Eighth Edition

STATISTICS

Concepts and Controversies

DAVID S. MOORE • WILLIAM I. NOTZ

STATISTICS

Concepts and Controversies

Eighth Edition

David S. Moore
Purdue University

William I. Notz
The Ohio State University

W. H. Freeman and Company
New York

Senior Publisher:	Ruth Baruth
Acquisitions Editor:	Karen Carson
Marketing Manager:	Steve Thomas
Developmental Editors:	Anne Scanlan-Rohrer, Katrina Wilhelm
Senior Media Editor:	Laura Judge
Associate Media Editor:	Catriona Kaplan
Marketing Assistant:	Alissa Nigro
Photo Editor:	Bianca Moscatelli
Photo Researcher:	Deborah Anderson
Cover and Text Designer:	Vicki Tomaselli
Project Editor:	Elizabeth Geller
Illustrations:	Network Graphics
Illustration Coordinator:	Bill Page
Production Coordinator:	Susan Wein
Composition:	MPS Limited
Printing and Binding:	Quad Graphics

Library of Congress Control Number: 2012943500

Instructor Complimentary Copy (w/CrunchIT/EESEE Access Card):

ISBN-13: 978-1-4641-3728-0
ISBN-10: 1-4641-3728-5

Student Edition (w/CrunchIT/EESEE Access Card):

ISBN-13: 978-1-4641-2566-9
ISBN-10: 1-4641-2566-X

Printed in the United States of America

First printing

W. H. Freeman and Company
41 Madison Avenue
New York, NY 10010
Houndmills, Basingstoke RG21 6XS, England
www.whfreeman.com

BRIEF CONTENTS

*This material is optional.

CONTENTS

*This material is optional.

*This material is optional.

TO THE TEACHER: Statistics as a Liberal Discipline

Statistics: Concepts and Controversies (***SCC***) is a book on statistics as a liberal discipline, that is, as part of the general education of "non-mathematical" students. The book grew out of one of the author's experience in developing and teaching a course for freshmen and sophomores from Purdue University's School of Liberal Arts. We are pleased that other teachers have found *SCC* useful for unusually diverse audiences, extending as far as students of philosophy and medicine. This eighth edition is a revision of the text, with several new features. It retains, however, the goals of the original: to present statistics not as a technical tool but as part of the intellectual culture that educated people share.

Statistics among the liberal arts

Statistics has a widespread reputation as the least liberal of subjects. When statistics is praised, it is most often for its usefulness. Health professionals need statistics to read accounts of medical research; managers need statistics because efficient crunching of numbers will find its way to the bottom line; citizens need statistics to understand opinion polls and government statistics such as the unemployment rate and the Consumer Price Index. Because data and chance are omnipresent, our propaganda line goes, everyone will find statistics useful, and perhaps even profitable.

This is true. We would even argue that for most students the conceptual and verbal approach in *SCC* is better preparation for future encounters with statistical studies than the usual methods-oriented introduction. The joint curriculum committee of the American Statistical Association and the Mathematical Association of America recommends that any first course in statistics "emphasize the elements of statistical thinking" and feature "more data and concepts, fewer recipes and derivations." *SCC* does this, with the flavor appropriate to a liberal education: more concepts, more thinking, only simple data, fewer recipes, and no formal derivations.

There is, however, another justification for learning about statistical ideas: statistics belongs among the liberal arts. A liberal education emphasizes fundamental intellectual skills, that is, general methods of inquiry that apply in a wide variety of settings. The traditional liberal arts present such methods: literary and historical studies, the political and social analysis of human societies, the probing of nature by experimental science, the power of abstraction and deduction in mathematics. The case that statistics belongs among the liberal arts rests on the fact that reasoning from uncertain empirical data is a similarly general intellectual method. *Data*

and *chance,* the topics of this book, are pervasive aspects of our experience. Though we employ the tools of mathematics to work with data and chance, the mathematics implements ideas that are not strictly mathematical. In fact, psychologists argue convincingly that mastering formal mathematics does little to improve our ability to reason effectively about data and chance in everyday life.

SCC is shaped, as far as the limitations of the authors and the intended readers allow, by the view that statistics is an independent and fundamental intellectual method. The focus is on statistical thinking, on what others might call *quantitative literacy* or *numeracy*.

The nature of this book

There are books on statistical theory and books on statistical methods. This is neither. It is a book on statistical ideas and statistical reasoning and on their relevance to public policy and to the human sciences from medicine to sociology. We have included many elementary graphical and numerical techniques to give flesh to the ideas and muscle to the reasoning. Students learn to think about data by working with data. We have not, however, allowed technique to dominate concepts. Our intention is to teach verbally rather than algebraically, to invite discussion and even argument rather than mere computation, though some computation remains essential. The coverage is considerably broader than one might traditionally cover in a one-term course, as the table of contents reveals. In the spirit of general education, we have preferred breadth to detail.

Despite its informal nature, *SCC* is a textbook. It is organized for systematic study and has abundant exercises, many of which ask students to offer a discussion or make a judgment. Even those admirable individuals who seek pleasure in uncompelled reading should look at the exercises as well as the text. Teachers should be aware that the book is more serious than its low mathematical level suggests. The emphasis on ideas and reasoning asks more of the reader than many recipe-laden methods texts.

New in this edition

This new version of a classic text fits the current teaching environment while continuing to present statistics to "nonmathematical" readers as an aid to clear thinking in personal and professional life. The following new features and enhancements build on *SCC*'s strong pedagogical foundation:

- **Content.** The use of the Internet for surveys is increasing but is controversial. We have therefore included a discussion of Web-based surveys in Chapter 4. To help make the concept of reliability more precise, we introduce variance in Chapter 8. In Chapter 17, we now

include a mention of Bayes procedures as part of the discussion of personal probability.

- **Chapter summaries.** The "Statistics in Summary" sections at the end of each chapter now consist of two sections. One, titled "Chapter Specifics," summarizes the material presented in the chapter. The second section, titled "Link It," relates the chapter content to material in previous and upcoming chapters. The goal of this new format is to help students understand how individual chapters relate to each other and to the overall practice of statistics.
- **Examples and exercises.** Over one-third of the examples and exercises are revised to reflect current data and a variety of topics. They cover a wide range of application areas, adding interest and relevance for students. New example and exercise topics include Facebook and grades, low-fat foods and obesity, and texting.
- **Design.** The contemporary design incorporates colorful figures to aid students' understanding of text material. Sleek marginal notes invite students to explore "Statistics in Your World." Exploring the Web exercises are now labeled with a QR code icon, bringing students directly to the Book Companion Site for updated links and sources.
- **EESEE (Electronic Encyclopedia of Statistical Examples and Exercises) case studies.** Developed by The Ohio State University Department of Statistics, these electronic case studies provide students with a wide variety of timely, real examples with real data. EESEE case studies are available via an access code–protected Web site. Access codes are included with new copies of the eighth edition of *SCC*, or subscriptions can be purchased online. Over 40 new case studies have been added, and several are referenced in chapter-opening Case Studies.
- **Applets.** An applet icon signals where related, interactive statistical applets can be found on the Book Companion Site. The applets have been revised, and new applets have been created.

In addition to the new eighth-edition enhancements, *SCC* has retained the successful pedagogical features from previous editions:

- **Case Studies.** Beginning each chapter, Case Studies engage students in real-life scenarios related to the chapter concepts. The **Case Study Evaluated** section at the end of each chapter revisits the chapter-opening Case Study with follow-up questions, asking students to evaluate what they have learned from the chapter and to apply their knowledge to the Case Study.

- **Statistical Controversies.** These boxed features explore controversial topics and relate them to the chapter material. There is follow-up discussion and a proposed resolution to each of these topics in the back of the text, in the Resolving the Controversy section.

- **In the News exercises.** From popular news media outlets, these exercises use current events and cite recent data sources.
- **Now It's Your Turn exercises.** These appear after a worked example, allowing students to test their understanding. Full solutions to these exercises are provided in the back of the text.

- **Exploring the Web exercises.** Found in each chapter, these exercises point students to the Web to investigate topics and think critically about statistical data and concepts. These are now labeled with a QR code directing students to the Book Companion Site for updated links and sources.

Media and Supplements

For Students

STATSPORTAL

www.yourstatsportal.com (Access code or online purchase required.) StatsPortal is the digital gateway to *Statistics: Concepts and Controversies*, eighth edition (*SCC* 8e), and is designed to enrich the course and enhance students' study skills through a collection of Web-based tools. StatsPortal integrates a rich suite of diagnostic, assessment, tutorial, and enrichment features, enabling students to master statistics at their own pace. It is organized around three main teaching and learning components:

1. **Interactive e-Book** offers a complete and customizable online version of the text, fully integrated with all the media resources available with *SCC* 8e. The e-Book allows students to quickly search the text, highlight key areas, and add notes about what they're reading. Instructors can customize the e-Book to add, hide, and reorder content, add their own material, and highlight key text for students.
2. **Resources** organizes all the resources for *SCC* 8e into one location for ease of use. These resources include the following:
   - **Statistical Video Series** consisting of StatClips, StatClips Examples, and Statistically Speaking "Snapshots." View animated lecture videos, whiteboard lessons, and documentary-style footage that illustrate key statistical concepts and help students visualize statistics in real-world scenarios.

- **StatTutor Tutorials** offer audio-multimedia tutorials tied directly to the textbook, containing videos, applets, and animations.
- LEARNINGCurve is a formative quizzing system that offers immediate feedback at the question level to help students master course material.
- **Statistical applets** offer a series of interactive applets to help students master key statistical concepts and work exercises from the text.
- **CrunchIt!® statistical software** allows users to analyze data from any Internet location. Designed with the novice user in mind, the software is not only easily accessible but also easy to use. CrunchIt!® offers all the basic statistical routines covered in introductory statistics courses and more.
- **Stat Tutorials** are algorithmically generated quizzing with step-by-step feedback and are easily assignable for homework.
- **Stats@Work Simulations** put students in the role of statistical consultants, helping them better understand statistics interactively within the context of real-life scenarios.
- **EESEE case studies,** developed by The Ohio State University Department of Statistics, teach students to apply their statistical skills by exploring actual case studies using real data.
- **Data sets** are available in ASCII, Excel, TI, Minitab, SPSS (an IBM Company),* and JMP formats.
- **Statistical software manuals** for TI-83/84, Minitab, and Excel provide instruction, examples, and exercises using specific statistical software packages.
- **Tables**

Resources for Instructors Only

- **Instructor's Solutions Manual** includes teaching suggestions, chapter comments, and detailed solutions to all exercises.
- **Test Bank** offers hundreds of multiple-choice questions.
- **Lecture PowerPoint slides** offer a detailed lecture presentation of statistical concepts covered in each chapter of *SCC* 8e.

*SPSS was acquired by IBM in October 2009.

- SolutionMaster is a Web-based version of the solutions in the Instructor's Solutions Manual. This easy-to-use tool allows instructors to generate a solution file for any set of homework exercises. Solutions can be downloaded in PDF format for convenient printing and posting. For more information or a demonstration, contact your local W. H. Freeman sales representative.

3. **Assignments** organizes assignments and guides instructors through an easy-to-create assignment process providing access to questions from the Test Bank and exercises from the text, including many algorithmic problems. The Assignment Center enables instructors to create their own assignments from a variety of question types for machine-gradable assignments. This powerful assignment manager allows instructors to select their preferred policies in regard to scheduling, maximum attempts, time limitations, feedback, and more!

Companion Web site www.whfreeman.com/scc8e This open-access Web site includes statistical applets, data sets, self-quizzes, updated links for Web exercises, and more.

Special Software Packages Student versions of JMP, Minitab, and SPSS are available on a CD-ROM packaged with the textbook. This software is not sold separately and must be packaged with a text or a manual. Contact your local W. H. Freeman sales representative for information or visit **www.whfreeman.com.**

Video Tool Kit (Access code or online purchase required.) This new Statistical Video Series consists of three types of videos aimed to illustrate key statistical concepts and help students visualize statistics in real-world scenarios:

- **StatClips** lecture videos, created and presented by Alan Dabney, PhD, Texas A&M University, are innovative visual tutorials that illustrate key statistical concepts. In three to five minutes, each StatClips video combines dynamic animation, data sets, and interesting scenarios to help students understand the concepts in an introductory statistics course.
- In **StatClips Examples,** Alan Dabney walks students through step-by-step examples related to the StatClips lecture videos to reinforce the concepts through problem solving.
- **Snapshots** videos are abbreviated, student-friendly versions of the **Statistically Speaking** video series that bring the world of statistics into the classroom. In the same vein as the successful PBS series

Against All Odds: Inside Statistics, Statistically Speaking videos use new and updated documentary footage and interviews that show real people using data analysis to make important decisions in their careers and in their daily lives. From business to medicine, from the environment to understanding the census, Snapshots focus on why statistics is important for students' careers, and how statistics can be a powerful tool to understand their world.

Lab and Activities Manual by Dennis Pearl, The Ohio State University. This manual provides a variety of projects and exercises to help students develop a fuller appreciation of statistical concepts. It features computer lab and hands-on activities illustrating key concepts in the text, as well as additional end-of-chapter-type problems and activities. Additionally, there are exercises based on the statistical applets and EESEE Case Studies (both accessed through the book's Web site).

ISBN: 1-4641-4484-2

Software Manuals covering Minitab, Excel, and TI-83/84 are offered within StatsPortal. These manuals are also available in printed versions through custom publishing.

For Instructors

Instructor's Web site (www.whfreeman.com/scc8e) requires user registration as an instructor and features all of the student Web materials plus:

- Instructor version of **EESEE** (Electronic Encyclopedia of Statistical Examples and Exercises), with solutions to the exercises in the student version.
- **PowerPoint slides** containing all textbook figures and tables.
- **Lecture PowerPoint slides** offering a detailed lecture presentation of statistical concepts covered in each chapter of *SCC* 8e.

Printed Instructor's Solutions Manual and Test Bank includes full solutions to all exercises and hints on teaching from *SCC*. The test bank contains hundreds of multiple-choice questions to generate quizzes and tests for each chapter of the text. ISBN: 1-4292-7760-2 The test bank is also available electronically on CD-ROM (for Windows and Mac), allowing questions to be downloaded, edited, and resequenced. ISBN: 1-4292-7909-5

Enhanced Instructor's Resource CD-ROM allows instructors to **search** and **export** (by key term or chapter) the following material:

- Data sets
- All text images and tables
- Instructor's Solutions Manual and Test Bank files
- PowerPoint files and lecture slides

ISBN: 1-4292-7908-7

Course Management Systems W. H. Freeman and Company provides courses for Blackboard, WebCT (Campus Edition and Vista), Angel, Desire2Learn, Moodle, and Sakai course management systems. These are completely integrated solutions that you can easily customize and adapt to meet your teaching goals and course objectives. Visit **www.macmillanhighered.com/Catalog/other/Coursepack** for more information.

i-clicker is a two-way radio-frequency classroom response solution developed by educators for educators. University of Illinois physicists Tim Stelzer, Gary Gladding, Mats Selen, and Benny Brown created the i-clicker system after using competing classroom response solutions and discovering that they were neither classroom-appropriate nor student-friendly. Each step of i-clicker's development has been informed by teaching and learning. i-clicker is superior to other systems from both a pedagogical and a technical standpoint. To learn more about packaging i-clicker with this textbook, please contact your local W. H. Freeman sales representative or visit **www.iclicker.com.**

Acknowledgments

The staff of W. H. Freeman, especially Mary Louise Byrd, Pamela Bruton, Karen Carson, Elizabeth Geller, Bianca Moscatelli, Anne Scanlan-Rohrer, Vicki Tomaselli, Susan Wein, and Katrina Wilhelm, have done their usual excellent job in editing, designing, and producing the book. We also thank Jackie Miller for carefully checking the accuracy of the manuscript. We are grateful to many colleagues who commented on successive drafts of the manuscript.

John Deely, *Purdue University*
Linda M. Deptola, *Rappahannock Community College*
Robert Floden, *Michigan State University*
Brian Garant, *Prairie State College*
Mark A. Gebert, *University of Kentucky*
Michael Granaas, *University of South Dakota*
Brian Habing, *University of South Carolina*
Leslie Hendrix, *University of South Carolina*
Mic Jackson, *Earlham College*

Stacy Karl, *University of Minnesota*
Yoon G. Kim, *Humboldt State University*
Rose Martinez-Dawson, *Clemson University*
Philip Meyers, *Marymount Manhattan College*
Robert Adam Molnar, *Bellarmine University*
Jacquelyn O'Donohue, *Plymouth State University*
Robert Poulson, *Kansas State University*
Sarah Quesen, *West Virginia University*
Kelly Quinn, *University of Illinois at Chicago*
Mamunur Rashid, *Indiana University–Purdue University Indianapolis*
Donald Richards, *Pennsylvania State University*
Ned Schillow, *Lehigh Carbon Community College*
Dana Sylvan, *Hunter College, The City College of New York*
Dennis Wacker, *Saint Louis University*
Sheila O'Leary Weaver, *University of Vermont*
Andrew Wiesner, *Pennsylvania State University*
Hongling Yang, *University of Texas at El Paso*
Andrew Zieffler, *University of Minnesota*

We are also grateful to those who reviewed previous editions.

Marcus Agustin, *Southern Illinois University, Edwardsville*
Ma. Zenia Agustin, *Southern Illinois University, Edwardsville*
Eric Agyekum, *Malaspina University-College*
Georgiana Baker, *University of South Carolina*
Jennifer Beineke, *Western New England College*
Melody L. Boyd, *Temple University*
Patricia M. Buchanan, *Pennsylvania State University*
Barbara K. Caress, *The City College of New York*
Melissa Cass, *State University of New York at New Paltz*
Joanne Christopherson, *California State University, Fullerton*
Samuel J. Clark, *University of Washington*
Diane Conway, *Bowling Green State University*
Colette Currie, *National Louis University*
Jimmy Doi, *California Polytechnic State University*
John Dugan, *University of Idaho*
Rick L. Edgeman, *University of Idaho*
Christopher J. Ferguson, *Texas A&M International University*
Joseph Gershtenson, *Eastern Kentucky University*
Jane J. Gringauz, *Minneapolis Community and Technical College*
Timothy Grosse, *Jefferson Community College*
Debra Hall, *Indiana University–Purdue University Indianapolis*
Pamela Harman, *California State University, Los Angeles*
Richard John, *University of Southern California*
Patricia A. Kan, *Cleveland State University*
Bonnie Kegan, *University of Maryland, Baltimore*
Rasul A. Khan, *Cleveland State University*
Josh Klugman, *Temple University*
Patrick Lang, *Idaho State University*
Natalie Lochner, *Rollins College*
Ulric Lund, *California Polytechnic State University*
Megan E. Lutz, *Virginia Polytechnic Institute and State University*

Antoinette Marquard, *Cleveland State University*
Jason Martin, *Temple University*
Eric Matsuoka, *Leeward Community College*
James E. Mays, *Virginia Commonwealth University*
Michael McGill, *Virginia Polytechnic Institute and State University*
Alana Northrop, *California State University, Fullerton*
William Rayens, *University of Kentucky*
Jonathan P. Schinohofen, *Bluegrass Community and Technical College*
Majid Shahidi, *Cegep Vanier College*
Louis Soukup, *Bellevue University*
W. Scott Street IV, *Virginia Commonwealth University*
Richard Tardanico, *Florida International University*
Sharon Taylor, *Georgia Southern University*
Agnes Tuska, *California State University, Fresno*
Lewis VanBrackle, *Kennesaw State University*
Gregg G. Van Ryzin, *Rutgers, The State University of New Jersey*
C. K. Venkateswaran, *Baker College*
Nathalie Viau, *Cegep Vanier College*
Gary P. Visco, *University of Vermont*
Elizabeth Walters, *Loyola College in Maryland*
Paul Watson, *Jefferson Community College*
Bethany White, *University of Western Ontario*
George P. Yanev, *University of South Florida*
Yi Yang, *James Madison University*
Jill C. Zimmerman, *Manchester Community College*

APPLICATIONS INDEX

Example Exercise Project

Education and child development

Food science and nutrition

Manufacturing, marketing, and consumer behavior

Medicine and public health

Physical sciences

Psychology, religion, and human behavior

Public policy and political science

Sports and leisure

Transportation

PRELUDE: Making Sense of Statistics

Statistics is about data. Data are numbers, but they are not "just numbers." *Data are numbers with a context*. The number 10.5, for example, carries no information by itself. But if we hear that a friend's new baby weighed 10.5 pounds at birth, we congratulate her on the healthy size of the child. The context engages our background knowledge and allows us to make judgments. We know that a baby weighing 10.5 pounds is quite large, and that a human baby is unlikely to weigh 10.5 ounces or 10.5 kilograms. The context makes the number informative.

Statistics uses data to gain insight and to draw conclusions. The tools are graphs and calculations, but the tools are guided by ways of thinking that amount to educated common sense. Let's begin our study of statistics with a rapid and informal guide to coping with data and statistical studies in the news media and in the heat of political and social controversy. We will examine the examples introduced in this prelude in more detail later.

Data beat anecdotes

Belief is no substitute for arithmetic.

HENRY SPENCER

An anecdote is a striking story that sticks in our minds exactly because it is striking. Anecdotes humanize an issue, so news reports usually start (and often stop) with anecdotes. But anecdotes are weak ground for making up your mind—they are often misleading exactly because they are striking. Always ask if a claim is backed by data, not just by an appealing personal story.

Does living near power lines cause leukemia in children? The National Cancer Institute spent 5 years and $5 million gathering data on the question. Result: no connection between leukemia and exposure to magnetic fields of the kind produced by power lines. The editorial that accompanied the study report in the *New England Journal of Medicine* thundered, "It is time to stop wasting our research resources" on the question.

Now compare the impact of a television news report of a 5-year, $5 million investigation with that of a televised interview with an articulate mother whose child has leukemia and who happens to live near a power line. In the public mind, the anecdote wins every time. Be skeptical. Data are more reliable than anecdotes because they systematically describe an overall picture rather than focus on a few incidents.

We are tempted to add, "Data beat self-proclaimed experts." The idea of balance held by much of the news industry is to present a quick statement by an "expert" on either side. We never learn that one expert expresses the

consensus of an entire field of science, while the other is a quack with a special-interest axe to grind. As a result of the media's taste for conflict, the public now thinks that for every expert there is an equal and opposite expert. If you really care about an issue, try to find out what the data say and how good the data are. Many issues do remain unsettled, but many others are unsettled only in the minds of people who don't care about evidence. You can start by looking at the credentials of the "experts" and at whether the studies they cite have appeared in journals that require careful outside review before they publish a claim.

Where the data come from is important

Figures won't lie but liars will figure.

CHARLES GROSVENOR

Data are numbers, and numbers always seem solid. Some are and some are not. Where the data come from is the single most important fact about any statistical study. When Ann Landers asked readers of her advice column whether they would have children again and 70% of those who replied shouted "No," readers should have just amused themselves with Ann's excerpts from tear-stained letters describing what beasts the writers' children are. Ann Landers was in the entertainment business. Her invitation attracted parents who regretted having their children. Most parents don't regret having children. We know this because opinion polls have asked large numbers of parents, chosen at random to avoid attracting one opinion or another. Opinion polls have their problems, as we will see, but they beat just asking upset people to write in.

Even the most reputable publications have not been immune to bad data. The *Journal of the American Medical Association* once printed an article claiming that pumping refrigerated liquid through tubes in the stomach relieves ulcers. The patients did respond, but only because patients often respond to *any* treatment given with the authority of a trusted doctor. That is, placebos (dummy treatments) work. When a skeptic finally tried a properly controlled study in which some patients got the tube and some got a placebo, the placebo actually did a bit better. "No comparison, no conclusion" is a good starting point for judging medical studies. We would be skeptical about the ongoing interest in "natural remedies," for example. Few of these have passed a comparative trial to show that they are more than just placebos sold in bottles bearing pretty pictures of plants.

Beware the lurking variable

I have enough money to last me the rest of my life, unless I buy something.

JACKIE MASON

You read that crime is higher in counties with gambling casinos. A college teacher says that students who took a course online did better than the students in the classroom. Government reports emphasize that well-educated people earn a lot more than people with less education. Don't jump to conclusions. Ask first, "What is there that they didn't tell me that might explain this?"

Crime is higher in counties with casinos, but it is also higher in urban counties and in poor counties. What kind of counties are casinos in? Did these counties have high crime rates before the casino arrived? The online students did better, but they were older and better prepared than the in-class students. No wonder they did better. Well-educated people do earn a lot. But educated people have (on the average) parents with more education and more money than the parents of poorly educated people have. They grew up in nicer places and went to better schools. These advantages help them get more education and would help them earn more even without that education.

All these studies report a connection between two variables and invite us to conclude that one of these variables influences the other. "Casinos increase crime" and "Stay in school if you want to be rich" are the messages we hear. Perhaps these messages are true. But perhaps much of the connection is explained by other variables lurking in the background, such as the nature of counties that accept casinos and the advantages that highly educated people were born with. Good statistical studies look at lots of background variables. This is tricky, but you can at least find out if it was done.

Variation is everywhere

When the facts change, I change my mind. What do you do, sir?

JOHN MAYNARD KEYNES

If a thermometer under your tongue reads higher than 98.6°F, do you have a fever? Maybe not. People vary in their "normal" temperature. Your own temperature also varies—it is lower around 6 A.M. and higher around 6 P.M. The government announces that the unemployment rate rose a tenth of a percent last month and that new home starts fell by 3%. The stock market promptly jumps (or sinks). Stocks are jumpier than is sensible. The government data come from samples that give good estimates but not the exact truth. Another run of the same samples would give slightly different answers. And economic facts jump around anyway, due to weather, strikes, holidays, and all sorts of other reasons.

Many people join the stock market in overreacting to minor changes in data that are really nothing but background noise. Here is Arthur

Nielsen, head of the country's largest market research firm, describing his experience:

> Too many business people assign equal validity to all numbers printed on paper. They accept numbers as representing Truth and find it difficult to work with the concept of probability. They do not see a number as a kind of shorthand for a range that describes our actual knowledge of the underlying condition.

Variation is everywhere. Individuals vary; repeated measurements on the same individual vary; almost everything varies over time. Ignore the pundits who try to explain the deep reasons behind each day's stock market moves, or who condemn a team's ability and character after a game decided by a last-second shot that did or didn't go in.

Conclusions are not certain

As far as the laws of mathematics refer to reality they are not certain, and as far as they are certain they do not refer to reality.

ALBERT EINSTEIN

Because variation is everywhere, statistical conclusions are not certain. Most women who reach middle age have regular mammograms to detect breast cancer. Do mammograms really reduce the risk of dying of breast cancer? Statistical studies of high quality find that mammograms reduce the risk of death in women aged 50 to 64 years by 26%. That's an average over all women in the age group. Because variation is everywhere, the results are different for different women. Some women who have mammograms every year die of breast cancer, and some who never have mammograms live to 100 and die when they crash their motorcycles.

What the summary study actually said was "mammography reduces the risk of dying of breast cancer by 26 percent (95 percent confidence interval, 17 to 34 percent)." That 26% is, in Arthur Nielsen's words, "shorthand for a range that describes our actual knowledge of the underlying condition." The range is 17% to 34%, and we are 95% confident that the truth lies in that range. We're pretty sure, in other words, but not certain. Once you get beyond news reports, you can look for phrases like "95% confident" and "statistically significant" that tell us that a study did produce findings that, while not certain, are pretty sure.

Data reflect social values

It's easy to lie with statistics. But it is easier to lie without them.

FREDERICK MOSTELLER

Good data do beat anecdotes. Data are more objective than anecdotes or loud arguments about what might happen. Statistics certainly lies on the factual, scientific, rational side of public discourse. Statistical studies deserve more weight than most other evidence about controversial issues. There is, however, no such thing as perfect objectivity. Statistics shares a social context that influences what we decide to measure and how we measure it.

Suicide rates, for example, vary greatly among nations. It appears that much of the difference in the reported rates is due to social attitudes rather than to actual differences in suicide rates. Counts of suicides come from death certificates. The officials who complete the certificates (details vary depending on the state or nation) can choose to look more or less closely at, for example, drownings and falls that lack witnesses. Where suicide is stigmatized, deaths are more often reported as accidents. Countries that are predominantly Catholic have lower reported suicide rates than others, for example. Japanese culture has a tradition of honorable suicide as a response to shame. This tradition leads to better reporting of suicide in Japan because it reduces the stigma attached to suicide. In other nations, changes in social values may lead to higher suicide counts. It is becoming more common to view depression as a medical problem rather than a weakness of character and suicide as a tragic end to the illness rather than a moral flaw. Families and doctors then become more willing to report suicide as the cause of death.

Social values influence data on matters less sensitive than suicide. The percentage of people who are unemployed in the United States is measured each month by the Bureau of Labor Statistics, using a large and very professionally chosen sample of people across the country. But what does it mean to be "unemployed"? It means that you don't have a job even though you want a job and have *actively looked for work in the last four weeks*. If you went four weeks without seeking work, you are not unemployed; you are "out of the labor force." This definition of unemployment reflects the value we attach to working. A different definition might give a very different unemployment rate.

Our point is not that you should mistrust the unemployment rate. The definition of "unemployment" has been stable over time, so that we can see trends. The definition is reasonably consistent across nations, so that we can make international comparisons. The data are produced by professionals free of political interference. The unemployment rate is important and useful information. Our point is that not everything important can be reduced to numbers and that reducing things to numbers is done by people influenced by many pressures, conscious and unconscious.

STATISTICS AND YOU: What Lies ahead in This Book

This isn't a book about the tools of statistics. It is a book about statistical ideas and their impact on everyday life, public policy, and many different fields of study. You will learn some tools, of course. Life will be easier if you have in hand *a calculator with built-in statistical functions*. Specifically, you need a calculator that will find means, standard deviations, and correlations. Look for a calculator that claims to do "two-variable statistics" or mentions "correlation." If you have access to a computer with statistical software, so much the better. On the other hand, you need little formal mathematics. If you can read and use simple equations, you are in good shape. Be warned, however, that you will be asked to think. Thinking exercises the mind more deeply than following mathematical recipes. *Statistics: Concepts and Controversies* presents statistical ideas in four parts:

I. Data production describes methods for producing data that can give clear answers to specific questions. Where the data come from really is important—basic concepts about how to select samples and design experiments are the most influential ideas in statistics.

II. Data analysis concerns methods and strategies for exploring, organizing, and describing data using graphs and numerical summaries. You can learn to look at data intelligently even with quite simple tools.

III. Probability is the language we use to describe chance, variation, and risk. Because variation is everywhere, probabilistic thinking helps separate reality from background noise.

IV. Statistical inference moves beyond the data in hand to draw conclusions about some wider universe, taking into account that variation is everywhere and that conclusions are uncertain.

Ultimately, data are used to draw conclusions or make decisions. The process of reasoning from data consists of several steps that yield a case for the validity of the final conclusion. Each part of this book discusses issues that affect the quality of the steps in this process. It is easy to focus on mastering the details in each chapter and lose track of how these details contribute to the overall argument. To help you see how the individual chapters fit into the overall argument, we end each chapter with a section that we call "Link It," which briefly describes how the contents of the chapter fit into the overall reasoning process. You will find this section immediately before the "Statistics in Summary" subsection.

Statistical ideas and tools emerged only slowly from the struggle to work with data. Two centuries ago, astronomers and surveyors faced the problem of combining many observations that, despite the greatest care, did not exactly match. Their efforts to deal with variation in their data

produced some of the first statistical tools. As the social sciences emerged in the 19th century, old statistical ideas were transformed and new ones were invented to describe the variation in individuals and societies. The study of heredity and of variable populations in biology brought more advance. The first half of the 20th century gave birth to statistical designs for producing data and to statistical inference based on probability. By midcentury it was clear that a new discipline had been born. As all fields of study place more emphasis on data and increasingly recognize that variability in data is unavoidable, statistics has become a central intellectual method. Every educated person should be acquainted with statistical reasoning. Reading this book will enable you to make that acquaintance.

ABOUT THE AUTHORS

David S. Moore is Shanti S. Gupta Distinguished Professor of Statistics, Emeritus, at Purdue University and was 1998 president of the American Statistical Association. He received his AB from Princeton University and his PhD from Cornell University, both in mathematics. He has written many research papers in statistical theory and served on the editorial boards of several major journals. Professor Moore is an elected fellow of the American Statistical Association and of the Institute of Mathematical Statistics and an elected member of the International Statistical Institute. He has served as program director for statistics and probability at the National Science Foundation.

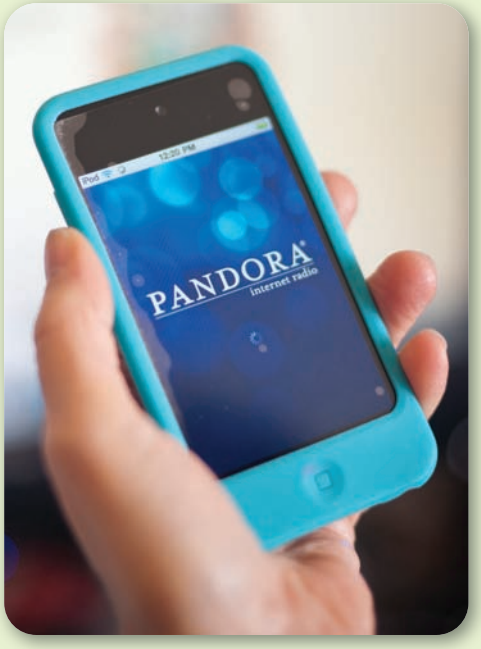
Professor Moore has devoted much of his career to the teaching of statistics. He was the content developer for the Annenburg/Corporation for Public Broadcasting college-level telecourse *Against All Odds: Inside Statistics* and for the series of video modules *Statistics: Decisions through Data,* intended to aid the teaching of statistics in schools. He is the author of influential articles on statistical education and of several leading textbooks. Professor Moore has served as president of the International Association for Statistical Education and has received the Mathematical Association of America's national award for distinguished college or university teaching of mathematics.

William I. Notz is Professor of Statistics at The Ohio State University. He received his BS in physics from Johns Hopkins University and his PhD in mathematics from Cornell University. His first academic job was as an assistant professor in the Department of Statistics at Purdue University. While there, he taught the introductory concepts course with Professor Moore, using the first edition of *Statistics: Concepts and Controversies.* As a result of this experience he developed an interest in statistical education. Professor Notz is a coauthor of EESEE (the Electronic Encyclopedia of Statistical Examples and Exercises) and has coauthored several textbooks.

Professor Notz's research interests have focused on experimental design and computer experiments. He is the author of several research papers and of a book on the design and analysis of computer experiments. He is an elected fellow of the American Statistical Association and an elected member of the International Statistical Institute. He has served as the editor of the journals *Technometrics* and *Journal of Statistics Education,* as well as on the editorial boards of several journals. At The Ohio State University, he has served as the Director of the Statistical Consulting Service, as acting chair of the Department of Statistics, and as associate dean in the College of Mathematical and Physical Sciences. He is a winner of The Ohio State University's Alumni Distinguished Teaching Award.

PART I

Producing Data



Richard B. Levine/Newscom

You and your friends are not typical. What you listen to on the radio, for example, is probably not what we listen to. Of course, we and our friends are also not typical. To get a true picture of the country as a whole (or even of college students), we must recognize that the picture may not resemble us or what we see around us. We need *data*. Data from Arbitron (a media research firm) show that the most popular radio formats are adult contemporary (a weekly cumulative audience of 72 million in Autumn 2009) and country (a weekly cumulative audience of 64 million). If you like pop contemporary hit radio (a weekly cumulative audience of 59 million) and we like all news (a weekly cumulative audience of only 15 million), we may have no clue about the tastes of radio audiences as a whole. If we are in the broadcasting business, or even if we are interested in pop culture, we must put our own tastes aside and look at the data.

You can find data in the library or on the Internet (that's where we found the radio format data). But how can we know whether data can be trusted? Good data are as much a human product as wool sweaters and tablet PCs. Sloppily produced data will frustrate you as much as a sloppily made sweater. You examine a sweater before you buy, and you don't buy if it is not well made. Neither should you use data that are not well made. The first part of this book shows how to tell if data are well made.

this page left intentionally blank

CHAPTER 1

Where Do Data Come From?

CASE STUDY You can read the newspaper and watch TV news for months without seeing an algebraic formula. No wonder algebra seems unconnected to life. You can't go a day, however, without meeting data and statistical studies. You hear that last month's unemployment rate was 10.0%. A news article says that 70% of people aged 18 to 24 believe that downloading music from the Internet is no different from buying a used CD or recording music borrowed from a friend, as opposed to 36% of people 65 or older. A longer article says that low-income children who received high-quality day care did better on academic tests given years later and were more likely to go to college and hold good jobs than other similar children.

Where do these data come from? Why can we trust them? Or maybe we can't trust them. Good data are the fruit of intelligent human effort. Bad data result from laziness or lack of understanding, or even the desire to mislead others. "Where do the data come from?" is the first question you should ask when someone throws a number at you.

Hemera/Thinkstock/Getty Images

During the 2008 presidential election, "Fork Over Your Vote" was a national project in which people chose who they thought should be the president by tossing a jelly bean into a jar. At the Coffee Scene in Pembroke Pines, Florida, customers got to choose either a blue or a red jelly bean with every purchase. Blueberry jelly beans were votes for Democrat Barack Obama. Cherry-strawberry candies were for McCain.

More than 100 restaurants in 34 states were involved in the project. Votes were counted every Monday. As of October 13, 2008, McCain had 53.5% of the national jelly bean vote. More than 211,000 bean ballots had been cast. The same jelly bean vote in the 2004 election correctly predicted President George W. Bush would win the presidency. Unfortunately for John McCain, the poll incorrectly predicted a victory for McCain in 2008 and it will be interesting to see if the project is repeated in the 2012 election. No information was available in January 2012 when we last checked.

What can we say about data from this poll? By the end of this chapter you will have learned some basic questions to ask about the data from the jelly bean poll. The answers to these questions will help us assess whether the data from the jelly bean poll are good or bad, as we will explore further in Chapter 2. ■

Talking about data: individuals and variables

Statistics is the science of data. We could almost say "the art of data," because good judgment and even good taste along with good math make good statistics. A big part of good judgment lies in deciding what you must measure in order to produce data that will shed light on your concerns. We begin with some vocabulary to describe the raw materials that go into data.

Individuals and variables

Individuals are the objects described by a set of data. Individuals may be people, but they may also be animals or things.

A **variable** is any characteristic of an individual. A variable can take different values for different individuals.

For example, here are the first lines of a professor's data set at the end of a statistics course:

NAME	MAJOR	POINTS	GRADE
ADVANI, SURA	COMM	397	B
BARTON, DAVID	HIST	323	C
BROWN, ANNETTE	LIT	446	A
CHIU, SUN	PSYC	405	B
CORTEZ, MARIA	PSYC	461	A

The *individuals* are students enrolled in the course. In addition to each student's name, there are three *variables*. The first says what major a student has chosen. The second variable gives the student's total points out of 500 for the course, and the third records the grade received.

Statistics deals with numbers, but not all variables are numerical. Some are "categorical" and simply place an individual into one of several groups or categories. Of the three variables in the professor's data set, only total points has numbers as its values. Major and grade are categorical and to do statistics with these variables, we use *counts* or *percentages*. We might give the percentage of students who got an A, for example, or the percentage who are psychology majors.

Bad judgment in choosing variables can lead to data that cost lots of time and money but don't shed light on the world. What constitutes good judgment can be controversial. Here are examples of the challenges in deciding what data to collect.

EXAMPLE 1 Who recycles?

Who takes the trouble to recycle? Researchers spent lots of time and money weighing the stuff put out for recycling in two neighborhoods in a California city; call them Upper Crust and Lower Mid. The *individuals* here are households, because trash and recycling pickup are done for residences, not for people one at a time. The *variable* measured was the weight in pounds of the curbside recycling basket each week.

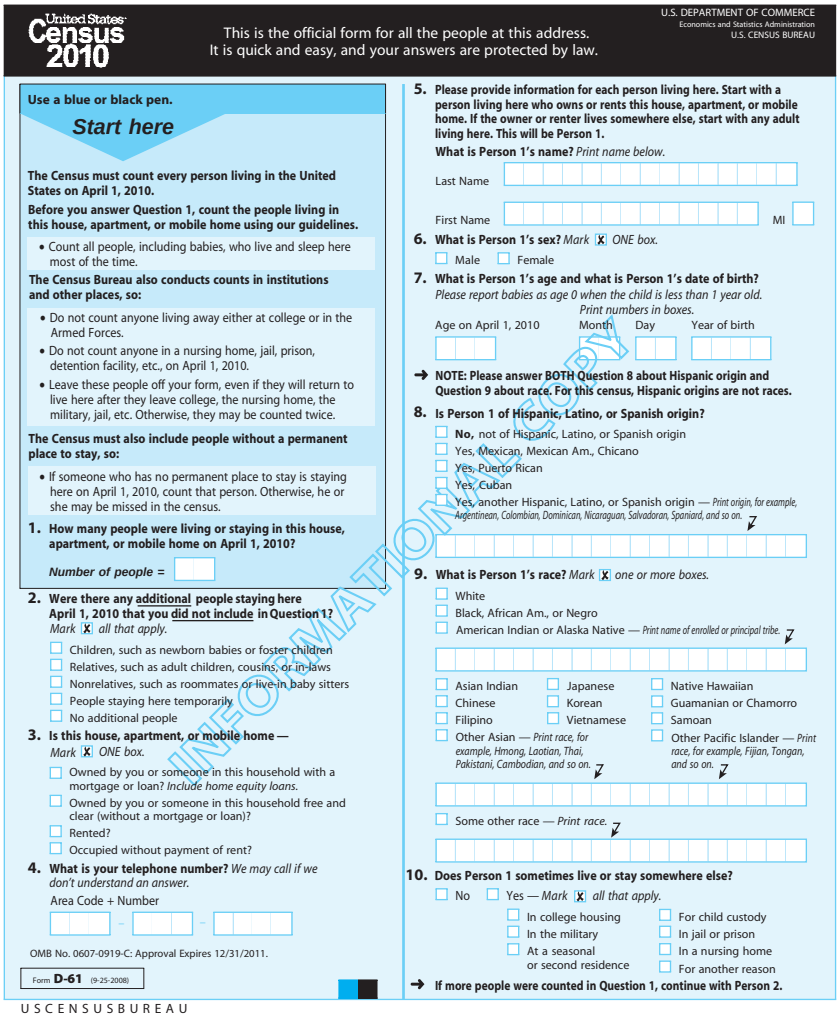
The Upper Crust households contributed more pounds per week on the average than did the folk in Lower Mid. Can we say that the rich are more serious about recycling? No. Someone noticed that Upper Crust recycling baskets contained lots of heavy glass wine bottles. In Lower Mid, they put out lots of light plastic soda bottles and light metal beer and soda cans. The conclusion: weight tells us little about commitment to recycling.

EXAMPLE 2 What's your race?

The U.S. census asks, "What is this person's race?" for every person in every household. "Race" is a *variable,* and the Census Bureau must say exactly how to measure it. The census form does this by giving a list of races. Years of political squabbling lie behind this list.

How many races shall we list, and what names shall we use for them? Shall we have a category for people of mixed race? Asians wanted more national categories, such as Filipino and Vietnamese, for the growing Asian population. Pacific Islanders wanted to be separated from the larger Asian group. Black leaders did not want a mixed-race category, fearing that many blacks would choose it and so reduce the official count of the black population.

The 2010 census form (see Figure 1.1) ended up with six Asian groups (plus "Other Asian") and three Pacific Island groups (plus "Other Pacific Islander"). There is no "mixed race" group, but you can mark more than one race. That is, people claiming mixed race can count as both, so that the total of the racial group counts in 2010 is larger than the population count. Unable to decide what the proper term for blacks should be, the Census Bureau settled on "Black, African American, or Negro." What about Hispanics? That's a separate question, because Hispanics can be of



United States Census 2010

This is the official form for all the people at this address.
It is quick and easy, and your answers are protected by law.

U.S. DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. CENSUS BUREAU

Use a blue or black pen.

Start here

The Census must count every person living in the United States on April 1, 2010.

Before you answer Question 1, count the people living in this house, apartment, or mobile home using our guidelines.

- Count all people, including babies, who live and sleep here most of the time.

The Census Bureau also conducts counts in institutions and other places, so:

- Do not count anyone living away either at college or in the Armed Forces.
- Do not count anyone in a nursing home, jail, prison, detention facility, etc., on April 1, 2010.
- Leave these people off your form, even if they will return to live here after they leave college, the nursing home, the military, jail, etc. Otherwise, they may be counted twice.

The Census must also include people without a permanent place to stay, so:

- If someone who has no permanent place to stay is staying here on April 1, 2010, count that person. Otherwise, he or she may be missed in the census.

1. How many people were living or staying in this house, apartment, or mobile home on April 1, 2010?

Number of people =

2. Were there any additional people staying here April 1, 2010 that you did not include in Question 1? *Mark [X] all that apply.*

- ☐ Children, such as newborn babies or foster children
- ☐ Relatives, such as adult children, cousins, or in-laws
- ☐ Nonrelatives, such as roommates or live-in baby sitters
- ☐ People staying here temporarily
- ☐ No additional people

3. Is this house, apartment, or mobile home — *Mark [X] ONE box.*

- ☐ Owned by you or someone in this household with a mortgage or loan? *Include home equity loans.*
- ☐ Owned by you or someone in this household free and clear (without a mortgage or loan)?
- ☐ Rented?
- ☐ Occupied without payment of rent?

4. What is your telephone number? *We may call if we don't understand an answer.*

Area Code + Number

OMB No. 0607-0919-C: Approval Expires 12/31/2011.

Form **D-61** (9-25-2008)

5. Please provide information for each person living here. Start with a person living here who owns or rents this house, apartment, or mobile home. If the owner or renter lives somewhere else, start with any adult living here. This will be Person 1.

What is Person 1's name? *Print name below.*

Last Name

First Name MI

6. What is Person 1's sex? *Mark [X] ONE box.*

☐ Male ☐ Female

7. What is Person 1's age and what is Person 1's date of birth?

Please report babies as age 0 when the child is less than 1 year old.

Print numbers in boxes.

Age on April 1, 2010 | Month | Day | Year of birth

➜ **NOTE: Please answer BOTH Question 8 about Hispanic origin and Question 9 about race. For this census, Hispanic origins are not races.**

8. Is Person 1 of Hispanic, Latino, or Spanish origin?

- ☐ **No,** not of Hispanic, Latino, or Spanish origin
- ☐ Yes, Mexican, Mexican Am., Chicano
- ☐ Yes, Puerto Rican
- ☐ Yes, Cuban
- ☐ Yes, another Hispanic, Latino, or Spanish origin — *Print origin, for example, Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.*

9. What is Person 1's race? *Mark [X] one or more boxes.*

- ☐ White
- ☐ Black, African Am., or Negro
- ☐ American Indian or Alaska Native — *Print name of enrolled or principal tribe.*
- ☐ Asian Indian
- ☐ Chinese
- ☐ Filipino
- ☐ Japanese
- ☐ Korean
- ☐ Vietnamese
- ☐ Other Asian — *Print race, for example, Hmong, Laotian, Thai, Pakistani, Cambodian, and so on.*
- ☐ Native Hawaiian
- ☐ Guamanian or Chamorro
- ☐ Samoan
- ☐ Other Pacific Islander — *Print race, for example, Fijian, Tongan, and so on.*
- ☐ Some other race — *Print race.*

10. Does Person 1 sometimes live or stay somewhere else?

☐ No ☐ Yes — *Mark [X] all that apply.*

- ☐ In college housing
- ☐ In the military
- ☐ At a seasonal or second residence
- ☐ For child custody
- ☐ In jail or prison
- ☐ In a nursing home
- ☐ For another reason

➜ **If more people were counted in Question 1, continue with Person 2.**

U S C E N S U S B U R E A U

Figure 1.1 The first page of the 2010 census form, mailed to all households in the country. The 2010 census form can be found online at **2010.census.gov/2010census/about/interactive-form.php**.

any race. Again unable to choose a short name that would satisfy everyone, the Census Bureau asked if you are of "Hispanic, Latino, or Spanish origin."

The fight over "race" reminds us that data reflect society. Race is a social idea, not a biological fact. In the census, you say what race you consider yourself to be. Race is a sensitive issue in the United States, so the fight is no surprise and the Census Bureau's diplomacy seems a good compromise.

Observational studies

As Yogi Berra, the former catcher and manager of the New York Yankees who is renowned for his humorous quotes, said, "You can observe a lot by watching." Sometimes all you can do is watch. To learn how chimpanzees in the wild behave, watch. To study how a teacher and young children interact in a schoolroom, watch. It helps if the watcher knows what to look for. The chimpanzee expert may be interested in how males and females interact, in whether some chimps in the troop are dominant, in whether the chimps hunt and eat meat. Indeed, chimps were thought to be vegetarians until Jane Goodall watched them carefully in Gombe National Park, Tanzania. Now it is clear that meat is a natural part of the chimpanzee diet.

At first, the observer may not know what to record. Eventually, patterns seem to emerge and we can decide what variables we want to measure. How often do chimpanzees hunt? Alone or in groups? How large are hunting groups? Males alone, or both males and females? How much of the diet is meat? Observation that is organized and measures clearly defined variables is more convincing than just watching. Here is an example of highly organized (and expensive) observation.

EXAMPLE 3 Do power lines cause leukemia in children?

Graham Bell/Alamy

Electric currents generate magnetic fields. So living with electricity exposes people to magnetic fields. Living near power lines increases exposure to these fields. Really strong fields can disturb living cells in laboratory studies. What about the weaker fields we experience if we live near power lines? Some data suggested that more children in these locations might develop leukemia, a cancer of the blood cells.

We can't do experiments that deliberately expose children to magnetic fields for weeks and months at a time. It's hard to compare cancer rates among children who happen to live in more and less exposed locations because leukemia is quite rare and locations vary a lot in many ways other than magnetic fields. It is easier to start with children who have leukemia and compare them with children who don't. We can look at lots of possible causes—diet, pesticides, drinking water, magnetic fields, and others—to see where children with leukemia differ from those without. Some of these broad studies suggested a closer look at magnetic fields.

A really careful look at magnetic fields took five years and cost $5 million. The researchers compared 638 children who had leukemia and 620 who did not. They went into the homes and actually measured the magnetic fields in the children's bedrooms, in other rooms, and at the front door. They recorded facts about nearby power lines for the family home and also for the mother's residence when she was pregnant. Result: no evidence of more than a chance connection between magnetic fields and childhood leukemia.

"No evidence" that magnetic fields are connected with childhood leukemia doesn't prove that there is no risk. It says only that a very careful study could not find any risk that stands out from the play of chance that distributes leukemia cases across the landscape. In other words, the study could not rule out chance as a plausible explanation for what was observed. Critics continue to argue that the study failed to measure some important variables or that the children studied don't fairly represent all children. Nonetheless, a carefully designed observational study is a great advance over haphazard and sometimes emotional counting of cancer cases.

Response variable and observational study

A **response** is a variable that measures an outcome or result of a study. An **observational study** observes individuals and measures variables of interest but does not intervene in order to influence the responses. The purpose of an observational study is to describe some group or situation.

You just don't understand A sample survey of journalists and scientists found quite a communications gap. Journalists think that scientists are arrogant, while scientists think that journalists are ignorant. We won't take sides, but here is one interesting result from the survey: 82% of the scientists agree that the "media do not understand statistics well enough to explain new findings" in medicine and other fields.

Sample surveys

You don't have to eat the entire pot of soup to know it needs more salt. That is the idea of sampling: to gain information about the whole by examining only a part. **Sample surveys** are an important kind of observational study. They survey some group of individuals by studying only some of its members, selected not because they are of special interest but because they represent the larger group. Here is the vocabulary we use to discuss sampling.

Populations and samples

The **population** in a statistical study is the entire group of individuals about which we want information.

A **sample** is the part of the population from which we actually collect information and is used to draw conclusions about the whole.

Notice that the *population* is the group we want to study. If we want information about all U.S. college students, that is our population even if students at only one college are available for sampling. To make sense of any sample result, you must know what population the sample represents. Did a preelection poll, for example, ask the opinions of all adults? Or citizens only? Registered voters only? Democrats only? The *sample* consists of the people we actually have information about. If the poll can't contact some of the people it selected, those people aren't in the sample.

The distinction between population and sample is basic to statistics. The following examples illustrate this distinction and also introduce some major uses of sampling. These brief descriptions also indicate the variables measured for each individual in the sample.

EXAMPLE 4 Public opinion polls

Polls such as those conducted by Gallup and many news organizations ask people's opinions on a variety of issues. The *variables* measured are responses to questions about public issues. Though most noticed at election time, these polls are conducted on a regular basis throughout the year. For a typical opinion poll:

Population: U.S. residents 18 years of age and over. Noncitizens and even illegal immigrants are included.

Sample: Between 1000 and 1500 people interviewed by telephone.

EXAMPLE 5 The Current Population Survey

Government economic and social data come from large sample surveys of a nation's individuals, households, or businesses. The monthly Current Population Survey (CPS) is the most important government sample survey in the United States. Many of the *variables* recorded by the CPS concern the employment or unemployment of everyone over 16 years old in a household. The government's monthly unemployment rate comes from