AMATYC) to provide traveling workshops and summer institutes in statistics. For over ten years, he has served as Vice-Chair of Undergraduate Studies at the UCLA Department of Statistics, and he is director of the UCLA Center for the Teaching of Statistics. In 2012, Rob was elected Fellow of the American Statistical Association.

In his free time, Rob plays the cello and is an ardent reader of fiction.

Colleen Ryan

Colleen N. Ryan has taught statistics, chemistry, and physics to diverse community college students for decades. She taught at Oxnard College from 1975 to 2006, where she earned the Teacher of the Year Award. Colleen currently teaches statistics part-time at California Lutheran University. She often designs her own lab activities. Her passion is to discover new ways to make statistical theory practical, easy to understand, and sometimes even fun.

Colleen earned a B.A. in physics from Wellesley College, an M.A.T. in physics from Harvard University, and an M.A. in chemistry from Wellesley College. Her first exposure to statistics was with Frederick Mosteller at Harvard.

In her spare time, Colleen sings, has been an avid skier, and enjoys time with her family.

Rebecca K. Wong

Rebecca K. Wong has taught mathematics and statistics at West Valley College for more than twenty years. She enjoys designing activities to help students actively explore statistical concepts and encouraging students to apply those concepts to areas of personal interest.

Rebecca earned a B.A. in mathematics and psychology from the University of California, Santa Barbara, an M.S.T. in mathematics from Santa Clara University, and an Ed.D. in Educational Leadership from San Francisco State University. She has been recognized for outstanding teaching by the National Institute of Staff and Organizational Development and the California Mathematics Council of Community Colleges.

When not teaching, Rebecca is an avid reader and enjoys hiking trails with friends.

Contents

This page intentionally left blank

Preface

About This Text

The primary focus of this text is still, as in the first edition, data. We live in a data-driven economy and, more and more, in a data-centered culture. We don't choose whether we interact with data; the choice is made for us by websites that track our browsing patterns, membership cards that track our spending habits, cars that transmit our driving patterns, and smart phones that record our most personal moments.

The silver lining of what some have called the Data Deluge is that we all have access to rich and valuable data relevant in many important fields: environment, civics, social sciences, economics, health care, entertainment. This text teaches students to learn from such data and, we hope, to become cognizant of the role of the data that appear all around them. We want students to develop a data habit of mind in which, when faced with decisions, claims, or just plain curiosity, they know to reach for an appropriate data set to answer their questions. More important, we want them to have the skills to access these data and the understanding to analyze the data critically. Clearly, we've come a long way from the "mean median mode" days of rote calculation. To survive in the modern economy requires much more than knowing how to plug numbers into a formula. Today's students must know which questions can be answered by applying which statistic, and how to get technology to compute these statistics from within complex data sets.

What's New in the Second Edition

The second edition remains true to the goals of the first edition: to provide students with the tools they need to make sense of the world by teaching them to collect, visualize, analyze, and interpret data. With the help of several wise and careful readers and class testers, we have fine-tuned the second edition to better achieve this vision. In some sections, we have rewritten explanations or added new ones. In others, we have more substantially reordered content.

More precisely, in this new edition you will find

- Coverage of two-proportion confidence intervals in Chapters 7 and 8.
- An increase of more than 150 homework exercises in this edition, with more than 400 total new, revised, and updated exercises. We've added larger data sets to Chapters 2, 3, 4, and 9. We've also added exercises to Section 2.5 and more Chapter Review exercises throughout.
- New or updated examples in each chapter, with current topics such as views of stem cell research (Chapter 7) and online classes (Chapter 10).
- A more careful and thorough integration of technology in many examples.
- Two new case studies: Student-to-Teacher Ratios in Chapter 2 and Dodging the Question in Chapter 8.
- A more straightforward implementation of simulations to understand probability in Chapter 5.
- A more unified presentation of hypothesis testing in Chapter 8 that better joins conceptual understanding with application.
- A greater number of "Looking Back" and "Caution" marginal boxes to help direct students' studies.
- Updated technology guides to match current hardware and software.

Approach

Our text is concept-based, as opposed to method-based. We teach useful statistical methods, but we emphasize that applying the method is secondary to understanding the concept.

In the real world, computers do most of the heavy lifting for statisticians. We therefore adopt an approach that frees the instructor from having to teach tedious procedures and leaves more time for teaching deeper understanding of concepts. Accordingly, we present formulas as an aid to understanding the concepts, rather than as the focus of study.

We believe students need to learn how to

- Determine which statistical procedures are appropriate.
- Instruct the software to carry out the procedures.
- Interpret the output.

We understand that students will probably see only one type of statistical software in class. But we believe it is useful for students to compare output from several different sources, so in some examples we ask them to read output from two or more software packages.

One of the authors (Rob Gould) served on a panel of co-authors for the first edition of the collegiate version of the American Statistical Association–endorsed *Guidelines for Assessment and Instruction in Statistics Education* (*GAISE*). We firmly believe in its main goals and have adopted them in the preparation of this book.

- We emphasize understanding over rote performance of procedures.
- We use real data whenever possible.
- We encourage the use of technology both to develop conceptual understanding and to analyze data.
- We believe strongly that students learn by doing. For this reason, the homework problems offer students both practice in basic procedures and challenges to build conceptual understanding.

Coverage

The first few chapters of this book are concept-driven and cover exploratory data analysis and inferential statistics—fundamental concepts that every introductory statistics student should learn. The last part of the book builds on that strong conceptual foundation and is more methods-based. It presents several popular statistical methods and more fully explores methods presented earlier.

Our ordering of topics is guided by the process through which students should analyze data. First, they explore and describe data, possibly deciding that graphics and numerical summaries provide sufficient insight. Then they make generalizations (inferences) about the larger world.

Chapters 1–4: Exploratory Data Analysis. The first four chapters cover data collection and summary. Chapter 1 introduces the important topic of data collection and compares and contrasts observational studies with controlled experiments. This chapter also teaches students how to handle raw data so that the data can be uploaded to their statistical software. Chapters 2 and 3 discuss graphical and numerical summaries of single variables based on samples. We emphasize that the purpose is not just to produce a graph or a number but, instead, to explain what those graphs and numbers say about the world. Chapter 4 introduces simple linear regression and presents it as a technique for providing graphical and numerical summaries of relationships between two numerical variables.

We feel strongly that introducing regression early in the text is beneficial in building student understanding of the applicability of statistics to real-world scenarios. After completing the chapters covering data collection and summary, students have acquired the skills and sophistication they need to describe two-variable associations and to generate informal hypotheses. Two-variable associations provide a rich context for class discussion and allow the course to move from fabricated problems (because one-variable analyses are relatively rare in the real world) to real problems that appear frequently in everyday life.

Chapters 5–8: Inference. These chapters teach the fundamental concepts of statistical inference. The main idea is that our data mirror the real world, but imperfectly; although our estimates are uncertain, under the right conditions we can quantify our uncertainty. Verifying that these conditions exist and understanding what happens if they are not satisfied are important themes of these chapters.

Chapters 9–10: Methods. Here we return to the themes covered earlier in the text and present them in a new context by introducing additional statistical methods, such as estimating population means and analyzing categorical variables. We also provide (in Section 10.3) guidance for reading scientific literature, to offer students the experience of critically examining real scientific papers.

Organization

Our preferred order of progressing through the text is reflected in the Contents, but there are some alternative pathways as well.

10-week Quarter. The first eight chapters provide a full, one-quarter course in introductory statistics. If time remains, cover Sections 9.1 and 9.2 as well, so that students can solidify their understanding of confidence intervals and hypothesis tests by revisiting the topic with a new parameter.

Proportions First. Ask two statisticians, and you will get three opinions on whether it is best to teach means or proportions first. We have come down on the side of proportions for a variety of reasons. Proportions are much easier to find in popular news media (particularly around election time), so they can more readily be tied to students' everyday lives. Also, the mathematics and statistical theory are simpler; because there's no need to provide a separate estimate for the population standard deviation, inference is based on the Normal distribution, and no further approximations (that is, the t-distribution) are required. Hence, we can quickly get to the heart of the matter with fewer technical diversions.

The basic problem here is how to quantify the uncertainty involved in estimating a parameter and how to quantify the probability of making incorrect decisions when posing hypotheses. We cover these ideas in detail in the context of proportions. Students can then more easily learn how these same concepts are applied in the new context of means (and any other parameter they may need to estimate).

Means First. Conversely, many people feel that there is time for only one parameter and that this parameter should be the mean. For this alternative presentation, cover Chapters 6, 7, and 9, in that order. On this path, students learn about survey sampling and the terminology of inference (population vs. sample, parameter vs. statistic) and then tackle inference for the mean, including hypothesis testing.

To minimize the coverage of proportions, you might choose to cover Chapter 6, Section 7.1 (which treats the language and framework of statistical inference in detail), and then Chapter 9. Chapters 7 and 8 develop the concepts of statistical inference more slowly than Chapter 9, but essentially, Chapter 9 develops the same ideas in the context of the mean.

If you present Chapter 9 before Chapters 7 and 8, we recommend that you devote roughly twice as much time to Chapter 9 as you have devoted to previous chapters, because many challenging ideas are explored in this chapter. If you have already covered Chapters 7 and 8 thoroughly, Chapter 9 can be covered more quickly.

Features

We've incorporated into this text a variety of features to aid student learning and to facilitate its use in any classroom.

Integrating Technology

Modern statistics is inseparable from computers. We have worked to make this text accessible for any classroom, regardless of the level of in-class exposure to technology, while still remaining true to the demands of the analysis. We know that students sometimes do not have access to technology when doing homework, so many exercises provide output from software and ask students to interpret and critically evaluate that given output.

Using technology is important because it enables students to handle real data, and real data sets are often large and messy. The following features are designed to guide students.

- **TechTips** outline steps for performing calculations using TI-84® (including TI-84 + C®) graphing calculators, Excel®, Minitab®, and StatCrunch®. We do not want students to get stuck because they don't know how to reproduce the results we show in the text, so whenever a new method or procedure is introduced, an icon, Tech, refers students to the TechTips section at the end of the chapter. Each set of TechTips contains at least one mini-example, so that students are not only learning to use the technology but also practicing data analysis and reinforcing ideas discussed in the text. Most of the provided TI-84 steps apply to all TI-84 calculators, but some are unique to the TI-84 + C calculator.
- **Check Your Tech** examples help students understand that statistical calculations done by technology do not happen in a vacuum and assure them that they can get the same numerical values by hand. Although we place a higher value on interpreting results and verifying conditions required to apply statistical models, the numerical values are important, too.
- All **data sets** used in the exposition and exercises are available on the companion website at www.pearsonglobaleditions.com/gould.

Guiding Students

- Each chapter opens with a **Theme**. Beginners have difficulty seeing the forest for the trees, so we use a theme to give an overview of the chapter content.
- Each chapter begins by posing a real-world **Case Study**. At the end of the chapter, we show how techniques covered in the chapter helped solve the problem presented in the Case Study.
- **Margin Notes** draw attention to details that enhance student learning and reading comprehension.

  The data icon appears alongside examples or discussions to indicate that the original data are available on the companion website.

  Caution notes provide warnings about common mistakes or misconceptions.

Looking Back reminders refer students to earlier coverage of a topic.

Details clarify or expand on a concept.

- KEY POINT **Key Points** highlight essential concepts to draw special attention to them. Understanding these concepts is essential for progress.
- **Snapshots** break down key statistical concepts introduced in the chapter, quickly summarizing each concept or procedure and indicating when and how it should be used.
- An abundance of worked-out **examples** model solutions to real-world problems relevant to students' lives. Each example is tied to an end-of-chapter exercise so that students can practice solving a similar problem and test their understanding. Within the exercise sets, the icon TRY indicates which problems are tied to worked-out examples in that chapter, and the numbers of those examples are indicated.
- The **Chapter Review** that concludes each chapter provides a list of important new terms, student learning objectives, a summary of the concepts and methods discussed, and sources for data, articles, and graphics referred to in the chapter.

Active Learning

- For each chapter we've included an activity, **Exploring Statistics**, that students are intended to do in class as a group. We have used these activities ourselves, and we have found that they greatly increase student understanding and keep students engaged in class.
- All exercises are located at the end of the chapter. **Section Exercises** are designed to begin with a few basic problems that strengthen recall and assess basic knowledge, followed by mid-level exercises that ask more complex, open-ended questions. **Chapter Review Exercises** provide a comprehensive review of material covered throughout the chapter.

The exercises emphasize good statistical practice by requiring students to verify conditions, make suitable use of graphics, find numerical values, and interpret their findings in writing. All exercises are paired so that students can check their work on the odd-numbered exercise and then tackle the corresponding even-numbered exercise. The answers to all odd-numbered exercises appear in the back of the text.

Challenging exercises, identified with an asterisk (*), ask open-ended questions and sometimes require students to perform a complete statistical analysis. For exercises marked with a , accompanying data sets are available in MyStatLab and on the companion website.

- Most chapters include select exercises marked with a g within the exercise set, to indicate that problem-solving help is available in the **Guided Exercises** section. If students need support while doing homework, they can turn to the Guided Exercises to see a step-by-step approach to solving the problem.

Acknowledgments

We are grateful for the attention and energy that a large number of people devoted to making this a better book. We extend our gratitude to Elaine Newman (Sonoma State University) and Ann Cannon (Cornell College), who checked the accuracy of this text and its many exercises. Thanks also to David Chelton, our developmental editor, to Carol Merrigan, who handled production, to Peggy McMahon, project manager, and to Connie Day, our copyeditor. Many thanks to John Norbutas for his technical advice and help with the TechTips. We thank Suzanna Bainbridge, our acquisitions editor, Justin Billing, editorial assistant, and Deirdre Lynch, editor-in-chief, for signing us up and sticking with us, and we are grateful to Dona Kenly and Erin Kelly for their market development efforts.

We extend our sincere thanks for the suggestions and contributions made by the following reviewers of this edition:

Lloyd Best, *Pacific Union College*
Mario Borha, *Loyola University of Chicago*
David Bosworth, *Hutchinson Community College*
Beth Burns, *Bowling Green State University*
Jim Johnston, *Concord University*
Manuel Lopez, *Cerritos College*
Ralph Padgett Jr., *University of California – Riverside*
Ron Palcic, *Johnson County Community College*
Patrick Perry, *Hawaii Pacific University*
Victor I. Piercey, *Ferris State University*
Danielle Rivard, *Post University*
Alex Rolon, *North Hampton Community College*
Ali Saadat, *University of California – Riverside*
Carol Saltsgaver, *University of Illinois – Springfield*
Kelly Sakkinen, *Lake Land College*
Sharon l. Sullivan, *Catawba College*
Manuel Uy, *College of Alameda*
Mahbobeh Vezvaei, *Kent State University*
Jerimi Ann Walker, *Moraine Valley Community College*
Dottie Walton, *Cuyahoga Community College*
Judy H. Williams, *Tidewater Community College*
Jane Marie Wright, *Suffolk County Community College*
Cathleen Zucco-Teveloff, *Rider University*

We would also like to extend our sincere thanks for the suggestions and contributions that the following reviewers, class testers, and focus group attendees made to the previous edition.

Arun Agarwal, *Grambling State University*
Anne Albert, *University of Findlay*
Michael Allen, *Glendale Community College*
Eugene Allevato, *Woodbury University*
Dr. Jerry Allison, *Trident Technical College*
Polly Amstutz, *University of Nebraska*
Patricia Anderson, *Southern Adventist University*
MaryAnne Anthony-Smith, *Santa Ana College*
David C. Ashley, *Florida State College at Jacksonville*
Diana Asmus, *Greenville Technical College*
Kathy Autrey, *Northwestern State University of Louisiana*
Wayne Barber, *Chemeketa Community College*
Roxane Barrows, *Hocking College*
Jennifer Beineke, *Western New England College*
Diane Benner, *Harrisburg Area Community College*
Norma Biscula, *University of Maine, Augusta*
K.B. Boomer, *Bucknell University*
David Bosworth, *Hutchinson Community College*
Diana Boyette, *Seminole Community College*
Elizabeth Paulus Brown, *Waukesha County Technical College*
Leslie Buck, *Suffolk Community College*
R.B. Campbell, *University of Northern Iowa*
Stephanie Campbell, *Mineral Area College*
Ann Cannon, *Cornell College*
Rao Chaganty, *Old Dominion University*
Carolyn Chapel, *Western Technical College*
Christine Cole, *Moorpark College*
Linda Brant Collins, *University of Chicago*
James A. Condor, *Manatee Community College*
Carolyn Cuff, *Westminster College*
Phyllis Curtiss, *Grand Valley State University*
Monica Dabos, *University of California, Santa Barbara*
Greg Davis, *University of Wisconsin, Green Bay*
Bob Denton, *Orange Coast College*
Julie DePree, *University of New Mexico–Valencia*
Jill DeWitt, *Baker Community College of Muskegon*
Paul Drelles, *West Shore Community College*
Keith Driscoll, *Clayton State University*
Rob Eby, *Blinn College*
Nancy Eschen, *Florida Community College at Jacksonville*
Karen Estes, *St. Petersburg College*
Mariah Evans, *University of Nevada, Reno*
Harshini Fernando, *Purdue University North Central*
Stephanie Fitchett, *University of Northern Colorado*
Elaine B. Fitt, *Bucks County Community College*
Michael Flesch, *Metropolitan Community College*
Melinda Fox, *Ivy Tech Community College, Fairbanks*
Joshua Francis, *Defiance College*
Michael Frankel, *Kennesaw State University*
Heather Gamber, *Lone Star College*
Debbie Garrison, *Valencia Community College, East Campus*
Kim Gilbert, *University of Georgia*
Stephen Gold, *Cypress College*
Nick Gomersall, *Luther College*
Mary Elizabeth Gore, *Community College of Baltimore County–Essex*
Ken Grace, *Anoka Ramsey Community College*
Larry Green, *Lake Tahoe Community College*
Jeffrey Grell, *Baltimore City Community College*

Albert Groccia, *Valencia Community College, Osceola Campus*
David Gurney, *Southeastern Louisiana University*
Chris Hakenkamp, *University of Maryland, College Park*
Melodie Hallet, *San Diego State University*
Donnie Hallstone, *Green River Community College*
Cecil Hallum, *Sam Houston State University*
Josephine Hamer, *Western Connecticut State University*
Mark Harbison, *Sacramento City College*
Beverly J. Hartter, *Oklahoma Wesleyan University*
Laura Heath, *Palm Beach State College*
Greg Henderson, *Hillsborough Community College*
Susan Herring, *Sonoma State University*
Carla Hill, *Marist College*
Michael Huber, *Muhlenberg College*
Kelly Jackson, *Camden County College*
Bridgette Jacob, *Onondaga Community College*
Robert Jernigan, *American University*
Chun Jin, *Central Connecticut State University*
Maryann Justinger, Ed.D., *Erie Community College*
Joseph Karnowski, *Norwalk Community College*
Susitha Karunaratne, *Purdue University North Central*
Mohammed Kazemi, *University of North Carolina–Charlotte*
Robert Keller, *Loras College*
Omar Keshk, *Ohio State University*
Raja Khoury, *Collin County Community College*
Brianna Killian, *Daytona State College*
Yoon G. Kim, *Humboldt State University*
Greg Knofczynski, *Armstrong Atlantic University*
Jeffrey Kollath, *Oregon State University*
Erica Kwiatkowski-Egizio, *Joliet Junior College*
Sister Jean A. *Lanahan, OP, Molloy College*
Katie Larkin, *Lake Tahoe Community College*
Michael LaValle, *Rochester Community College*
Deann Leoni, *Edmonds Community College*
Lenore Lerer, *Bergen Community College*
Quan Li, *Texas A&M University*
Doug Mace, *Kirtland Community College*
Walter H. Mackey, *Owens Community College*
Keith McCoy, *Wilbur Wright College*
Elaine McDonald-Newman, *Sonoma State University*
William McGregor, *Rockland Community College*
Bill Meisel, *Florida State College at Jacksonville*
Bruno Mendes, *University of California, Santa Cruz*
Wendy Miao, *El Camino College*
Robert Mignone, *College of Charleston*
Ashod Minasian, *El Camino College*
Megan Mocko, *University of Florida*
Sumona Mondal, *Clarkson University*
Kathy Mowers, *Owensboro Community and Technical College*
Mary Moyinhan, *Cape Cod Community College*
Junalyn Navarra-Madsen, *Texas Woman's University*
Azarnia Nazanin, *Santa Fe College*
Stacey O. Nicholls, *Anne Arundel Community College*
Helen Noble, *San Diego State University*
Lyn Noble, *Florida State College at Jacksonville*
Keith Oberlander, *Pasadena City College*
Pamela Omer, *Western New England College*
Nabendu Pal, *University of Louisiana at Lafayette*
Irene Palacios, *Grossmont College*
Adam Pennell, *Greensboro College*
Joseph Pick, *Palm Beach State College*
Philip Pickering, *Genesee Community College*
Robin Powell, *Greenville Technical College*
Nicholas Pritchard, *Coastal Carolina University*
Linda Quinn, *Cleveland State University*
William Radulovich, *Florida State College at Jacksonville*
Mumunur Rashid, *Indiana University of Pennsylvania*
Fred J. Rispoli, *Dowling College*
Nancy Rivers, *Wake Technical Community College*
Corlis Robe, *East Tennesee State University*
Thomas Roe, *South Dakota State University*
Dan Rowe, *Heartland Community College*
Carol Saltsgaver, *University of Illinois–Springfield*
Radha Sankaran, *Passaic County Community College*
Delray Schultz, *Millersville University*
Jenny Shook, *Pennsylvania State University*
Danya Smithers, *Northeast State Technical Community College*
Larry Southard, *Florida Gulf Coast University*
Dianna J. Spence, *North Georgia College & State University*
René Sporer, *Diablo Valley College*
Jeganathan Sriskandarajah, *Madison Area Technical College–Traux*
David Stewart, *Community College of Baltimore County–Cantonsville*
Linda Strauss, *Penn State University*
John Stroyls, *Georgia Southwestern State University*
Joseph Sukta, *Moraine Valley Community College*
Lori Thomas, *Midland College*
Malissa Trent, *Northeast State Technical Community College*
Ruth Trygstad, *Salt Lake Community College*
Gail Tudor, *Husson University*
Manuel T. Uy, *College of Alameda*
Lewis Van Brackle, *Kennesaw State University*
Mahbobeh Vezvaei, *Kent State University*
Joseph Villalobos, *El Camino College*
Barbara Wainwright, *Sailsbury University*
Henry Wakhungu, *Indiana University*
Dottie Walton, *Cuyahoga Community College*
Jen-ting Wang, *SUNY, Oneonta*
Jane West, *Trident Technical College*
Michelle White, *Terra Community College*
Bonnie-Lou Wicklund, *Mount Wachusett Community College*
Sandra Williams, *Front Range Community College*
Rebecca Wong, *West Valley College*
Alan Worley, *South Plains College*
Jane-Marie Wright, *Suffolk Community College*
Haishen Yao, *CUNY, Queensborough Community College*
Lynda Zenati, *Robert Morris Community College*
Yan Zheng-Araujo, *Springfield Community Technical College*
Cathleen Zucco-Teveloff, *Rider University*
Mark A. Zuiker, *Minnesota State University,*

Acknowledgments for the Global Edition

Contributors
Vikas Arora, Statistician
Kiran Paul, Statistician

Reviewers
Santhosh Kumar, *Christ University*
Kiran Paul, Statistician
Chirag Trivedi, *R. J. Tibrewal Commerce College*

Resources for Success

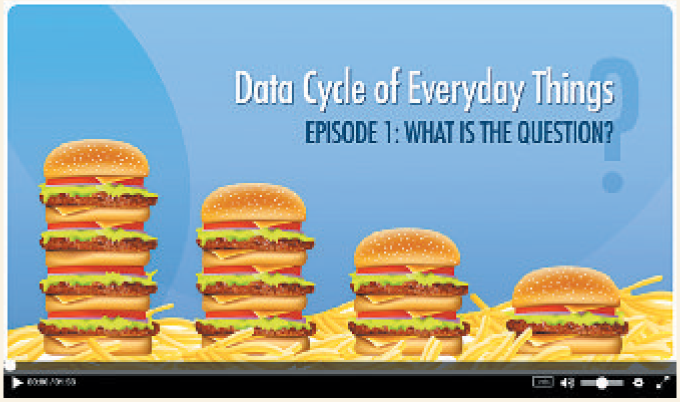
MyStatLab® Online Course for *Essential Statistics: Exploring the World through Data*, Second Edition, Gould/Ryan/Wong

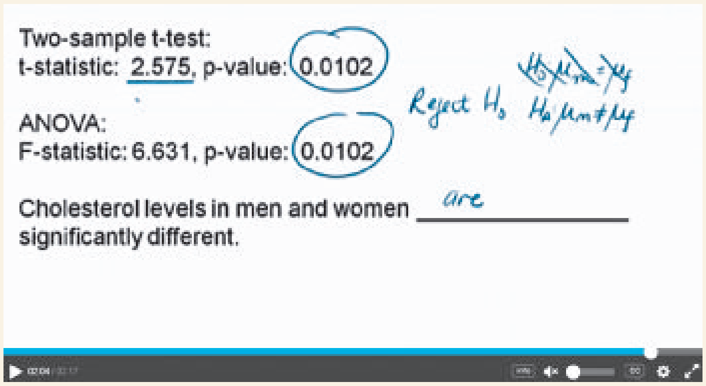
(access code required)

MyStatLab is available to accompany Pearson's market-leading text offerings. To give students a consistent tone, voice, and teaching method, each text's flavor and approach is tightly integrated throughout the accompanying MyStatLab course, making learning the material as seamless as possible.

Data Cycle of Everyday Things Videos

Data Cycle of Everyday Things Videos demonstrate for students that data can be, and is, a part of everyday life! Through a series of fun and engaging episodes, students learn to collect, analyze, and apply data to answer any range of real-world statistical questions.



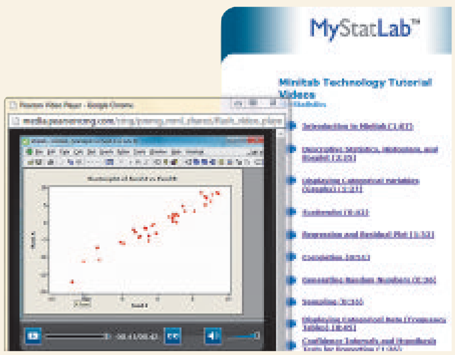


Chapter Review Videos

Chapter Review Videos walk students through solving a selection of the more complex problems posed in each chapter, providing a review of the chapter's key concepts and methods and offering students support where they most need it.

Technology Tutorial Videos and Study Cards

Technology Tutorials provide brief video walkthroughs and step-by-step instructional study cards on common statistical procedures for Minitab®, Excel®, and the TI-83/84 graphing calculator.



www.mystatlab.com

BREAKTHROUGH
To improving results

Resources for Success

Instructor Resources

Additional resources can be downloaded from **www.pearsonglobaleditions.com/gould** or MyStatLab.

Instructor's Solutions Manual

Instructor's Solutions Manual contains worked-out solutions to all the text exercises.

TestGen

TestGen® (**www.pearsoned.com/testgen**)
Updated to more closely mirror the 2nd Edition book, *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.

Online Test Bank

Online Test Bank (download only) includes tests for each chapter with questions aimed to reinforce the text's learning objectives.

PowerPoint® Lecture Slides

PowerPoint Lecture Slides, aligned with the text, provide an overview of each chapter, stressing important definitions and offering additional examples. Multiple-choice questions are included for class assessment.

Learning Catalytics

Learning Catalytics is a web-based engagement and assessment tool. As a "bring-your-own-device" direct response system, Learning Catalytics offers a diverse library of dynamic question types that allow students to interact with and think critically about statistical concepts. As a real-time resource, instructors can take advantage of critical teaching moments both in the classroom or through assignable and gradeable homework.

Student Resources

Additional resources for student success via MyStatLab.

Study Cards for Statistics Software

Study Cards for Statistics Software This series of study cards, available for Excel, Minitab, JMP®, SPSS®, R, StatCrunch, and the TI-84 graphing calculators, provides students with easy, step-by-step guides to the most common statistics software.

PowerPoint Slides

PowerPoint Slides provide an overview of each chapter, stressing important definitions and offering additional examples. These slides are an excellent resource for both traditional and online students.

Chapter Review Videos

Chapter Review Videos walk students through solving a selection of the more complex problems posed in each chapter, thereby offering a review of the chapter's key concepts and providing students support where they most need it.

The Data Cycle of Everyday Things Videos

The Data Cycle of Everyday Things Videos demonstrate for students that data collection and data analysis can be applied to answer questions about everyday life. Through a series of fun and engaging episodes, students learn to collect, analyze, and apply data to answer any range of real-world statistical questions.

Index of Applications

ENTERTAINMENT

ENVIRONMENT

FINANCE

FOOD AND DRINK

GAMES

GENERAL INTEREST

HEALTH

LAW

POLITICS

PSYCHOLOGY

SOCIAL ISSUES

SPORTS

SURVEYS AND OPINION POLLS

TECHNOLOGY

TRANSPORTATION

Essential Statistics: Exploring the World through Data

Second Edition

1 Introduction to Data

THEME

Statistics is the science of data, so we must learn the types of data we will encounter and the methods for collecting data. The method used to collect data is very important because it determines what types of conclusions we can reach and, as you'll learn in later chapters, what types of analyses we can do. By organizing the data we've collected, we can often spot patterns that are not otherwise obvious.

This text will teach you to examine data to better understand the world around you. If you know how to sift data to find patterns, can communicate the results clearly, and understand whether you can generalize your results to other groups and contexts, you will be able to make better decisions, offer more convincing arguments, and learn things you did not know before. Data are everywhere, and making effective use of them is such a crucial task that one prominent economist has proclaimed statistics one of the most important professions of the decade (*McKinsley Quarterly* 2009).

The use of statistics to make decisions and convince others to take action is not new. Some statisticians date the current practice of statistics back to the mid-nineteenth century. One famous example occurred in 1854, when the British were fighting the Russians in the brutal Crimean War. A British newspaper had criticized the military medical facilities, and a young but well-connected nurse, Florence Nightingale, was appointed to study the situation and, if possible to improve it.

Nightingale carefully recorded the numbers of deaths, the causes of the deaths, and the times and dates of the deaths. She organized these data graphically, and these graphs enabled her to see a very important pattern: A large percentage of deaths were due to contagious disease, and many deaths could be prevented by improving sanitary conditions. Within six months, Nightingale had reduced the death rate by half. Eventually she convinced Parliament and military authorities to completely reorganize the medical care they provided. Accordingly, she is credited with inventing modern hospital management.

In modern times, we have equally important questions to answer. Do cell phones cause brain tumors? Are alcoholic drinks healthful in moderation? Which diet works best for losing weight? What percentage of the public is concerned about job security? **Statistics**—the science (and art!) of collecting and analyzing observations to learn about ourselves, our surroundings, and our universe—helps answer questions such as these.

Data are the building blocks of statistics. This chapter introduces some of the basic types of data and explains how we collect them, store them, and organize them. These ideas and skills will provide a basic foundation for your study of the rest of the text.

CASE STUDY

Deadly Cell Phones?

In September 2002, Dr. Christopher Newman, a resident of Maryland, sued Motorola, Verizon, and other wireless carriers, accusing them of causing a cancerous brain tumor behind his right ear. As evidence, his lawyers cited a study by Dr. Lennart Hardell. Hardell had studied a large number of people with brain tumors and had found that a greater percentage of them used cell phones than of those who did not have brain tumors (CNN 2002; Brody 2002).

Speculation that cell phones might cause brain cancer began as early as 1993, when (as CNN reports) the interview show *Larry King Live* featured a man who claimed that his wife died because of cancer caused by her heavy cell phone use. However, more recent studies have contradicted Hardell's results, as well as earlier reports about the health risks of heavy cell phone use.

The judge in Dr. Newman's trial was asked to determine whether Hardell's study was compelling enough to support allowing the trial to proceed. Part of this determination involved evaluating the method that Hardell used to collect data. If you were the judge, how would you rule? You will learn the judge's ruling at the end of the chapter. You will also see how the methods used to collect data about important *cause-and-effect relationships*—such as that which Dr. Newman alleged to exist between cell phone use and brain cancer—can affect the conclusions we can draw.

SECTION 1.1

What Are Data?

The study of statistics rests on two major concepts: variation and data. **Variation** is the more fundamental of these concepts. To illustrate this idea, draw a circle on a piece of paper. Now draw another one, and try to make it look just the same. Now another. Are all three exactly the same? We bet they're not. They might be slightly different sizes, for instance, or slightly different versions of round. This is an example of variation. How can you reduce this variation? Maybe you can get a penny and outline the penny. Try this three times. Does variation still appear? Probably it does, even if you need a magnifying glass to see, say, slight variations in the thickness of the penciled line.

Data are observations that you or someone else records. The drawings in Figure 1.1 are data that record our attempts to draw three circles that look the same. Analyzing pictorial data such as these is not easy, so we often try to quantify such observations—that is, to turn them into numbers. How would you measure whether these three circles are the same? Perhaps you would compare diameters or circumferences, or somehow try to measure how and where these circles depart from being perfect circles. Whatever technique you chose, these measurements could also be considered data.

Details

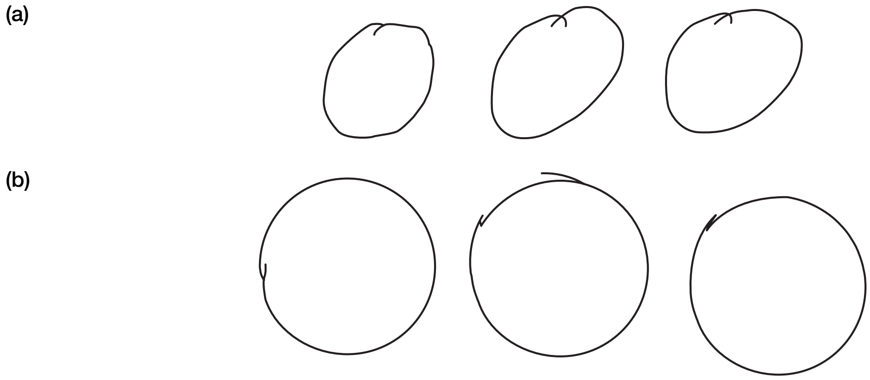
Data Are What Data Is

If you want to be "old school" grammatically correct, then the word *data* is plural. So we say "data *are*" and not "data *is*." The singular form is *datum*. However, this usage is changing over time, and some dictionaries now say that *data* can be used as both a singular and a plural noun.

Data are more than just numbers, though. David Moore, a well-known statistician, defined data as "numbers in context." By this he meant that data consist not only of the numbers we record, but also of the story behind the numbers. For example,

10.00, 9.88, 9.81, 9.81, 9.75, 9.69, 9.5, 9.44, 9.31

are just numbers. But in fact these numbers represent "Weight in pounds of the ten heaviest babies in a sample of babies born in North Carolina in 2004." Now these numbers have a context and have been elevated into data. See how much more interesting data are than numbers?



▶ **FIGURE 1.1** **(a)** Three circles drawn by hand. **(b)** Three circles drawn using a coin. It is clear that the circles drawn by hand show more variability than the circles drawn with the aid of a coin.

These data were collected by the state of North Carolina in part to help researchers understand the factors that contribute to low-weight and premature births. If doctors understand the causes of premature birth, they can work to prevent it—perhaps by helping expectant mothers change their behavior, perhaps by medical intervention, and perhaps by a combination of both.

KEY POINT Data are "numbers in context."

In the last few years, our culture and economy have been inundated with data. The magazine *The Economist* has called this surge of data the "data deluge." One reason for the rising tide of data is the application of automated data collection devices. These range from automatic sensors that simply record everything they see and store the data on a computer, to websites and smart phone apps that record every transaction their users make. Google, for example, saves every search you make and combines this with data on which links you click in order to improve the way it presents information (and also, of course, to determine which advertisements will appear on your search results).

Thanks to small, portable sensors, you can now join the "Personal Data Movement." Members of this movement record data about their daily lives and analyze it in order to improve their health, to run faster, or just to make keepsakes—a modern-day scrapbook. Maybe you or a friend uses a Nike Fuel Band to keep track of regular runs. One of the authors of this text carries a FitBit in his pocket to record his daily activity. From this he learned that on days he lectures, he typically takes 7600 steps, and on days that he does not lecture, he typically only takes 4900 steps. Some websites, such as your.flowingdata.com, make use of Twitter to help users collect, organize, and understand whatever personal data they choose to record.

Of course, it is not only machines that collect data. Humans still actively collect data with the intent of better understanding some phenomenon or making a discovery. Marketers prepare focus groups and surveys to describe the market for a new product. Sports analysts collect data to help their teams' coaches win games, or to help fantasy football league players. Scientists perform experiments to test theories and to measure changes in the economy or the climate. In this text you'll learn about the many ways in which data are used.

The point is that we have reached a historical moment where almost everything can be thought of as data. And once you find a way of capturing data about something in your world, you can organize, sort, visualize, and analyze those data to gain deeper understanding about the world around you.

What Is Data Analysis?

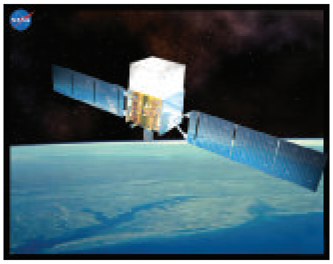
In this text you will study the science of data. Most important, you will learn to analyze data. What does this mean? You are analyzing data when you examine data of some sort and explain what they tell us about the real world. In order to do this, you must first learn about the different types of data, how data are stored and structured, and how they are summarized. The process of summarizing data takes up a big part of this text; indeed, we could argue that the entire text is about summarizing data, either through creating a visualization of the data or distilling them down to a few numbers that we hope capture their essence.

KEY POINT Data analysis involves creating summaries of data and explaining what these summaries tell us about the real world.

SECTION 1.2

Classifying and Storing Data

▲ **FIGURE 1.2** A photo of Carhenge, Nebraska.



▲ **FIGURE 1.3** Satellites in NASA's Earth Observing Mission record ultraviolet reflections and transmit these data back to Earth. Such data are used to construct images of our planet. Earth Observer (http://eos.gsfc.nasa.gov/).

Details

More Grammar

We're using the word *sample* as a noun—it is an object, a collection of data that we study. Later we'll also use the word *sample* as a verb—that is, to describe an action. For example, we'll sample ice cream cones to measure their weight.

Details

Quantitative and Qualitative Data

Some statisticians use the word *quantitative* to refer to numerical variables (think "quantity") and *qualitative* to refer to categorical variables (think "quality"). We prefer *numerical* and *categorical*. Both sets of terms are commonly used, and you should be prepared to hear and see both.

The first step in understanding data is to understand the different types of data you will encounter. As you've seen, data are numbers in context. But that's only part of the story; data are also recorded observations. Your photo from your vacation to Carhenge in Nebraska is data (Figure 1.2). The ultraviolet images streaming from the Earth Observer Satellite system are data (Figure 1.3). These are just two examples of data that are not numbers. Statisticians work hard to help us analyze complex data, such as images and sound files, just as easily as we study numbers. Most of the methods involve recoding the data into numbers. For example, your photos can be digitized in a scanner, converted into a very large set of numbers, and then analyzed. You might have a digital camera that gives you feedback about the quality of a photo you've taken. If so, your camera is not only collecting data but also analyzing it!

Almost always, our data sets will consist of characteristics of people or things (such as gender and weight). These characteristics are called **variables**. Variables are not "unknowns" like those you studied in algebra. We call these characteristics variables because they have variability: The values of the variable can be different from person to person.

KEY POINT Variables in statistics are different from variables in algebra. In statistics, variables record characteristics of people or things.

When we work with data, they are grouped into a collection, which we call either a **data set** or a **sample**. The word *sample* is important, because it implies that the data we see are just one part of a bigger picture. This "bigger picture" is called a **population**. Think of a population as the Data Set of Everything—it is the data set that contains all of the information about everyone or everything with respect to whatever variable we are studying. Quite often, the population is really what we want to learn about, and we learn about it by studying the data in our sample. However, many times it is enough just to understand and describe the sample. For example, you might collect data from students in your class simply because you want to know about the students in your class, and not because you wish to use this information to learn about all students at your school. Sometimes, data sets are so large that they effectively *are* the population, as you'll soon see in the data reflecting births in North Carolina.

Two Types of Variables

The variables you'll find in your data set come in two basic types, which can themselves be broken into smaller divisions, as we'll discuss later.

Numerical variables describe quantities of the objects of interest. The values will be numbers. The weight of an infant is an example of a numerical variable.

Categorical variables describe qualities of the objects of interest. These values will be categories. The sex of an infant is an example of a categorical variable. The possible values are the categories "male" and "female." Eye color of an infant is another example; the categories might be brown, blue, black, and so on. You can often identify categorical variables because their values are *usually* words, phrases, or letters. (We say "usually" because we sometimes use numbers to represent a word or phrase. Stay tuned.)