



**USED**

**DAVID S. MOORE**

**The Basic Practice  
of Statistics**

**SECOND EDITION**




The cover features a collection of colorful M&M's candies in various colors (red, orange, green, purple, yellow, black) scattered across the page. Each candy is outlined with a red hand-drawn scribble. The candies are set against a background of purple and white, with some areas featuring a black and white halftone pattern. The title 'The Basic Practice of Statistics' is written in a blue serif font, and 'SECOND EDITION' is in a red sans-serif font. The author's name 'DAVID S. MOORE' is in a large red sans-serif font, and 'Purdue University' is in a smaller black sans-serif font. The publisher's name 'W. H. Freeman and Company' and 'New York' are in a black sans-serif font. The overall design is vibrant and artistic.

# The Basic Practice of Statistics

SECOND EDITION

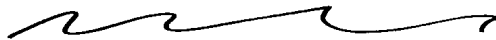
**DAVID S. MOORE**

Purdue University

 W. H. Freeman and Company  
New York

Acquisitions editor: Patrick Farace  
Publisher: Michelle Russel Julet  
Marketing manager: Kimberly Manzi  
Development editor: Don Gecewicz  
New media producer: Brian J. McCooey  
Project editor: Diane Cimino Maass  
Cover and text designer: Vicki Tomaselli  
Cover and interior illustrations: Mark Chickinelli  
Production coordinator: Julia DeRosa  
Illustration and composition: Publication Services  
Manufacturing: RR Donnelley & Sons Company

"M&M'S" CHOCOLATE CANDIES IS A REGISTERED TRADEMARK OF MARS, INCORPORATED  
AND USED WITH PERMISSION OF THE OWNER. COPYRIGHT ©MARS, INC.



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

Moore, David S.  
The basic practice of statistics / David S. Moore—2d ed.  
p. cm.  
Includes index.  
ISBN 0-7167-3627-6  
1. Statistics. I. Title.  
QA276.12.M648 2000  
519.5—dc21

99-18033

CIP

©1995, 2000 by W H. Freeman and Company. All rights reserved.

No part of this book may be reproduced by any mechanical, photographic, or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted, or otherwise copied for public or private use, without written permission from the publisher.

Printed in the United States of America

Second printing 2000

# PREFACE

*The Basic Practice of Statistics* (BPS) is an introduction to statistics for students in two-year and four-year colleges and universities that emphasizes working with data and statistical ideas. In this preface I describe for instructors the nature and features of the book.

## Guiding principles

Once upon a time, basic statistics courses emphasized only probability and inference. BPS reflects the current standard, in which data analysis and the design of data production join probability-based inference as major content areas. Statisticians have reached a broad consensus on the nature of first courses for general college audiences. As Richard Schaffer says in discussing a survey paper of mine, “With regard to the content of an introductory statistics course, statisticians are in closer agreement today than at any previous time in my career.”<sup>1</sup> Figure 1 is an outline of the consensus as summarized by the joint curriculum committee of the American Statistical Association and the Mathematical Association of America.<sup>2</sup>

I was a member of the ASA/MAA committee, and I agree with their conclusions. Fostering active learning is the business of the teacher (though an emphasis on working with data helps). BPS is guided by the first two recommendations. Although the book is elementary in the level of mathematics

1. **Emphasize the elements of statistical thinking:**
  - (a) the need for data;
  - (b) the importance of data production;
  - (c) the omnipresence of variability;
  - (d) the measuring and modeling of variability.
2. **Incorporate more data and concepts, fewer recipes and derivations. Wherever possible, automate computations and graphics.** An introductory course should:
  - (a) rely heavily on *real* (not merely realistic) data;
  - (b) emphasize *statistical* concepts, e.g., causation vs. association, experimental vs. observational and longitudinal vs. cross-sectional studies;
  - (c) rely on computers rather than computational recipes;
  - (d) treat formal derivations as secondary in importance.
3. **Foster active learning**, through the following alternatives to lecturing:
  - (a) group problem solving and discussion;
  - (b) laboratory exercises;
  - (c) demonstrations based on class-generated data;
  - (d) written and oral presentations;
  - (e) projects, either group or individual.

**Figure 1** Recommendations of the ASA/MAA Joint Curriculum Committee

required and in the statistical procedures presented, it aims to give students both an understanding of the main ideas of statistics and useful skills for working with data. Examples and exercises, though intended for beginners, use real data and give enough background to allow students to consider the meaning of their calculations. I often ask for conclusions that are more than a number (or “reject  $H_0$ ”). Some exercises require judgment in addition to right-or-wrong calculations and conclusions. I hope that teachers will encourage discussion of results in class.

Chapters 1 and 2 present the methods and unifying ideas of data analysis. Students appreciate the usefulness of data analysis; that they can actually do it relieves a bit of their anxiety about statistics. I hope that they will grow accustomed to examining data and will continue to do so even when formal inference to answer a specific question is the ultimate goal. Chapter 3 discusses random sampling and randomized comparative experiments. These are among the most important ideas in statistics and are often unjustly neglected in beginning instruction. Chapter 4 builds on the ideas of Chapter 3 and the data-analytic tools of Chapter 1 to present the central idea of a sampling distribution and (informally) the language and most important facts of probability theory. The optional Chapter 5 presents additional material for use in courses that require a greater emphasis on formal probability. Chapter 6, which describes the reasoning of statistical inference, is the cornerstone of the rest of the book. The remaining chapters present methods of inference for various settings, with a strong emphasis on practical aspects of using these methods. Chapters 7 and 8 discuss one-sample and two-sample procedures. Chapters 9, 10, and 11 (which can be read independently of each other in any order) offer a choice of somewhat more advanced topics. Working with data in real settings is a theme throughout the text.

### Technology

Automating calculations increases the students' ability to complete problems, reduces their frustration, and helps them concentrate on ideas and problem recognition rather than mechanics. *All students should have at least a “two-variable statistics” calculator* with functions for correlation and the least-squares regression line as well as for the mean and standard deviation. Graphing calculators offer considerably more capability. Because students have calculators, the text doesn't discuss “computing formulas” for the sample standard deviation or the least-squares regression line. *Statistical software* has considerable advantages over calculators: easier data entry and editing and much better graphics. I encourage the use of software whenever facilities permit. BPS does not, however, assume that students will use software. Instructors will find using software easier because the CD-ROM that accompanies each copy of the book contains the data sets for exercises and tables in several common formats, along with several other features (see supplement section).

Output from several technologies, from graphing calculators through spreadsheets to statistical packages such as Minitab, appears in the text. The variety is deliberate: students see that the methods they meet in BPS are universal and that basic knowledge makes output from any technology easy to read and use.

**Accessibility**

The intent of BPS is to be modern *and* accessible. In comparison with the longer *Introduction to the Practice of Statistics* (IPS),<sup>3</sup> BPS has more concise discussions, less optional material, and more stopping places for the reader. BPS requires less advanced reading and study skills than IPS, but retains the essential content and the emphasis on working with data and thinking about statistical problems. The short Apply Your Knowledge exercise sections that follow every major new idea allow a quick check of basic mastery. Each chapter ends with Statistics in Summary, an explicit list of goals and often a pictorial review of main ideas. Please see the “Guided Book Tour” which follows the Preface.

**Why did you do that?**

There is no single best way to organize our presentation of statistics to beginners. That said, my choices are based on more than a whim. Here are comments on several “frequently asked questions” about the order and selection of material in BPS.

**Why does the distinction between population and sample not appear in Chapters 1 and 2?** This is a sign that there is more to statistics than inference. In fact, statistical inference is appropriate only in rather special circumstances. Chapters 1 and 2 present tools and tactics for describing data—any data. These tools and tactics do not depend on the idea of inference from sample to population. Many data sets in these chapters (for example, the several sets of data about the 50 states) do not lend themselves to inference because they represent an entire population. John Tukey of Bell Labs and Princeton, the philosopher of modern data analysis, insists that the population/sample distinction be avoided when it is not relevant. He uses the word “batch” for data sets in general. I see no need for a special word, but I think Tukey is right.

**Why not begin with data production?** It is certainly reasonable to do so—the natural flow of a planned study is from design to data analysis to inference. I choose to place the design of data production (Chapter 3) after data analysis (Chapters 1 and 2) to emphasize that data-analytic techniques apply to any data. One of the primary purposes of statistical designs for producing data is to make inference possible, so the discussion in Chapter 3 is a natural transition to inference ideas.

**Why do normal distributions appear in Chapter 1?** Density curves such as the normal curves are just another tool to describe the distribution of a quantitative variable, along with stemplots, histograms, and boxplots. It is becoming common for software to offer to make density curves from data just as it offers histograms. I prefer not to suggest that this material is essentially tied to probability, as the traditional order does. I also want students to think about the flow from graphs to numerical summaries to mathematical models, and density curves are the common mathematical model for the overall pattern of a distribution. Finally, I would like to break up the indigestible lump of probability that troubles students so much. Meeting normal distributions early does this and strengthens the “probability distributions are like data distributions” way of approaching probability.

**Why not delay correlation and regression until late in the course, as is traditional?** BPS begins by offering experience working with data and gives a conceptual structure for this non-mathematical but very important part of statistics. Students profit from more experience with data and from seeing the conceptual structure worked out in relations among variables as well as in describing single-variable data. Moreover, correlation and regression as descriptive tools (Chapter 2) have wider scope than an emphasis on inference (Chapter 11) allows. The very important discussion of lurking variables, for example, fits poorly with inference. I consider Chapter 2 essential and Chapter 11 optional.

**Where did all the probability go?** Much of it went to the optional Chapter 5. Students don't mind skipping a chapter and are sometimes bothered by optional material within chapters, so in this new edition all of Chapter 4 is required and all of Chapter 5 is optional.

Experienced teachers recognize that students find probability difficult. Research on learning confirms our experience. Even students who can do formally posed probability problems often have a very fragile conceptual grasp of probability ideas. Attempting to present a substantial introduction to probability in a data-oriented statistics course for students who are not mathematically trained is difficult to do. Formal probability does not help these students master the ideas of inference (at least not as much as we teachers imagine), and it depletes reserves of mental energy that might better be applied to essentially statistical ideas.

I have therefore presented little formal probability in the core Chapter 4. The chapter follows a straight line from the idea of probability as long term regularity, through concrete ways of assigning probabilities, to the central idea of the sampling distribution of a statistic. The law of large numbers and the central limit theorem appear in the context of discussing the sampling distribution of a sample mean. What is omitted here is mostly "general probability rules," including conditional probability, and combinatorics. The general rules appear in Chapter 5. Combinatorics is a different (and even harder) topic.

**Why didn't you cover Topic X?** Introductory texts ought not to be encyclopedic. Including each reader's favorite topic (control charts, nonparametric tests, geometric distributions, and so on) results in a text that is formidable in size and intimidating to students. I chose topics on two grounds: they are the most commonly used in practice, and they are suitable vehicles for learning broader statistical ideas. There are studies of usage in many fields of application. For example, Emerson and Colditz<sup>4</sup> report that just descriptive statistics,  $t$  procedures, and two-way tables would give full access to 73% of the articles in the *New England Journal of Medicine*. That suggests a reasonable semester course from BPS: Chapters 1 to 4 and 6 to 9.

### Acknowledgments

I am grateful to many colleagues from two-year and four-year colleges and universities who commented on successive drafts of the manuscript:

Elizabeth Applebaum,  
*Park College*

Edgar Avelino,  
*Langara College*

Smiley Cheng,  
*University of Manitoba*

James Curl,  
*Modesto Junior College*

David Gurney,  
*Southeastern Louisiana University*

Donald Harden,  
*Georgia State University*

Sue Holt,  
*Cabrillo College*

Elizabeth Houseworth,  
*University of Oregon*

T. Henry Jablonski, Jr.,  
*Eastern Tennessee State University*

Tom Kaupe,  
*Shoreline Community College*

James Lang,  
*Valencia Community College*

Donald Loftgaarden,  
*University of Montana*

Steve Marsden,  
*Glendale College*

Darcy Mays,  
*Virginia Commonwealth University*

Amy Salvati,  
*Adirondack Community College*

N. Paul Schembari,  
*East Stroudsburg University*

W. Robert Stephenson,  
*Iowa State University*

Martin Tanner,  
*Northwestern University*

Bruce Torrence,  
*Randolph-Macon College*

Mike Turegon,  
*Oklahoma City Community College*

Jean Werner,  
*Mansfield University of Pennsylvania*

Dex Whittinghill,  
*Rowan University*

Rodney Wong,  
*University of California at Davis*

I would also like to thank reviewers of the first edition:

Douglas M. Andrews  
*Wittenberg University*

Rebecca Busam  
*The Ohio State University*

Michael Butler  
*College of the Redwoods*

Carolyn Pilers Dobler  
*Gustavus Adolphus College*

Joel B. Greenhouse  
*Carnegie Mellon University*

Larry Griffey  
*Florida Community College  
at Jacksonville*

Brenda Gunderson  
*The University  
of Michigan*

Catherine Cummins Hayes  
*University of Mobile*

Tim Hesterberg  
*Franklin & Marshall College*

Ronald La Porte  
*Macomb Community College*

Ken McDonald  
*Northwest Missouri State University*

William Notz  
*The Ohio State University*

Mary Parker  
*Austin Community College*

Kenneth Ross  
*University of Oregon*

Calvin Schmall  
*Solano Community College*

Frank Soler  
*De Anza College*

Linda Sorenson  
*Algoma University*

Tom Sutton  
*Mohawk College*



Sue Holt of Cabrillo College deserves special thanks for careful comments based on long use of BPS in a two-year college. I am particularly grateful to Patrick Farace, Diane Cimino Maass and the other editorial and design professionals who have contributed greatly to the attractiveness of this book.

Finally, I am indebted to many statistics teachers with whom I have discussed the teaching of our subject over many years; to the people from diverse fields with whom I have worked to understand data; and especially to students whose compliments and complaints have changed and improved my teaching. Working with teachers, colleagues in other disciplines, and students constantly reminds me of the importance of hands-on experience with data and of statistical thinking in an era when computer routines quickly handle statistical details.

David S. Moore

# SUPPLEMENTS

A full range of supplements is available to help teachers and students use *The Basic Practice of Statistics*. **Complimentary supplements** available for the instructor include:

- **Instructor CD-ROM** contains the Instructor's version of EESEE with solutions, Presentation Manager Pro software to develop presentations from pre-loaded text figures as well as the user's own content (saved locally or found on the World Wide Web) and masters to the text figures to create presentations with the user's own presentation software. ISBN: 0-7167-3612-8
- **Instructor's Guide with Solutions** by Darryl Nester of Bluffton College and David Moore. The Instructor's Guide includes teaching suggestions, chapter comments, and worked-out solutions to all exercises. Solutions include graphics and statistical software output from various statistical packages where appropriate. ISBN: 0-7167-3616-0
- **Test Bank** by William Notz and Michael Fligner of The Ohio State University. Hundreds of page-referenced multiple-choice questions allow instructors to generate quizzes and tests. Available in a printed version as well as electronic versions for Mac and Windows platforms. The electronic version publishes texts in print, or on a LAN system, or on-line. ISBNs: Printed: 0-7167-3618-7, Windows: 0-7167-3619-5, Mac: 0-7167-3620-9

## And for students. . . .

- **Student CD-ROM** developed by Sumanas Multimedia Development Services. Packaged with every copy of the text, the CD contains a number of components that illuminate and expand the themes of the text:
  - The Encyclopedia of Statistical Examples and Exercises (EESEE), developed by faculty of the Department of Statistics at Ohio State University, is a rich repository of case studies that apply the concepts of *The Basic Practice of Statistics*. Many of the diverse examples are illustrated by photos or video clips. Each case is accompanied by practice problems and a majority include data sets, which are available in Data Desk, Minitab, and JMP formats.
  - Most of the textbook's data sets are included in the various formats: ASCII, Excel, and Minitab.
  - Q&A self-quizzes for each chapter give students feedback for correct and incorrect answers. A printable report card records students' performances on the quizzes.
  - A CD/Web update section links to new and updated material as well as providing the latest statistics news via the World Wide Web.

- Additional text sections covering control charts and nonparametric tests are included on the CD.
- **Student Study Guide** by William Notz and Michael Fligner, Ohio State University, and Rebecca Busam, University of Texas, Austin. The study guide helps students learn and review the basic concepts of the textbook in a printed format. It explains crucial concepts in each section of the text and provides solutions to key text problems and step-through models of important statistical techniques. ISBN: 0-7167-3617-9
- **Minitab Manual** by Betsy Greenberg, University of Texas. Written specifically for students who are using Minitab with *The Basic Practice of Statistics*, this manual offers careful illustrations and exercises that allow the student to unlock the power of Minitab for statistical analysis. Appendices include lists of Minitab functions, commands, and macros for easy reference. Both windows and session commands are covered. ISBN: 0-7167-3613-X
- **New Excel Manual** by Fred Hoppe, McMaster University. Providing exercises and applications for each chapter, this manual demonstrates how Excel's ability to organize data into spreadsheets allows easy analysis and graphic exploration. Each chapter focuses on the manner in which Excel displays and analyses data, the basic themes of *The Basic Practice of Statistics*. ISBN: 0-7167-3611-X
- **New TI-83 Graphing Calculator Manual** by David Neal, Western Kentucky University. Offering detailed instructions on the use of the TI-83 graphing calculator with *The Basic Practice of Statistics*, this manual references text exercises to demonstrate solutions with TI-83 functions. Figures illustrate the calculator's output. A detailed list of TI-83 statistical functions is included. ISBN: 0-7167-3614-4
- **New SPSS Manual** by Paul Stephenson, Justine Ritchie, Neal Rogness, and Patricia Stephenson, Grand Valley State University. Written specifically for students who are using SPSS with *The Basic Practice of Statistics*, this manual demonstrates the software's ability to perform a wide variety of statistical techniques ranging from descriptive statistics to complex multivariate procedures. ISBN: 0-7167-3610-1
- **New SAS Manual** by Mike Evans, University of Toronto. This manual explores the uses of SAS analysis with *The Basic Practice of Statistics*. In-depth illustrations and exercises allow the student to unlock the power of SAS for use in statistical analysis. Appendices include lists of SAS functions, and commands for easy reference. ISBN: 0-7167-3609-8


# GUIDED BOOK TOUR

**New Redesigned Part and Chapter Openers** introduce the major concepts and themes of the three parts—analyzing and producing data in Part I, probability and inference in Part II, and more advanced inference in Part III. Chapter openers engage students by presenting capsule biographies of men and women who have been important in the development of statistics.

**New Four-Color Design** presents information in an appealing and stylish format making the text more accessible to students and easier to use. Statistical graphs can now be easily interpreted by using different colors, while four-color spot photographs enliven the pages with images from real applications of statistics.

**“A Bit More” Think Pieces** appear in the margin of each chapter and are accompanied by a light illustration. These brief notes offer interesting and often amusing anecdotes about statistics and its impact.

172 CHAPTER 3 • Producing Data



**Are these random digits really random?**

Not a chance. The random digits in Table B were produced by a computer program. Computer programs do exactly what you tell them to do. Give the program the same input and it will produce exactly the same “random” digits. Of course, clever people have devised computer programs that produce output that looks like random digits. These are called “pseudo-random numbers,” and that’s what Table B contains. Pseudo-random numbers work fine for statistical randomizing, but they have hidden nonrandom patterns that can mess up more refined uses.

• Each triple of entries is equally likely to be any of the 1000 possibilities 000, 001, . . . , 999; and so on. These “equally likely” facts make it easy to use Table B to choose an SRS. Here is an example that shows how.

**EXAMPLE 3.4 How to choose an SRS**

Joan’s small accounting firm serves 30 business clients. Joan wants to interview a sample of 5 clients in detail to find ways to improve client satisfaction. To avoid bias, she chooses an SRS of size 5.

**Step 1: Label.** Give each client a numerical label, using as few digits as possible. Two digits are needed to label 30 clients, so we use labels

01, 02, 03, . . . , 29, 30

It is also correct to use labels 00 to 29 or even another choice of 30 two-digit labels. Here is the list of clients, with labels attached:

01	A-1 Plumbing	16	JL Records
02	Accent Printing	17	Johnson Commodities
03	Action Sport Shop	18	Keiser Construction
04	Anderson Construction	19	Liu’s Chinese Restaurant
05	Bailey Trucking	20	Magic-Tan
06	Balloons Inc.	21	Peerless Machine
07	Bennett Hardware	22	Photo Arts
08	Bes’s Camera Shop	23	River City Books
09	Blue Print Specialties	24	Riverside Tavern
10	Central Tree Service	25	Rustic Boutique
11	Classic Flowers	26	Satellite Services
12	Computer Answers	27	Scotch Wash
13	Darlene’s Dolls	28	Sewer’s Center
14	Fleisch Realty	29	Tire Specialties
15	Hernandez Electronics	30	Von’s Video Store

**Step 2: Table.** Enter Table B anywhere and read two-digit groups. Suppose we enter at line 130, which is

69051 64817 87174 09517 84534 06489 87201 97245

The first 10 two-digit groups in this line are

69 05 16 48 17 87 17 40 95 17

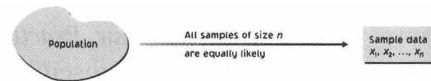
Each successive two-digit group is a label. The labels 00 and 31 to 99 are not used in this example, so we ignore them. The first 5 labels between 01 and 30 that we encounter in the table choose our sample. Of the first 10 labels in line 130, we ignore 5 because they are too high (over 30). The others are 05, 16, 17, 17, and 17. The clients labeled 05, 16, and 17 go into the sample. Ignore the second and third 17s

◀ **New and Revised Exercises and Examples** include a wide variety of real data in real settings from business and economics, demographics, public health and medicine, agriculture, biological sciences, history, psychology, and more.

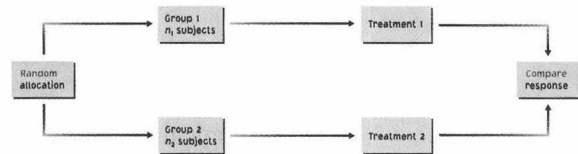
◀ **Step-by-Step Examples** work through procedures and join techniques and concepts together in an easy-to-follow format.

**New “Statistics in Summary”** sections at the end of each chapter review key concepts, often in easy-to-remember pictorial form. A detailed outline of chapter learning objectives follows.

**STATISTICS IN SUMMARY**  
Simple Random Sample



**STATISTICS IN SUMMARY**  
Randomized Comparative Experiment



case should show the sizes of the groups, the specific treatments, and the response variable.

- Use Table B of random digits to carry out the random assignment of subjects to groups in a completely randomized experiment.
- Recognize the placebo effect. Recognize when the double-blind technique should be used.
- Explain why a randomized comparative experiment can give good evidence for cause-and-effect relationships.

184 CHAPTER 3 • Producing Data

We hope to see a difference in the responses so large that it is unlikely to open just because of chance variation. We can use the laws of probability, we give a mathematical description of chance behavior, to learn if the treatment effects are larger than we would expect to see if only chance were operating. If they are, we call them **statistically significant**.

**STATISTICAL SIGNIFICANCE**

An observed effect so large that it would rarely occur by chance is called **statistically significant**.

If we observe statistically significant differences among the groups in a comparative randomized experiment, we have good evidence that the treatment actually caused these differences. You will often see the phrase “statistically significant” in reports of investigations in many fields of study. The great advantage of randomized comparative experiments is that they can produce data that give good evidence for a cause-and-effect relationship between the explanatory and response variables. We know that in general a strong association does not imply causation. A statistically significant association in data from a well-designed experiment does imply causation.



**Scratch my furry ears**

Rats and rabbits, specially bred to be uniform in their inherited characteristics, are the subjects in many experiments. It turns out that animals, like people, are quite sensitive to how they are treated. This creates some amusing opportunities for hidden bias. For example, human affection can change the cholesterol level of rabbits. Choose some rabbits at random and regularly remove them from their cages to have their heads scratched by friendly people. Leave other rabbits unloved. All the rabbits eat the same diet, but the rabbits that receive affection have lower cholesterol.

**APPLY YOUR KNOWLEDGE**

- Conserving energy. Example 3.13 describes an experiment to learn whether providing households with electronic indicators or charts will reduce their electricity consumption. An executive of the electric company objects to including a control group. He says, “It would be simpler to just compare electricity use last year (before the indicator or chart was provided) with consumption in the same period this year. If households use less electricity this year, the indicator or chart must be working.” Explain clearly why this design is inferior to that in Example 3.13.
- Exercise and heart attacks. Does regular exercise reduce the risk of a heart attack? Here are two ways to study this question. Explain clearly why the second design will produce more trustworthy data.
  - A researcher finds 2000 men over 40 who exercise regularly and have not had heart attacks. She matches each with a similar man who does not exercise regularly, and she follows both groups for 5 years.
  - Another researcher finds 4000 men over 40 who have not had heart attacks and are willing to participate in a study. She assigns 2000 of the men to a regular program of supervised exercise. The

**CHAPTER 3 Review Exercises**

- Repairing knees in comfort (EESEE). Injured knees are routinely repaired by arthroscopic surgery that does not require opening up the knee. Can we reduce patient discomfort by giving them a nonsteroidal anti-inflammatory drug (NSAID)? Eighty-three patients were placed in three groups. Group A received the NSAID both before and after surgery. Group B was given a placebo before and the NSAID after. Group C received a placebo both before and after surgery. The patients recorded a pain score by answering questions one day after the surgery.<sup>17</sup>



**Titles for Examples and Exercises** pique student interest by displaying the ties between statistics and a very wide variety of issues and topics.

**CD-ROM Icon** references exercises in BPS which are based on data from an EESEE case study on the CD-ROM.

**“Apply Your Knowledge” Exercises** immediately follow each new idea. These straightforward exercises embedded in the text allow rapid feedback and reinforcement of basic concepts and skills.

# APPLICATIONS

*The Basic Practice of Statistics* presents a wide variety of applications from diverse disciplines. The list below indicates the number of exercises and examples which relate to different fields.

## **Exercises by application:**

Agriculture: 19  
Biological and environmental sciences: 69  
Business and economics: 145  
Education: 63  
History and public policy: 13  
Paleontology: 2  
Physical science: 31  
People and places: 120  
Psychology and behavioral sciences: 24  
Public health and medicine: 111  
Sports: 17

## **Examples by application:**

Agriculture: 14  
Biological and environmental sciences: 18  
Business and economics: 32  
Education: 13  
History and public policy: 5  
Physical science: 15  
People and places: 20  
Psychology and behavioral sciences: 8  
Public health and medicine: 29  
Sports: 8

For a complete breakdown list of examples and exercises by chapter and number, please see the *Instructor's Guide* or W.H. Freeman's Web site: [www.whfreeman.com/statistics](http://www.whfreeman.com/statistics)

# STATISTICAL THINKING

## How to Tell the Facts from the Artifacts

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 can’t weigh 10.5 ounces or 10.5 kilograms. The context makes the number informative.

Statistics uses data to gain insight and to draw conclusions. Our 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 an informal look at some principles of statistical thinking.

### Data Illuminate

What percent of the American population do you think is black? What percent do you think is white? When white Americans were asked these questions, their average answers were 23.8% black and 49.9% white. In fact, the Census Bureau tells us that 11.8% of Americans are black and 74% are white.<sup>1</sup>

Race remains a central social issue in the United States. It is illuminating to see that whites think (wrongly) that they are a minority. The census data—what really is true about the U.S. population—are also illuminating. We wonder if knowing the facts might help change attitudes.

### Data Beat Anecdotes

An anecdote is a striking story that sticks in our minds exactly because it is striking. Anecdotes humanize an issue, but they can be misleading.

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.<sup>2</sup>

Now compare the effectiveness of a television news report of a 5-year, \$5 million investigation against 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. A statistically literate person, however, knows that data are more reliable than anecdotes because they systematically describe an overall picture rather than focusing on a few incidents.

### Beware the Lurking Variable

Air travelers would like their flights to arrive on time. Airlines collect data about on-time arrivals and report them to the Department of Transportation. Here are one month’s data for flights from several western cities for two airlines:

	On time	Delayed
Alaska Airlines	3274	501
America West	6438	787

You can see that the percentages of late flights were:

$$\text{Alaska Airlines } \frac{501}{3775} = 13.3\%$$

$$\text{America West } \frac{787}{7225} = 10.9\%$$

It appears that America West does better.

This isn’t the whole story, however. Almost all relationships between two variables are influenced by other variables lurking in the background. We have data on two variables, the airline and whether or not the flight was late. Let’s add data on a third variable, which city the flight left from.<sup>3</sup>

	Alaska Airlines		America West	
	On time	Delayed	On time	Delayed
Los Angeles	497	62	694	117
Phoenix	221	12	4840	415
San Diego	212	20	383	65
San Francisco	503	102	320	129
Seattle	1841	305	201	61
Total	3274	501	6438	787



The “Total” row shows that the new table describes the same flights as the earlier table. Look again at the percentages of late flights, first for Los Angeles:

$$\text{Alaska Airlines } \frac{62}{559} = 11.1\%$$

$$\text{America West } \frac{117}{811} = 14.4\%$$

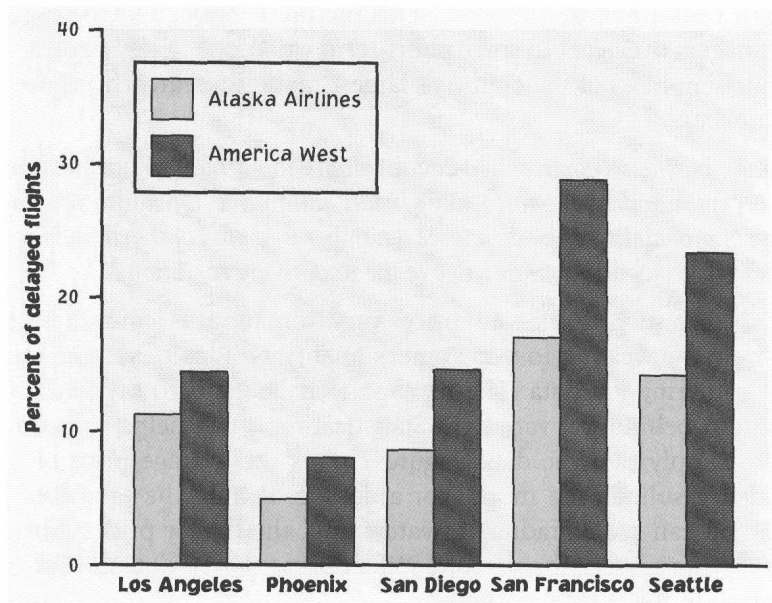
Alaska Airlines wins. The percentages delayed for Phoenix are:

$$\text{Alaska Airlines } \frac{12}{233} = 5.2\%$$

$$\text{America West } \frac{415}{5255} = 7.9\%$$

Alaska Airlines wins again. In fact, as Figure 1 shows, Alaska Airlines has a lower percentage of late flights at *every one* of these cities.

How can it happen that Alaska Airlines wins at every city but America West wins when we combine all the cities? Look at the data: America West flies most often from sunny Phoenix, where there are few delays. Alaska Airlines flies most often from Seattle, where fog and rain cause frequent delays. What city we fly from has a major influence on the chance of a delay, so including the city data reverses our conclusion. The message is worth repeating: almost all relationships between two variables are influenced by other variables lurking in the background.



**Figure 1** Comparing the percents of delayed flights for two airlines at five airports.