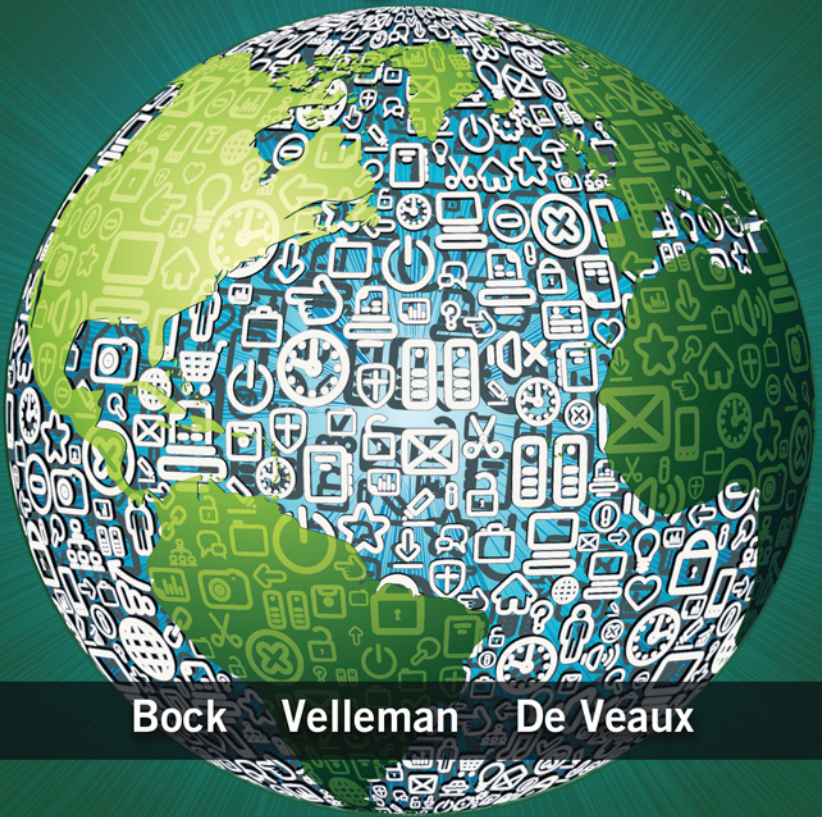
# STATS 4e

## Modeling the World



Bock Velleman De Veaux

# STATS 4e

## Modeling the World

# STATS 4e

## Modeling the World

**David E. Bock**
Ithaca High School (Retired)

**Paul F. Velleman**
Cornell University

**Richard D. De Veaux**
Williams College

PEARSON

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

**Editor-in-Chief:** Deirdre Lynch
**Executive Editor:** Christopher Cummings
**Senior Content Editor:** Chere Bemelmans
**Assistant Editor:** Sonia Ashraf
**Senior Managing Editor:** Karen Wernholm
**Associate Managing Editor:** Tamela Ambush
**Project Managers:** Sherry Berg and Sheila Spinney
**Digital Assets Manager:** Marianne Groth
**Supplements Production Coordinator:** Katherine Roz
**Manager, Multimedia Production:** Christine Stavrou
**Media Producer:** Vicki Dreyfus
**Software Development:** Bob Carroll and Mary Durnwald
**Senior Marketing Manager:** Erin K. Lane
**Marketing Coordinator:** Kathleen DeChavez
**Senior Author Support/Technology Specialist**: Joe Vetere
**Rights and Permissions Advisor:** Dana Weightman
**Image Manager:** Rachel Youdelman
**Procurement Specialist:** Debbie Rossi
**Associate Director of Design:** Andrea Nix
**Senior Designer and Cover Design:** Barbara Atkinson
**Text Design:** Studio Montage
**Production Management, Composition, and Illustrations:** PreMedia Global
**Cover Image:** iStockphoto/Thinkstock/Getty Images

The Pearson team would like to acknowledge Sheila Spinney and her many years of hard work and dedication to publishing. She was a joy to work with and a wonderful friend, and we will miss her greatly.



**Library of Congress Cataloging-in-Publication Data**

Bock, David E.
Stats : modeling the world / David E. Bock, Paul F. Velleman, Richard D. De Veaux.— 4th ed.
p. cm.
Includes index.
ISBN 978-0-321-85401-8
1. Graphic calculators—Textbooks. I. Velleman, Paul F., 1949- II. De Veaux, Richard D. III. Title.
QA276.12.B628 2010
519.5—dc22

2012005942



1 2 3 4 5 6 7 8 9 10—CRK— 17 16 15 14 13

PEARSON

**www.pearsonhighered.com**

ISBN 13: 978-0-321-85401-8
ISBN 10: 0-321-85401-2

*To Greg and Becca, great fun as kids and great friends as adults,*
*and especially to my wife and best friend, Joanna, for her*
*understanding, encouragement, and love*

*—Dave*

*To my sons, David and Zev, from whom I've learned so much,*
*and to my wife, Sue, for taking a chance on me*

*—Paul*

*To Sylvia, who has helped me in more ways than she'll ever know,*
*and to Nicholas, Scyrine, Frederick, and Alexandra,*
*who make me so proud in everything that they are and do*

*—Dick*

# Meet the Authors

**David E. Bock** taught mathematics at Ithaca High School for 35 years. He has taught Statistics at Ithaca High School, Tompkins-Cortland Community College, Ithaca College, and Cornell University. Dave has won numerous teaching awards, including the MAA's Edyth May Sliffe Award for Distinguished High School Mathematics Teaching (twice), Cornell University's Outstanding Educator Award (three times), and has been a finalist for New York State Teacher of the Year.

Dave holds degrees from the University at Albany in Mathematics (B.A.) and Statistics/Education (M.S.). Dave has been a reader and table leader for the AP Statistics exam, serves as a Statistics consultant to the College Board, and leads workshops and institutes for AP Statistics teachers. He also served as K–12 Education and Outreach Coordinator and senior lecturer for the Mathematics Department at Cornell University. His understanding of how students learn informs much of this book's approach.

Dave and his wife relax by biking or hiking, and when not at home near Ithaca can often be found in North Carolina's Blue Ridge Mountains. They have a son, a daughter, and four grandchildren.

**Paul F. Velleman** has an international reputation for innovative Statistics education. He is the author and designer of the multimedia Statistics program *ActivStats*, for which he was awarded the EDUCOM Medal for innovative uses of computers in teaching statistics, and the ICTCM Award for Innovation in Using Technology in College Mathematics. He also developed the award-winning statistics program, *Data Desk*, and the Internet site Data and Story Library (DASL) (lib.stat.cmu.edu/DASL/), which provides data sets for teaching Statistics. Paul's understanding of using and teaching with technology informs much of this book's approach.

Paul teaches Statistics at Cornell University in the Department of Statistical Sciences, for which he has been awarded the MacIntyre prize for Exemplary Teaching. He holds an A.B. from Dartmouth College in Mathematics and Social Science, and M.S. and Ph.D. degrees in Statistics from Princeton University, where he studied with John Tukey. His research often deals with statistical graphics and data analysis methods. Paul co-authored (with David Hoaglin) *ABCs of Exploratory Data Analysis.* Paul is a Fellow of the American Statistical Association and of the American Association for the Advancement of Science. Paul is the father of two boys.

**Richard D. De Veaux** is an internationally known educator and consultant. He has taught at the Wharton School and the Princeton University School of Engineering, where he won a "Lifetime Award for Dedication and Excellence in Teaching." Since 1994, he has taught at Williams College. He is currently the C. Carlisle and Margaret Tippit Professor of Statistics at Williams College. Dick has won both the Wilcoxon and Shewell awards from the American Society for Quality. He is an elected member of the International Statistics Institute (ISI) and a fellow of the American Statistical Association (ASA). In 2008, he was named Statistician of the Year by the Boston Chapter of the ASA. Dick is also well known in industry, where for more than 25 years he has consulted for such Fortune 500 companies as American Express, Hewlett-Packard, Alcoa, DuPont, Pillsbury, General Electric, and Chemical Bank. Because he consulted with Mickey Hart on his book *Planet Drum*, he has also sometimes been called the "Official Statistician for the Grateful Dead." His real-world experiences and anecdotes illustrate many of this book's chapters.

Dick holds degrees from Princeton University in Civil Engineering (B.S.E.) and Mathematics (A.B.) and from Stanford University in Dance Education (M.A.) and Statistics (Ph.D.), where he studied dance with Inga Weiss and Statistics with Persi Diaconis. His research focuses on the analysis of large data sets and data mining in science and industry.

In his spare time, he is an avid cyclist and swimmer. He also is the founder and bass for the doo-wop group, the Diminished Faculty, and is a frequent singer and soloist with various local choirs, including the Choeur Vittoria of Paris, France. Dick is the father of four children.

# Table of Contents

## Part V From the Data at Hand to the World at Large



## Part VI Learning About the World



## Part VII Inference When Variables Are Related



### Appendixes



*Optional chapter.

# Preface

## About the Book

Yes, a preface is supposed to be "about this book" – and we'll get there – but first we want to talk about the bigger picture: the ongoing growth of interest in Statistics. From the hit movie *Moneyball* to Nate Silver's success at predicting elections to *Wall Street Journal* and *New York Times* articles touting the explosion of job opportunities for graduates with degrees in Statistics, public awareness of the widespread applicability, power, and importance of statistical analysis has never been higher. Each year, more students sign up for Stats courses and discover what drew us to this field: it's interesting, stimulating, and even fun. Statistics helps students develop key tools and critical thinking skills needed to become well-informed consumers, parents, and citizens. We think Statistics isn't as much a math course as a civics course, and we're delighted that our books can play a role in preparing a generation for life in the Information Age.

## New to the Fourth Edition

This new edition of *Stats: Modeling the World* extends the series of innovations pioneered in our books, teaching Statistics and statistical thinking as it is practiced today. We've made some important revisions and additions, each with the goal of making it even easier for students to put the concepts of Statistics together into a coherent whole.

- ***Chapter 1 (and beyond).*** Now Chapter 1 gets down to business immediately, looking at data rather than just presenting the book's features. And throughout the book we've rewritten many other sections to make them clearer and more interesting. Several chapters lead with new up-to-the-minute motivating examples and follow through with analyses of the data, and many other new examples provide a basis for sample problems and exercises.
- ***What If.*** We close most chapters by looking at a simulation that explores or extends an important concept. Starting with Chapter 2, students see the power of simulation as they gain additional insights or get a sneak preview of important ideas yet to come. These *What If* elements offer great fodder for class discussions while paving the way for better grasp of such critical concepts as independence, sampling variability, the Central Limit Theorem, and statistical significance.
- ***Practice Exams.*** At the end of each of the book's seven parts you'll find a practice exam, consisting of both multiple choice and free response questions. These cumulative exams encourage students to keep important concepts and skills in mind throughout the course while helping them synthesize their understanding as they build connections among the various topics.
- ***What Have We Learned?*** We've revised our chapter-ending study guides to better help students review the key concepts and terms.
- ***Updated examples, exercises, and data.*** We've updated our innovative *Think/Show/Tell Step-by-Step* examples with new contexts and data. We've added hundreds of new exercises and updated continuing exercises with the most recent data. Whenever possible, we've provided those data on the DVD and the book's website. Most of the examples and exercises are based on recent news stories, research articles, and other real-world sources. We've listed many of those sources so students can explore them further.
- ***Updated TI Tips.*** Each chapter's easy-to-read "TI Tips" now show students how to use TI-84 Plus statistics functions with the StatWizard operating system.
- ***Streamlined design.*** This edition sports a new design that clarifies the purpose of each text element. The major theme of each chapter is easier to follow without distraction.

To better help students know where to focus their study efforts, essential supporting material is shaded, while enriching—and often entertaining—side material is not.

## Our Goal: Read This Book!

The best text in the world is of little value if students don't read it. Starting with the first edition, our goal has been to create a book that students would willingly read, easily learn from, and even like. We've been thrilled with the glowing feedback we've received from instructors and students using the first three editions of *Stats: Modeling the World*. Our conversational style, our interesting anecdotes and examples, and even our humor[1] engage students' interest as they learn statistical thinking. We hear from grateful instructors that their students actually do read this book (sometimes even voluntarily reading ahead of the assignments). And we hear from (often amazed) students that they actually enjoyed their textbook.

Here are some of the ways we have made *Stats: Modeling the World*, Fourth Edition engaging:

- ***Readability.*** You'll see immediately that this book doesn't read like other Statistics texts. The style is both colloquial and informative, enticing students to actually read the book to see what it says.
- ***Informality.*** Our informal style doesn't mean that the subject matter is covered superficially. Not only have we tried to be precise, but wherever possible we offer deeper explanations and justifications than those found in most introductory texts.
- ***Focused lessons.*** The chapters are shorter than in most other texts, making it easier for both instructors and students to focus on one topic at a time.
- ***Consistency.*** We've worked hard to demonstrate how to do Statistics well. From the very start and throughout the book we model the importance of plotting data, of checking assumptions and conditions, and of writing conclusions that are clear, concise, and in context.
- ***The need to read.*** Because the important concepts, definitions, and sample solutions aren't set in boxes, students won't find it easy to just to skim this book. We intend that it be read, so we've tried to make the experience enjoyable.

## Continuing Features

Along with the improvements we've made, you'll still find the many engaging, innovative, and pedagogically effective features responsible for the success of our earlier editions.

- ***Think, Show, Tell.*** The worked examples repeat the mantra of *Think, Show*, and *Tell* in every chapter. They emphasize the importance of thinking about a Statistics question (What do we know? What do we hope to learn? Are the assumptions and conditions satisfied?) and reporting our findings (the *Tell* step). The *Show* step contains the mechanics of calculating results and conveys our belief that it is only one part of the process.
- ***Step-by-Step*** examples guide students through the process of analyzing a problem by showing the general explanation on the left and the worked-out solution on the right. The result: better understanding of the concept, not just number crunching.

[1] And, yes, those footnotes!

- ***For Example.*** In every chapter, an interconnected series of *For Example* elements present a continuing discussion, recapping a story and moving it forward to illustrate how to apply each new concept or skill.
- ***Just Checking.*** At key points in each chapter, we ask students to pause and think with questions designed to be a quick check that they understand the material they've just read. Answers are at the end of the exercise sets in each chapter so students can easily check themselves.
- ***Updated TI Tips.*** Each chapter's easy-to-read "TI Tips" now show students how to use TI-84 Plus statistics functions with the StatWizard operating system. (Help using a TI-Nspire appears in Appendix B, and help with a TI-89 is on the book's companion website www.pearsonhighered.com/bock.) As we strive for sound understanding of formulas and methods, we want students to use technology for actual calculations. We do emphasize that calculators are just for "Show"—they cannot Think about what to do nor Tell what it all means.
- ***Math Boxes.*** In many chapters we present the mathematical underpinnings of the statistical methods and concepts. By setting these proofs, derivations, and justifications apart from the narrative, we allow students to continue to follow the logical development of the topic at hand, yet also explore the underlying mathematics for greater depth.
- ***ActivStats Pointers.*** Margin pointers alert students to *ActivStats* videos, simulations, animations, and activities that enhance learning by paralleling the book's discussions.
- ***TI-Nspire Activities.*** Other margin pointers identify demonstrations and investigations for TI-Nspire handhelds to enhance each chapter. They're found at the book's website (www.pearsonhighered.com/bock).
- ***What Can Go Wrong?*** Each chapter still contains our innovative *What Can Go Wrong?* sections that highlight the most common errors people make and the misconceptions they have about Statistics. Our goals are to help students avoid these pitfalls and to arm them with the tools to detect statistical errors and to debunk misuses of statistics, whether intentional or not.
- ***Exercises.*** We've maintained the pairing of examples so that each odd-numbered exercise (with an answer in the back of the book) is followed by an even-numbered exercise illustrating the same concept. Exercises are ordered by approximate level of complexity.
- ***Reality Check.*** We regularly remind students that Statistics is about understanding the world with data. Results that make no sense are probably wrong, no matter how carefully we think we did the calculations. Mistakes are often easy to spot with a little thought, so we ask students to stop for a reality check before interpreting their result.
- ***Notation Alerts.*** Clear communication is essential in Statistics, and proper notation is part of the vocabulary students need to learn. We've found that it helps to call attention to the letters and symbols statisticians use to mean very specific things.
- ***On the Computer.*** Because real-world data analysis is done on computers, at the end of each chapter we summarize what students can find in most statistics software, usually with an annotated example.

## Our Approach

We've been guided in the choice of topics and emphasis on clear communication by the requirements of the Advanced Placement Statistics course. In our order of presentation, we have tried to ensure that each new topic fits logically into the growing structure of understanding that we hope students will build.

## GAISE Guidelines

We have worked to provide materials to help each class, in its own way, follow the guidelines of the GAISE (Guidelines for Assessment and Instruction in Statistics Education) project sponsored by the American Statistical Association. That report urges that Statistics education should

1. emphasize Statistical literacy and develop Statistical thinking,
2. use real data,
3. stress conceptual understanding rather than mere knowledge of procedures,
4. foster active learning,
5. use technology for developing concepts and analyzing data, and
6. make assessment a part of the learning process.

## Mathematics

Mathematics traditionally appears in Statistics texts in several roles:

1. It can provide a concise, clear statement of important concepts.
2. It can embody proofs of fundamental results.
3. It can describe calculations to be performed with data.

Of these, we emphasize the first. Mathematics can make discussions of Statistics concepts, probability, and inference clear and concise. We have tried to be sensitive to those who are discouraged by equations by also providing verbal descriptions and numerical examples.

This book is not concerned with proving theorems about Statistics. Some of these theorems are quite interesting, and many are important. Often, though, their proofs are not enlightening to introductory Statistics students, and can distract the audience from the concepts we want them to understand. However, we have not shied away from the mathematics where we believed that it helped clarify without intimidating. You will find some important proofs, derivations, and justifications in the Math Boxes that accompany the development of many topics.

Nor do we concentrate on calculations. Although statistics calculations are generally straightforward, they are also usually tedious. And, more to the point, they are often unnecessary. Today, virtually all statistics are calculated with technology, so there is little need for students to work by hand. The equations we use have been selected for their focus on understanding concepts and methods.

# Technology and Data

To experience the real world of Statistics, it's best to explore real data sets using modern technology. This fact permeates *Stats: Modeling the World*, Fourth Edition, where we use real data for the book's examples and exercises. Technology lets us focus on teaching statistical thinking rather than getting bogged down in calculations. The questions that motivate each of our hundreds of examples are not "How do you find the answer?" but "How do you think about the answer?"

***Technology.*** We assume that students are using some form of technology in this Statistics course. That could include a graphing calculator along with a statistics package or spreadsheet. Rather than adopt any particular software, we discuss generic computer output. "TI-Tips"—included in most chapters—show students how to use statistics features of the TI-84 Plus series. The DVD includes *ActivStats* and the software package Data Desk. In Appendix B, we offer general guidance (by chapter) to help students get started on

common software platforms (StatCrunch, Excel, MINITAB, Data Desk, JMP, and SPSS) and a TI-Nspire. The book's website includes additional guidance for students using a TI-89.

***Data.*** Because we use technology for computing, we don't limit ourselves to small, artificial data sets. In addition to including some small data sets, we have built examples and exercises on real data with a moderate number of cases—usually more than you would want to enter by hand into a program or calculator. These data are included on the DVD as well as on the book's website, www.pearsonhighered.com/bock.

## On the DVD

The DVD includes a wealth of supporting materials.

***ActivStats.*** The award-winning ActivStats multimedia program complements the book with videos of real-word stories, worked examples, animated expositions of each of the major Statistics topics, and tools for performing simulations, visualizing inference, and learning to use statistics software. The new version of *ActivStats* includes

- Improved navigation and a cleaner design that makes it easier to find and use tools;
- More than 1000 homework exercises;
- Video clips, animated activities, teaching applets, and more than 300 data sets.

***Data Desk.*** This full-featured statistics software package is both powerful and easy to use.

***Additional tech guidance*** for the TI-89 calculators.

***Additional chapters.*** Two additional chapters cover **Analysis of Variance** (Chapter 27) and **Multiple Regression** (Chapter 28). These topics point the way to further study in Statistics.

# Supplements

## For the Student

**Stats: Modeling the World, Fourth Edition**, for-sale student edition (ISBN-13: 978-0-321-85401-8; ISBN-10: 0-321-85401-2)

**Graphing Calculator Manual** (download only) by John Diehl (Hinsdale Central High School) and Patricia Humphrey (Georgia Southern University) is organized to follow the sequence of topics in the text, and is an easy-to-follow, step-by-step guide on how to use the TI-84 Plus, TI-Nspire, and Casio graphing calculators. It provides worked-out examples to help students fully understand and use the graphing calculator. Available for download from www.pearsonhighered.com/mathstatsresources.

**Study card** for the De Veaux/Velleman/Bock Statistics Series is a resource for students containing important formulas, definitions, and tables that correspond precisely to the De Veaux/Velleman/Bock Statistics series. This card can work as a reference for completing homework assignments or as an aid in studying. (ISBN-13: 978-0-321-82626-8; ISBN-10: 0-321-82626-4)

**Videos** for the Bock/Velleman/De Veaux Series, Fourth Edition, available to stream from within MyStatLab®.

## For the Instructor

**Instructor's Edition** contains answers to all exercises. (ISBN-13: 978-0-321-85858-0; ISBN-10: 0-321-85858-1)

**Instructor's Solutions Manual** (download only), by William Craine, contains detailed solutions to all of the exercises. The Instructor's Solutions Manual is available to download from within MyStatLab® and in the Instructor Resource Center at **www.pearsonhighered.com/irc**.

**Online Test Bank and Resource Guide**, by William Craine, with contributions from Corey Andreasen, Jared Derksen, John Diehl, and Jane Viau, is completely revised and expanded for the fourth edition. The Test Bank and Resource Guide contains chapter-by-chapter comments on major concepts; tips on presenting topics (and what to avoid); teaching examples; suggested assignments; Web links and lists of other resources; additional sets of chapter quizzes, unit tests, and investigative tasks; TI-Nspire activities; and suggestions for projects. We've added more worksheets on key topics, correspondence to AP exam questions in each chapter, and reading guides to the fourth edition. An indispensable guide to help instructors prepare for class, the previous editions were soundly praised by new instructors of Statistics and seasoned veterans alike. The Online Test Bank and Resource Guide is available to download from within MyStatLab® and in the Instructor Resource Center at **www.pearsonhighered.com/irc**.

**Instructor's Podcasts** (10 points in 10 minutes). These audio podcasts focus on key points in each chapter to help you with class preparation. They can be easily downloaded from MyStatLab and the Instructor Resource Center at **www.pearsonhighered.com/irc**.

## Technology Resources

### MyStatLab™ Online Course (access code required)

MyStatLab is a course management systems that delivers **proven results** in helping individual students succeed.

- MyStatLab can be successfully implemented in any environment—lab-based, hybrid, fully online, traditional—and demonstrates the quantifiable difference that integrated usage has on student retention, subsequent success, and overall achievement.
- MyStatLab's comprehensive online gradebook automatically tracks students' results on tests, quizzes, homework, and in the study plan. Instructors can use the gradebook to provide positive feedback or intervene if students have trouble. Gradebook data can be easily exported to a variety of spreadsheet programs, such as Microsoft Excel.

MyStatLab provides **engaging experiences** that personalize, stimulate, and measure learning for each student.

- **Tutorial Exercises with Multimedia Learning Aids:** The homework and practice exercises in MyStatLab align with the exercises in the textbook, and they regenerate algorithmically to give students unlimited opportunity for practice and mastery. Exercises offer immediate helpful feedback, guided solutions, sample problems, animations, videos, and eText clips for extra help at point-of-use.
- **Adaptive Study Plan:** Pearson now offers an optional focus on adaptive learning in the study plan to allow students to work on just what they need to learn when it makes the most sense to learn it. The adaptive study plan maximizes students' potential for understanding and success.
- **Additional Statistics Question Libraries:** In addition to algorithmically regenerated questions that are aligned with your textbook, MyStatLab courses come with two additional question libraries. **450 Getting Ready for Statistics** questions offer the developmental math topics students need for the course. These can be assigned as a prerequisite to other assignments, if desired. The **1000 Conceptual Question Library** require students to apply their statistical understanding.
- **StatCrunch®:** MyStatLab includes a web-based statistical software, StatCrunch, within the online assessment platform so that students can easily analyze data sets from exercises and the text. In addition, MyStatLab includes access to **www.StatCrunch.com**, a website where users can access tens of thousands of shared data sets, conduct online surveys, perform complex analyses using the powerful statistical software, and generate compelling reports.
- **Integration of Statistical Software:** Knowing that students often use external statistical software, we make it easy to copy our data sets, both from the ebook and the

MyStatLab questions, into software such as StatCrunch, Minitab, Excel, and more. Students have access to a variety of support tools—Technology Instruction Videos, Technology Study Cards, and Manuals for select titles—to learn how to effectively use statistical software.

- **StatTalk Videos:** Fun-loving statistician Andrew Vickers takes to the streets of Brooklyn, NY, to demonstrate important statistical concepts through interesting stories and real-life events. This series of 24 videos will actually help you understand statistics. Accompanying assessment questions and instructor's guide available.
- **Expert Tutoring:** Although many students describe the whole of MyStatLab as "like having your own personal tutor," students also have access to live tutoring from Pearson. Qualified statistics instructors provide tutoring sessions for students via MyStatLab.

And, MyStatLab comes from a **trusted partner** with educational expertise and an eye on the future.

- Knowing that you are using a Pearson product means knowing that you are using quality content. That means that our eTexts are accurate, that our assessment tools work, and that our questions are error-free. And whether you are just getting started with MyStatLab, or have a question along the way, we're here to help you learn about our technologies and how to incorporate them into your course.

To learn more about how MyStatLab combines proven learning applications with powerful assessment, visit **www.mystatlab.com** or contact your Pearson representative.

## MyStatLab™ Ready to Go Course (access code required)

These new Ready to Go courses provide students with all the same great MyStatLab features that you're used to, but make it easier for instructors to get started. Each course includes preassigned homeworks and quizzes to make creating your course even simpler. Ask your Pearson representative about the details for this particular course or to see a copy of this course.

## MyMathLab®Plus/MyStatLab™Plus

MyLabsPlus combines proven results and engaging experiences from MyMathLab® and MyStatLab™ with convenient management tools and a dedicated services team. Designed to support growing math and statistics programs, it includes additional features such as:

- **Batch Enrollment:** Your school can create the login name and password for every student and instructor, so everyone can be ready to start class on the first day. Automation of this process is also possible through integration with your school's Student Information System.
- **Login from Your Campus Portal:** You and your students can link directly from your campus portal into your MyLabsPlus courses. A Pearson service team works with your institution to create a single sign-on experience for instructors and students.
- **Advanced Reporting:** MyLabsPlus's advanced reporting allows instructors to review and analyze students' strengths and weaknesses by tracking their performance on tests, assignments, and tutorials. Administrators can review grades and assignments across all courses on your MyLabsPlus campus for a broad overview of program performance.
- **24/7 Support:** Students and instructors receive 24/7 support, 365 days a year, by email or online chat.

MyLabsPlus is available to qualified adopters. For more information, visit our website at www.mylabsplus.com or contact your Pearson representative.

## MathXL® for Statistics Online Course (access code required)

MathXL® is the homework and assessment engine that runs MyStatLab. (MyStatLab is MathXL plus a learning management system.)

With MathXL for Statistics, instructors can:

- Create, edit, and assign online homework and tests using algorithmically generated exercises correlated at the objective level to the textbook.
- Create and assign their own online exercises and import TestGen tests for added flexibility.
- Maintain records of all student work, tracked in MathXL's online gradebook.

With MathXL for Statistics, students can:

- Take chapter tests in MathXL and receive personalized study plans and/or personalized homework assignments based on their test results.
- Use the study plan and/or the homework to link directly to tutorial exercises for the objectives they need to study.
- Students can also access supplemental animations and video clips directly from selected exercises.
- Knowing that students often use external statistical software, we make it easy to copy our data sets, both from the ebook and the MyStatLab questions, into software like StatCrunch®, Minitab, Excel and more.

MathXL for Statistics is available to qualified adopters. For more information, visit our website at www.mathxl.com, or contact your Pearson representative.

## StatCrunch®

StatCrunch is powerful web-based statistical software that allows users to perform complex analyses, share data sets, and generate

compelling reports of their data. The vibrant online community offers tens of thousands of data sets for students to analyze.

- **Collect.** Users can upload their own data to StatCrunch or search a large library of publicly shared data sets, spanning almost any topic of interest. Also, an online survey tool allows users to quickly collect data via web-based surveys.
- **Crunch.** A full range of numerical and graphical methods allow users to analyze and gain insights from any data set. Interactive graphics help users understand statistical concepts, and are available for export to enrich reports with visual representations of data.
- **Communicate.** Reporting options help users create a wide variety of visually-appealing representations of their data.

Full access to StatCrunch is available with a MyStatLab kit, and StatCrunch is available by itself to qualified adopters. StatCrunch Mobile is now available to access from your mobile device. For more information, visit our website at **www.StatCrunch.com**, or contact your Pearson representative.

## StatCrunch® eBook

This interactive, online textbook includes StatCrunch, a powerful, web-based statistical software. Embedded StatCrunch buttons allow users to open all data sets and tables from the book with the click of a button and immediately perform an analysis using StatCrunch.

## TestGen®

TestGen® (**www.pearsoned.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 testbank are available for download from Pearson Education's online catalog.

## PowerPoint® Lecture Slides

PowerPoint® Lecture Slides provide an outline to use in a lecture setting, presenting definitions, key concepts, and figures from the text. These slides are available within MyStatLab and in the Instructor Resource Center at **www.pearsonhighered.com/irc**.

## Companion DVD

A multimedia program on DVD accompanies student books or may be purchased separately. It is available per student or as a lab version (per work station). The DVD holds a number of supporting materials, including:

- ***ActivStats® for Data Desk.*** The award-winning *ActivStats* multimedia program supports learning chapter by chapter with the book. It complements the book with videos of real-word stories, worked examples, animated expositions of each of the major Statistics topics, and tools for performing simulations, visualizing inference, and learning to use statistics software. *ActivStats* includes 17 short video clips; 170 animated activities and teaching applets; 300 data sets; 1,000 homework exercises, many with links to Data Desk files; interactive graphs, simulations, activities for the TI-Nspire graphing calculator, visualization tools, and much more.
- ***Data Desk*** statistics package.
- ***Data.*** Data for exercises marked T are available on the DVD and website formatted for CSV, TXT, and TI calculators suitable for virtually any statistics software.
- ***Additional Chapters.*** Two additional chapters cover **Analysis of Variance** (Chapter 27) and **Multiple Regression** (Chapter 28). These topics point the way to further study in Statistics.

**ActivStats®** The award-winning ActivStats multimedia program supports learning chapter by chapter with the book and is available as a standalone DVD. It complements the book with videos of real-word stories, worked examples, animated expositions of each of the major Statistics topics, and tools for performing simulations, visualizing inference, and learning to use statistics software. ActivStats includes 17 short video clips; 170 animated activities and teaching applets; 300 data sets; 1,000 homework exercises; interactive graphs, simulations, visualization tools, and much more.

**Companion Website (www.pearsonhighered.com/bock)** provides additional resources for instructors and students.

**The Student Edition of MINITAB** is a condensed edition of the Professional release of MINITAB statistical software that offers the full range of statistical methods and graphical capabilities, along with worksheets that can include up to 10,000 data points. Individual copies of the software can be bundled with the text. (ISBN 13: 978-0-13-143661-9; ISBN-10: 0-13-143661-9)

**JMP Student Edition** is an easy-to-use, streamlined version of JMP desktop statistical discovery software from SAS Institute, Inc. and is available for bundling with the text. (ISBN 13: 978-0-321-89164-8; ISBN-10: 0-321-89164-3)

**XLStat for Pearson** is an Excel add-in that enhances the analytical capabilities of Excel. XLStat is used by leading businesses and universities around the world. Available for bundling with this text (ISBN-13: 978-0-321-75932-0; ISBN-10: 0-321-75932-X). For more information, visit **www.pearsonhighered.com/xlstat**.

# Acknowledgments

Many people have contributed to this book in all four of its editions. This edition would have never seen the light of day without the assistance of the incredible team at Pearson. Our Editor in Chief, Deirdre Lynch, was central to the genesis, development, and realization of the book from day one. Chris Cummings, Executive Editor, provided much needed support. Chere Bemelmans, Senior Content Editor, kept us on task as much as humanly possible. Sheila Spinney, Senior Production Project Manager, and Sherry Berg, Project Manager, kept the cogs from getting into the wheels where they often wanted to wander. Sonia Ashraf, Assistant Editor, and Kathleen DeChavez, Marketing Assistant, were essential in managing all of the behind-the-scenes work that needed to be done. Christine Stavrou, Manager–Media Production, put together a top-notch media package for this book. Barbara T. Atkinson, Senior Designer, and Studio Montage are responsible for the wonderful way the book looks. Debbie Rossi, Manufacturing Buyer, worked miracles to get this book and DVD in your hands, and Greg Tobin, President, EMSS, was supportive and good-humored throughout all aspects of the project. Special thanks go out to PreMedia Global, the compositor, for the wonderful work they did on this book, and in particular to Nancy Kincade, Project Manager, for her close attention to detail.

We'd also like to thank our accuracy checkers whose monumental task was to make sure we said what we thought we were saying. They are Douglas Cashing, St. Bonaventure University; Mark Littlefield, Newburyport High School; Stanley Seltzer, Ithaca College; and Susan Blackwell, First Flight High School.

We extend our sincere thanks for the suggestions and contributions made by the following reviewers, focus group participants, and class-testers:

John Arko
*Glenbrook South High School, IL*

Kathleen Arthur
*Shaker High School, NY*

Allen Back
*Cornell University, NY*

Beverly Beemer
*Ruben S. Ayala High School, CA*

Judy Bevington
*Santa Maria High School, CA*

Susan Blackwell
*First Flight High School, NC*

Gail Brooks
*McLennan Community College, TX*

Walter Brown
*Brackenridge High School, TX*

Darin Clifft
*Memphis University School, TN*

Bill Craine
*Lansing High School, NY*

Sybil Coley
*Woodward Academy, GA*

Kevin Crowther
*Lake Orion High School, MI*

Caroline DiTullio
*Summit High School, NJ*

Jared Derksen
*Rancho Cucamonga High School, CA*

Sam Erickson
*North High School, WI*

Laura Estersohn
*Scarsdale High School, NY*

Laura Favata
*Niskayuna High School, NY*

David Ferris
*Noblesville High School, IN*

Linda Gann
*Sandra Day O'Connor High School, TX*

Randall Groth
*Illinois State University, IL*

Donnie Hallstone
*Green River Community College, WA*

Howard W. Hand
*St. Marks School of Texas, TX*

Bill Hayes
*Foothill High School, CA*

Miles Hercamp
*New Palestine High School, IN*

Michelle Hipke
*Glen Burnie Senior High School, MD*

Carol Huss
*Independence High School, NC*

Sam Jovell
*Niskayuna High School, NY*

Peter Kaczmar
*Lower Merion High School, PA*

John Kotmel
*Lansing High School, NY*

Beth Lazerick
*St. Andrews School, FL*

Michael Legacy
*Greenhill School, TX*

Guillermo Leon
*Coral Reef High School, FL*

John Lieb
*The Roxbury Latin School, MA*

Mark Littlefield
*Newburyport High School, MA*

Martha Lowther
*The Tatnall School, DE*

John Maceli
*Ithaca College, NY*

Jim Miller
*Alta High School, UT*

Timothy E. Mitchell
*King Philip Regional High School, MA*

Maxine Nesbitt
*Carmel High School, IN*

Elizabeth Ann Przybysz
*Dr. Phillips High School, FL*

Diana Podhrasky
*Hillcrest High School, TX*

Rochelle Robert
*Nassau Community College, NY*

Karl Ronning
*Davis Senior High School, CA*

Bruce Saathoff
*Centennial High School, CA*

Agatha Shaw
*Valencia Community College, FL*

Murray Siegel
*Sam Houston State University, TX*

Chris Sollars
*Alamo Heights High School, TX*

Darren Starnes
*The Webb Schools, CA*

Jane Viau
*The Frederick Douglass Academy, NY*

*David Bock*
*Paul Velleman*
*Richard De Veaux*

chapter 1

# Stats Starts Here[1]

"But where shall I begin?" asked Alice. "Begin at the beginning," the King said gravely, "and go on till you come to the end: then stop."

*—Lewis Carroll, Alice's Adventures in Wonderland*

Statistics gets no respect. People say things like "You can prove anything with Statistics." People will write off a claim based on data as "just a statistical trick." And a Statistics course may not be your friends' first choice for a fun elective.

But Statistics *is* fun. That's probably not what you heard on the street, but it's true. Statistics is about how to think clearly with data. We'll talk about data in more detail soon, but for now, think of **data** as any collection of numbers, characters, images, or other items that provide information about something. Whenever there are data and a need for understanding the world, you'll find Statistics. A little practice thinking statistically is all it takes to start seeing the world more clearly and accurately.

## So, What Is (Are?) Statistics?

**Q:** What is Statistics?
**A:** Statistics is a way of reasoning, along with a collection of tools and methods, designed to help us understand the world.
**Q:** What are statistics?
**A:** Statistics (plural) are particular calculations made from data.
**Q:** So what is data?
**A:** You mean, "what *are* data?" Data is the plural form. The singular is datum.
**Q:** OK, OK, so what are data?
**A:** Data are values along with their context.

It seems every time we turn around, someone is collecting data on us, from every purchase we make in the grocery store, to every click of our mouse as we surf the Web.

Consider the following:

- If you have a Facebook account, you have probably noticed that the ads you see online tend to match your interests and activities. Coincidence? Hardly. According to the *Wall Street Journal* (10/18/2010),[2] much of your personal information has probably been sold to marketing or tracking companies. Why would Facebook give you a free account and let you upload as much as you want to its site? Because your data are valuable! Using your Facebook profile, a company might build a profile of your

[1]We could have called this chapter "Introduction," but nobody reads the introduction, and we wanted you to read this. We feel safe admitting this here, in the footnote, because nobody reads footnotes either.

[2]blogs.wsj.com/digits/2010/10/18/referers-how-facebook-apps-leak-user-ids/


*Frazz © 2013 Jef Mallett. Distributed by Universal Uclick. Reprinted with permission. All rights reserved.*

interests and activities: what movies and sports you like; your age, sex, education level, and hobbies; where you live; and, of course, who your friends are and what *they* like. From Facebook's point of view, your data are a potential gold mine. Gold ore in the ground is neither very useful nor pretty. But with skill, it can be turned into something both beautiful and valuable. What we're going to talk about in this book is how you can mine your own data and learn valuable insights about the world.

- Like many other retailers, Target stores create customer profiles by collecting data about purchases using credit cards. Patterns the company discovers across similar customer profiles enable it to send you advertising and coupons that promote items you might be particularly interested in purchasing. As valuable to the company as these marketing insights can be, some may prove startling to individuals. Recently coupons Target sent to a Minneapolis girl's home revealed she was pregnant before her father knew![3]
- How dangerous is texting while driving? Researchers at the University of Utah tested drivers on simulators that could present emergency situations. They compared reaction times of sober drivers, drunk drivers, and texting drivers.[4] The results were striking. The texting drivers actually responded more slowly and were more dangerous than those who were above the legal limit for alcohol.

In this book, you'll learn how to design and analyze experiments like this. You'll learn how to interpret data and to communicate the message you see to others. You'll also learn how to spot deficiencies and weaknesses in conclusions drawn by others that you see in newspapers and on the Internet every day. Statistics can help you become a more informed citizen by giving you the tools to understand, question, and interpret data.

**Are You a Statistic?**
The ads say, "Don't drink and drive; you don't want to be a statistic." But you can't be a statistic.

We say: "Don't be a datum."

## Statistics in a Word

It can be fun, and sometimes useful, to summarize a discipline in only a few words. So,

Economics is about . . . *Money (and why it is good).*

Psychology: *Why we think what we think (we think).*

Biology: *Life.*

Anthropology: *Who?*

History: *What, where, and when?*

Philosophy: *Why?*

Engineering: *How?*

Accounting: *How much?*

In such a caricature, Statistics is about . . . ***Variation.***

**Statistics Is about Variation**
Data vary because we don't see everything and because even what we do see and measure, we measure imperfectly.

So, in a very basic way, Essential Statistics is about the real, imperfect world in which we live.

[3]http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/

[4]"Text Messaging During Simulated Driving," Drews, F. A. et al. Human Factors: hfs.sagepub.com/content/51/5/762

Data vary. Ask different people the same question and you'll get a variety of answers. Statistics helps us to make sense of the world described by our data by seeing past the underlying variation to find patterns and relationships. This book will teach you skills to help with this task and ways of thinking about variation that are the foundation of sound reasoning about data.

## But What *Are* Data?

Amazon.com opened for business in July 1995, billing itself as "Earth's Biggest Bookstore." By 1997, Amazon had a catalog of more than 2.5 million book titles and had sold books to more than 1.5 million customers in 150 countries. In 2010, the company's sales reached $34.2 billion (a nearly 40% increase from the previous year). Amazon has sold a wide variety of merchandise, including a $400,000 necklace, yak cheese from Tibet, and the largest book in the world. How did Amazon become so successful and how can it keep track of so many customers and such a wide variety of products? The answer to both questions is *data.*

But what are data? Think about it for a minute. What exactly *do* we mean by "data"? Do data have to be numbers? The amount of your last purchase in dollars is numerical data. But your name and address in Amazon's database are also data even though they are not numerical. What about your ZIP code? That's a number, but would Amazon care about, say, the *average* ZIP code of its customers?

Let's look at some hypothetical values that Amazon might collect:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 105-2686834-3759466 | Ohio | Nashville | Kansas | 10.99 | 440 | N | B00000I5Y6 | Katherine H. |
| 105-9318443-4200264 | Illinois | Orange County | Boston | 16.99 | 312 | Y | B000002BK9 | Samuel P. |
| 105-1872500-0198646 | Massachusetts | Bad Blood | Chicago | 15.98 | 413 | N | B000068ZVQ | Chris G. |
| 103-2628345-9238664 | Canada | Let Go | Mammals | 11.99 | 902 | N | B0000010AA | Monique D. |
| 002-1663369-6638649 | Ohio | Best of Kansas | Kansas | 10.99 | 440 | N | B002MXA7Q0 | Katherine H. |

A S ***Activity*: What Is (Are) Data?** Do you really know what are data and what are just numbers?

Try to guess what they represent. Why is that hard? Because there is no *context*. If we don't know what values are measured and what is measured about them, the values are meaningless. We can make the meaning clear if we organize the values into a **data table** such as this one:

| Order Number | Name | State/Country | Price | Area Code | Previous Album Download | Gift? | ASIN | New Purchase Artist |
|---|---|---|---|---|---|---|---|---|
| 105-2686834-3759466 | Katherine H. | Ohio | 10.99 | 440 | Nashville | N | B0000015Y6 | Kansas |
| 105-9318443-4200264 | Samuel R | Illinois | 16.99 | 312 | Orange County | Y | B000002BK9 | Boston |
| 105-1372500-0198646 | Chris G. | Massachusetts | 15.98 | 413 | Bad Blood | N | B000068ZVQ | Chicago |
| 103-2628345-9238664 | Monique D. | Canada | 11.99 | 902 | Let Go | N | B0000010AA | Mammals |
| 002-1663369-6638649 | Katherine H. | Ohio | 10.99 | 440 | Best of Kansas | N | B002MXA7Q0 | Kansas |

The W's:
Who
What
and in what units
When
Where
Why
How

Now we can see that these are purchase records for album download orders from Amazon. The column titles tell what has been recorded. Each row is about a particular purchase.

What information would provide a **context**? Newspaper journalists know that the lead paragraph of a good story should establish the "Five W's": *who, what, when, where,* and (if possible) *why*. Often, we add *how* to the list as well. The answers to the first two questions are essential. If we don't know *what* values are measured and *who* those values are measured on, the values are meaningless.

## *Who* and *What*

In general, the rows of a data table correspond to individual **cases** about *Who*m (or about which—if they're not people) we record some characteristics. Cases go by different names, depending on the situation.

- Individuals who answer a survey are called **respondents**.
- People on whom we experiment are **subjects** or (in an attempt to acknowledge the importance of their role in the experiment) **participants**.
- Animals, plants, websites, and other inanimate subjects are often called **experimental units**.
- Often we simply call cases what they are: for example, *customers, economic quarters*, or *companies*.
- In a database, rows are called **records**—in this example, purchase records. Perhaps the most generic term is *cases*, but in any event the rows represent the *who* of the data.

The characteristics recorded about each individual are called **variables.** These are usually shown as the columns of a data table, and they should have a name that identifies *What* has been measured. *Name*, *Price*, *Area Code*, and whether the purchase was a *Gift* are some of the variables Amazon collected data for. Variables may seem simple, but we'll need to take a closer look soon.

We must know *who* and *what* to analyze data. Without knowing these two, we don't have enough information to start. Of course, we'd always like to know more. The more we know about the data, the more we'll understand about the world. If possible, we'd like to know the *when* and *where* of data as well. Values recorded in 1803 may mean something different than similar values recorded last year. Values measured in Tanzania may differ in meaning from similar measurements made in Mexico. And knowing *why* the data were collected can tell us much about its reliability and quality.

Often, the cases are a **sample** of cases selected from some larger **population** that we'd like to understand. Amazon certainly cares about its customers, but also wants to know how to attract all those other Internet users who may never have made a purchase from Amazon's site. To be able to generalize from the sample of cases to the larger population, we'll want the sample to be *representative* of that population—a kind of snapshot image of the larger world.

A S ***Activity*: Consider the context** . . .Can you tell who's *Who* and what's *What*? And *Why*? This activity offers real-world examples to help you practice identifying the context.

### For Example IDENTIFYING THE "WHO"

In December 2011, *Consumer Reports* published an evaluation of 25 tablets from a variety of manufacturers.

**QUESTION:** Describe the population of interest, the sample, and the *Who* of the study.

**ANSWER:** The magazine is interested in the performance of tablets currently offered for sale. It tested a sample of 25 tablets, which are the "Who" for these data. Each tablet selected represents all tablets of that model offered by that manufacturer.

# How the Data Are Collected

*How* the data are collected can make the difference between insight and nonsense. As we'll see later, data that come from a voluntary survey on the Internet are almost always worthless. One primary concern of Statistics is the design of sound methods for collecting data.[5] Throughout this book, whenever we introduce data, we'll provide a margin note listing the W's (and H) of the data. Identifying the W's is a habit we recommend.

The first step of any data analysis is to know what you are trying to accomplish and what you want to know. To help you use Statistics to understand the world and make decisions, we'll lead you through the entire process of *thinking* about the problem, *showing* what you've found, and *telling* others what you've learned. Every guided example in this book is broken into these three steps: *Think*, *Show*, and *Tell*. Identifying the problem and the *who* and *what* of the data is a key part of the *Think* step of any analysis. Make sure you know these before you proceed to *Show* or *Tell* anything about the data.

A S ***Activity*: Collect data in an experiment on yourself.** With the computer, you can experiment on yourself and then save the data. Go on to the subsequent related activities to check your understanding.

# More About Variables (*What*?)

**Privacy and the Internet** You have many Identifiers: a social security number, a student ID number, possibly a passport number, a health insurance number, and probably a Facebook account name. Privacy experts are worried that Internet thieves may match your identity in these different areas of your life, allowing, for example, your health, education, and financial records to be merged. Even online companies such as Facebook and Google are able to link your online behavior to some of these identifiers, which carries with it both advantages and dangers. The National Strategy for Trusted Identities in Cyberspace (www.wired.com/images_blogs/threatlevel/2011/04/NSTICstrategy_041511.pdf) proposes ways that we may address this challenge in the near future.

The Amazon data table displays information about several variables: *Order Number*, *Name*, *State/Country*, *Price*, and so on. These identify *what* we know about each individual. Variables such as these can play different roles, depending on how we plan to use them. While some are merely identifiers, others may be categorical or quantitative. Making that distinction is an important step in our analysis.

## Identifiers

For some variables, such as a *student ID*, each individual receives a unique value. We call a variable like this, an **identifier variable**. Identifiers are useful, but not typically for analysis.

Amazon wants to know who you are when you sign in again and doesn't want to confuse you with some other customer. So it assigns you a unique identifier. Amazon also wants to send you the right product, so it assigns a unique Amazon Standard Identification Number (ASIN) to each item it carries. Identifier variables themselves don't tell us anything useful about their categories because we know there is exactly one individual in each. You'll want to recognize when a categorical variable is playing the role of an identifier so you aren't tempted to analyze it.

## Categorical Variables

Some variables just tell us what group or category each individual belongs to. Are you male or female? Pierced or not? What color are your eyes? We call variables like these **categorical variables**.[6] Some variables are clearly categorical, like the variable *State/Country*. Its values are text and those values tell us what category the particular case falls into. Descriptive responses to questions are often categories. For example, the responses to the questions "Who is your cell phone provider?" or "What is your marital status?" yield categorical values. But numerals are often used to label categories, so categorical variable values can also be numerals. For example, Amazon collects telephone area codes that *categorize* each phone number into a geographical region. So area code is considered a categorical variable even though it has numeric values.

[5] Coming attractions: to be discussed in Part III. We sense your excitement.

[6] You may also see them called *qualitative* variables.

## Quantitative Variables

When a variable contains measured numerical values with measurement *units*, we call it a **quantitative variable**. Quantitative variables typically record an amount or degree of something. For a quantitative variable, its measurement **units** provide a meaning for the numbers. Even more important, units such as yen, cubits, carats, angstroms, nanoseconds, miles per hour, or degrees Celsius tell us the *scale* of measurement, so we know how far apart two values are. Without units, the values of a measured variable have no meaning. It does little good to be promised a raise of 5000 a year if you don't know whether it will be paid in Euros, dollars, pennies, yen, or Estonian krooni.

## Either/Or?

Some variables with numeric values can be treated as either categorical or quantitative depending on what we want to know. Amazon could record your *Age* in years. That seems quantitative, and it would be if the company wanted to know the average age of those customers who visit their site after 3 A.M. But suppose Amazon wants to decide which album to feature on its site when you visit. Then thinking of your age in one of the categories Child, Teen, Adult, or Senior might be more useful. So, sometimes whether a variable is treated as categorical or quantitative is more about the question we want to ask rather than an intrinsic property of the variable itself.

Suppose a course evaluation survey asks, "How valuable do you think this course will be to you?" 1 = Worthless; 2 = Slightly; 3 = Middling; 4 = Reasonably; 5 = Invaluable. Is *Educational Value* categorical or quantitative? A teacher might just count the number of students who gave each response for her course, treating *Educational Value* as a categorical variable. Or if she wants to see whether the course is improving, she might treat the responses as the *amount* of perceived value—in effect, treating the variable as quantitative.

But what are the units? There is certainly an *order* of perceived worth: Higher numbers indicate higher perceived worth. A course that averages 4.5 seems more valuable than one that averages 2, but the teacher will have to imagine that it has "educational value units," whatever they are. Because there are no natural units, she should be cautious. Variables that report order without natural units are often called *ordinal variables*. But saying "that's an ordinal variable" doesn't get you off the hook. You must still look to the *why* of your study and understand what you want to learn from the variable to decide whether to treat it as categorical or quantitative.

A S ***Activity*: Recognize variables measured in a variety of ways.** This activity shows examples of the many ways to measure data.

A S ***Activities*: Variables.** Several activities show you how to begin working with data in your statistics package.

### For Example IDENTIFYING "WHAT" AND "WHY" OF TABLETS

**RECAP:** A *Consumer Reports* article about 25 tablet computers lists each tablet's manufacturer, cost, battery life (hrs.), operating system (iOS/Android/RIM), and overall performance score (0–100).

**QUESTION:** Are these variables categorical or quantitative? Include units where appropriate, and describe the "Why" of this investigation.

**ANSWER:** The variables are

- manufacturer (categorical)
- cost (quantitative, $)
- battery life (quantitative, hrs.)
- operating system (categorical)
- performance score (quantitative, no units)

The magazine hopes to provide consumers with the information to choose a good tablet.

# Just Checking

In the 2004 Tour de France, Lance Armstrong made history by winning the race for an unprecedented sixth time. In 2005, he became the only 7-time winner and set a new record for the fastest average speed—41.65 kilometers per hour. A cancer survivor, Armstrong became an international celebrity. But it was all too good to be true. In 2012, following revelations of doping, the International Cycling Union stripped Armstrong of all of his titles and records and banned him from professional cycling for life.

You can find data on all the Tour de France races on the DVD. Keep in mind that the entire data set has over 100 entries.

1. List as many of the W's as you can for this data set.
2. Classify each variable as categorical or quantitative; if quantitative, identify the units.

| Year | Winner | Country of Origin | Total Time (h/min/s) | Avg. Speed (km/h) | Stages | Total Distance Ridden (km) | Starting Riders | Finishing Riders |
|---|---|---|---|---|---|---|---|---|
| 1903 | Maurice Garin | France | 94.33.00 | 25.3 | 6 | 2428 | 60 | 21 |
| 1904 | Henri Cornet | France | 96.05.00 | 24.3 | 6 | 2388 | 88 | 23 |
| 1905 | Louis Trousseller | France | 112.18.09 | 27.3 | 11 | 2975 | 60 | 24 |
| ⋮ | | | | | | | | |
| 1999 | Lance Armstrong (DQ) | USA | 91.32.16 | 40.30 | 20 | 3687 | 180 | 141 |
| 2000 | Lance Armstrong (DQ) | USA | 92.33.08 | 39.56 | 21 | 3662 | 180 | 128 |
| 2001 | Lance Armstrong (DQ) | USA | 86.17.28 | 40.02 | 20 | 3453 | 189 | 144 |
| 2002 | Lance Armstrong (DQ) | USA | 82.05.12 | 39.93 | 20 | 3278 | 189 | 153 |
| 2003 | Lance Armstrong (DQ) | USA | 83.41.12 | 40.94 | 20 | 3427 | 189 | 147 |
| 2004 | Lance Armstrong (DQ) | USA | 83.36.02 | 40.53 | 20 | 3391 | 188 | 147 |
| 2005 | Lance Armstrong (DQ) | USA | 86.15.02 | 41.65 | 21 | 3608 | 189 | 155 |
| ⋮ | | | | | | | | |
| 2011 | Cadel Evans | Australia | 86.12.22 | 39.788 | 21 | 3430 | 198 | 167 |
| 2012 | Bradley Wiggins | Great Britain | 87.34.47 | 39.928 | 20 | 3497 | 219 | 153 |
| 2013 | Chris Froome | Great Britain | 83.56.40 | 40.551 | 21 | 3404 | 219 | 170 |

**A S** ***Self-Test*: Review concepts about data.**
Like the Just Checking sections of this textbook, but interactive. (Usually, we won't reference the *ActivStats* self-tests here, but look for one whenever you'd like to check your understanding or review material.

**There's a World of Data on the Internet** These days, one of the richest sources of data is the Internet. With a bit of practice, you can learn to find data on almost any subject. Many of the data sets we use in this book were found in this way. The Internet has both advantages and disadvantages as a source of data. Among the advantages are the fact that often you'll be able to find even more current data than those we present. The disadvantage is that references to Internet addresses can "break" as sites evolve, move, and die.

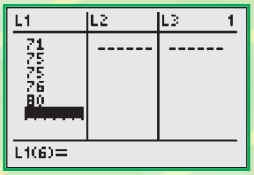
Our solution to these challenges is to offer the best advice we can to help you search for the data, wherever they may be residing. We usually point you to a website. We'll sometimes suggest search terms and offer other guidance.

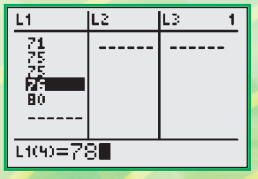
Some words of caution, though: Data found on Internet sites may not be formatted in the best way for use in statistics software. Although you may see a data table in standard form, an attempt to copy the data may leave you with a single column of values. You may have to work in your favorite statistics or spreadsheet program to reformat the data into variables. You will also probably want to remove commas from large numbers and extra symbols such as money indicators ($, ¥, £); few statistics packages can handle these.

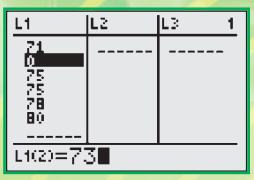
## WHAT CAN GO WRONG?

- **Don't label a variable as categorical or quantitative without thinking about the question you want it to answer.** The same variable can sometimes take on different roles.
- **Just because your variable's values are numbers, don't assume that it's quantitative.** Categories are often given numerical labels. Don't let that fool you into thinking they have quantitative meaning. Look at the context.
- **Always be skeptical.** One reason to analyze data is to discover the truth. Even when you are told a context for the data, it may turn out that the truth is a bit (or even a lot) different. Think about *how* the data were collected. People who want to influence what you think may slant the context. A survey that seems to be about all students may in fact report just the opinions of those who visited a fan website. The question that respondents answered may be posed in a way that influences responses.

## TI Tips WORKING WITH DATA





You'll need to be able to enter and edit data in your calculator. Here's how:

**TO ENTER DATA:** Hit the STAT button, and choose EDIT from the menu. You'll see a set of columns labeled L1, L2, and so on. Here is where you can enter, change, or delete a set of data.

Let's enter the heights (in inches) of the five starting players on a basketball team: 71, 75, 75, 76, and 80. Move the cursor to the space under L1, type in 71, and hit ENTER (or the down arrow). There's the first player. Now enter the data for the rest of the team.

**TO CHANGE A DATUM:** Suppose the 76" player grew since last season; his height should be listed as 78". Use the arrow keys to move the cursor onto the 76, then change the value and ENTER the correction.

**TO ADD MORE DATA:** We want to include the sixth man, 73" tall. It would be easy to simply add this new datum to the end of the list. However, sometimes the order of the data matters, so let's place this datum in numerical order. Move the cursor to the desired position (atop the first 75). Hit 2ND INS, then ENTER the 73 in the new space.

**TO DELETE A DATUM:** The 78" player just quit the team. Move the cursor there. Hit DEL. Bye.

**TO CLEAR THE DATALIST:** Finished playing basketball? Move the cursor atop the L1. Hit CLEAR, then ENTER (or down arrow). You should now have a blank datalist, ready for you to enter your next set of values.

**LOST A DATALIST?** Oops! Is L1 now missing entirely? Did you delete L1 by mistake, instead of just *clearing* it? Easy problem to fix: buy a new calculator. No? OK, then simply go to the STAT EDIT menu, and run SetUpEditor to recreate all the lists.

## What Have We Learned?

We've learned that data are information in a context.

- The W's help nail down the context *Who*, *What*, *When*, *Why*, *Where*, and *hoW*.
- We must know at least the *Who*, *What*, and *hoW* to be able to say anything useful based on the data. The *Who* are the cases. The *What* are the *variables*. A variable gives information about each of the cases. The *hoW* helps us decide whether we can trust the data.

We treat variables in two basic ways: as *categorical* or *quantitative*.

- Categorical variables identify a category for each case. Usually, we think about the counts of cases that fall into each category. (An exception is an identifier variable that just names each case.)
- Quantitative variables record measurements or amounts of something; they must have *units*.
- Sometimes we treat a variable as categorical or quantitative depending on what we want to learn from it, which means that some variables can't be pigeonholed as one type or the other. That's an early hint that in Statistics we can't always pin things down precisely.

## Terms

**Data** Systematically recorded information, whether numbers or labels, together with its context. (p. 1)

**Data table** An arrangement of data in which each row represents a case and each column represents a variable. (p. 3)

**Context** The context ideally tells *Who* was measured, *What* was measured, *How* the data were collected, *Where* the data were collected, and *When* and *Why* the study was performed. (p. 4)

**Case** A case is an individual about whom or which we have data. (*Who*). (p. 4)

**Respondent** Someone who answers, or responds to, a survey. (p. 4)

**Subject** A human experimental unit. Also called a participant. (p. 4)

**Participant** A human experimental unit. Also called a subject. (p. 4)

**Experimental unit** An individual in a study for which or for whom data values are recorded. Human experimental units are usually called subjects or participants. (p. 4)

**Record** Information about an individual in a database. (p. 4)

**Variable** A variable holds information about the same characteristic for many cases. (*What*). (p. 4)

**Sample** The cases we actually examine in seeking to understand the much larger population. (p. 4)

**Population** All the cases we wish we knew about. (p. 4)

**Identifier variable** A categorical variable that records a unique value for each case, used to name or identify it. (p. 5)

**Categorical variable** A variable that names categories (whether with words or numerals) is called categorical. (p. 5)

**Quantitative variable** A variable in which the numbers act as numerical values is called quantitative. Quantitative variables always have units. (p. 6)

**Units** A quantity or amount adopted as a standard of measurement, such as dollars, hours, or grams. (p. 6)