# A FIRST COURSE IN Statistics

SIXTH EDITION

JAMES T. McCLAVE / TERRY SINCICH

# A First Course in Statistics

**Sixth Edition**

James T. McClave
*Info Tech, Inc.*
*University of Florida*

Terry Sincich
*University of South Florida*

**PRENTICE HALL**
Upper Saddle River, New Jersey 07458

**Library of Congress Cataloging-in-Publication Data**
McClave, James T.
A first course in statistics / James T. McClave, Terry Sincich. — 6th ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-13-579277-0
1. Statistics. I. Sincich, Terry. II. Title.
QA276.M378 1997
519.5—dc20 96-34853
CIP

Executive Editor: Ann Heath
Editorial Director: Tim Bozik
Editor-in-Chief: Jerome Grant
Assistant Vice President of Production and Manufacturing: David W. Riccardi
Development Editor: Millicent Treloar
Text Design/Project Management/Composition: Elm Street Publishing Services, Inc.
Managing Editor: Linda Mihatov Behrens
Executive Managing Editor: Kathleen Schiaparelli
Marketing Manager: Evan Girard
Creative Director: Paula Maylahn
Art Director: Amy Rosen
Assistant to the Art Director: Rod Hernandez
Cover Designer: Jeanette Jacobs
Manufacturing Buyer: Alan Fischer
Manufacturing Manager: Trudy Pisciotti
Editorial Assistant: Mindy Ince
Cover Photo: Marcus Amon/Tony Stone Images
Interior Photos:
Pages ix, 1, 21: Paul Rees/Tony Stone Images
Pages 91, 145, 209, 245: Frank Stewart
Pages 287, 361: Mark E. Gibson/Visuals Unlimited
Page 399: Mitch Wojnarowicz/The Image Works

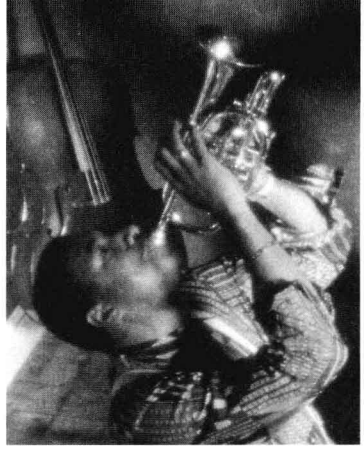

Simon & Schuster / A Viacom Company
Upper Saddle River, NJ 07458



Printed in the United States of America
10 9 8 7 6 5 4 3 2

ISBN 0-13-579277-0

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall Canada, Inc., *Toronto*
Prentice-Hall Hispanoamericano, S.A., *Mexico*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Simon & Schuster Asia Pte. Ltd., *Singapore*



# PREFACE

## STATISTICS IN THE 1990's

Most news reports today include results from scientific studies, usually statistical in nature. Consider this one published on page one of the February 12, 1996, *New York Times:* "As Patrons Age, Future of the Arts Is Uncertain." Many of the studies we see reported raise issues of public funding, consumer awareness, or public health and safety, all of which directly affect our daily lives and the decisions we make. In today's world, a solid understanding of the information in statistical reports is as important to artists, actors, and musicians as it is to the sociologists, economists, scientists, and others who produce the reports.

This sixth edition of *A First Course in Statistics* has been extensively revised to stress the development of statistical thinking, the assessment of credibility and value of the inferences made from data, both by those who consume them and those who produce them. This is a one-semester introductory text emphasizing inference, with extensive coverage of data collection and analysis as needed to evaluate the reported results of statistical studies. It covers basic statistical topics through simple linear regression. It assumes a mathematical background of basic algebra.

## NEW IN THE SIXTH EDITION

### Major Content Changes

- Chapter 1 has been entirely rewritten to set the groundwork for statistical thinking and to acquaint the student at an early point with the importance of data collection, measurement, types of data, experiments, surveys, and the validity of drawing inferences from data.
- Chapter 2 now more thoroughly covers descriptive analytical tools that are useful in examining data. Pie charts, bar graphs, frequency tables (for qualitative data), and dot plots (for quantitative data) have been added.
- Chapters 5 through 8 more heavily stress confidence intervals and rely more heavily on computer output than on formulas.
- Chapter 6 now provides an equal balance between applications relying on $p$-values and those relying on critical values in their interpretation.
- Chapter 7 includes an optional section on analysis of variance (ANOVA).
- Chapter 8 includes a section on contingency table analysis.
- Chapter 9 now incorporates computer printouts throughout the discussion of simple linear regression, and it includes a new optional section on rank correlation.

### Pedagogy

- All new, restructured cases, approximately two per chapter, summarize statistical studies on contemporary, controversial issues and include questions to the students prompting them to evaluate findings.
- More than 60 percent of the examples and exercises in the text are new or have been completely revised. Most employ the use of current (post-1990) real data taken from a wide variety of publications.
- New end-of-chapter "Quick Reviews" provide page references to important ideas in the chapter.
- New "Language Lab" feature explains the use of key symbols in formulas and provides a pronunciation guide.
- New "Student Projects" in each chapter emphasize gathering data, analyzing data, and/or report writing.
- Five entirely new, large data bases provide the foundation for the new "Exploring Data with a Computer" feature found in the end-of-chapter exercises, and the data bases are used in examples throughout the text.

## TRADITIONAL STRENGTHS

We have maintained the features of *A First Course in Statistics* that we believe make it unique among introductory statistics texts. These features, which assist the student in achieving an overview of statistics and an understanding of its relevance in the social and life sciences, business, and everyday life, are as follows:

### The Use of Examples as a Teaching Device

Almost all new ideas are introduced and illustrated by real data-based applications and examples. We believe that students better understand definitions, generalizations, and abstractions *after* seeing an application.

### Many Exercises—Labeled by Type

The text includes more than 800 exercises illustrated by applications in almost all areas of research. Many students have trouble learning the mechanics of statistical techniques when all problems are couched in terms of realistic applications. For this reason, all exercise sections are divided into two parts:

**Learning the Mechanics.** Designed as straightforward applications of new concepts, these exercises allow students to test their ability to comprehend a concept or a definition.

**Applying the Concepts.** Based on applications taken from a wide variety of journals, newspapers, and other sources, these exercises develop the student's skills at comprehending real-world problems that describe situations to which the techniques may be applied.

### Nonparametric Methods Integrated

Throughout the text, optional sections on alternative nonparametric procedures follow the relevant sections.

## FULL VERSION AVAILABLE

A full-sized one- or two-term general statistics text is also available: *Statistics,* seventh edition by James T. McClave (University of Florida), Frank Dietrich (Northern Kentucky University), and Terry Sincich (University of South Florida): © 1997, 823 pp. cloth, ISBN: 0-13-471542-X. Call Prentice Hall Faculty Services at 1–800–526–0485 or your local representative.

**Contents.** 1. Statistics, Data, and Statistical Thinking 2. Methods for Describing Sets of Data 3. Probability 4. Discrete Random Variables 5. Continuous Random Variables 6. Sampling Distributions 7. Inferences Based on a Single Sample: Estimation with Confidence Intervals 8. Inferences Based on a Single Sample: Tests of Hypotheses 9. Inferences Based on Two Samples: Confidence Intervals and Tests of Hypotheses 10. Analysis of Variance: Comparing More Than Two Means 11. Simple Linear Regression 12. Multiple Regression 13. Model Building 14. The Chi-Square Test and the Analysis of Contingency Tables 15. Nonparametric Statistics Appendix A. Tables Appendix B. Data Sets Appendix C. Calculation Formulas for Analysis of Variance Answers to Selected Odd-Numbered Exercises Index

## ACKNOWLEDGMENTS

This book reflects the efforts of a great many people over a number of years. First we would like to thank the following professors, whose reviews and comments on this and prior editions of *Statistics* and/or *A First Course in Statistics* have contributed to this sixth edition:

### Reviewers of Previous Editions

David Atkinson
Olivet Nazarene University

William H. Beyer
University of Akron

Patricia M. Buchanan
Pennsylvania State University

Kathryn Chaloner
University of Minnesota

John Dirkse
California State University—Bakersfield

N. B. Ebrahimi
Northern Illinois University

Dale Everson
University of Idaho

Rudy Gideon
University of Montana

Larry Griffey
Florida Community College

David Groggel
Miami University at Oxford

John E. Groves
California Polytechnic State University—San Luis Obispo

Jean L. Holton
Virginia Commonwealth University

John H. Kellermeier
State University College at Plattsburgh

Timothy J. Killeen
University of Connecticut

William G. Koellner
Montclair State University

James R. Lackritz
San Diego State University

Diane Lambert
AT&T/Bell Laboratories

James Lang
Valencia Junior College

Glenn Larson
University of Regina

John J. Lefante, Jr.
University of South Alabama

Pi-Erh Lin
Florida State University

R. Bruce Lind
University of Puget Sound

Rhonda Magel
North Dakota State University

Linda C. Malone
University of Central Florida

Allen E. Martin
California State University—
Los Angeles

Leslie Matekaitis
Cal Genetics

E. Donice McCune
Stephen F. Austin State University

Satya Narayan Mishra
University of South Alabama

A. Mukherjea
Unversity of South Florida

Bernard Ostle
University of Central Florida

William B. Owen
Central Washington University

Won J. Park
Wright State University

John J. Peterson
Smith Kline & French Laboratories

Andrew Rosalsky
University of Florida

C. Bradley Russell
Clemson University

Rita Schillaber
University of Alberta

James R. Schott
University of Central Florida

Susan C. Schott
University of Central Florida

George Schultz
St. Petersburg Junior College

Carl James Schwarz
University of Manitoba

Mike Seyfried
Shippensburg University

Lewis Shoemaker
Millersville University

Charles W. Sinclair
Portland State University

Robert K. Smidt
California Polytechnic State
University—San Luis Obispo

Vasanth B. Solomon
Drake University

W. Robert Stephenson
Iowa State University

Barbara Treadwell
Western Michigan University

Dan Voss
Wright State University

Augustin Vukov
University of Toronto

Dennis D. Wackerly
University of Florida

### Reviewers Involved with the Seventh Edition of *Statistics*

Bill Adamson
South Dakota State University

Ibrahim Ahmad
Northern Illinois University

Mary Sue Beersman
Northeast Missouri State University

Hanfeng Chen
Bowling Green State University

Gerardo Chin-Leo
The Evergreen State College

Linda Brant Collins
Iowa State University

John Egenolf
University of Alaska—Anchorage

Christine Franklin
University of Georgia

Victoria Marie Gribshaw
Seton Hill College

Shu-ping Hodgson
Central Michigan University

Soon Hong
Grand Valley State University

Ina Parks S. Howell
Florida International University

Timothy Killeen
University of Connecticut

Edwin G. Landauer
Clackamas Community College

Brenda Masters
Oklahoma State University

Mark M. Meerschaert
University of Nevada—Reno

Christopher Morrell
Loyola College in Maryland

Thomas O'Gorman
Northern Illinois University

Ronald Pierce
Eastern Kentucky University

Betty Rehfuss
North Dakota State University—Bottineau

Andrew Rosalsky
University of Florida

James R. Schott
University of Central Florida

Arvind K. Shah
University of South Alabama

Thaddeus Tarpey
Wright State University

Kathy Taylor
Clackamas Community College

Special thanks are due to our ancillary authors, Nancy Shafer Boudreau and Mark Dummeldinger, and to typist Brenda Dobson, who have worked with us for many years. Carl Richard Gumina has done an excellent job of accuracy checking this sixth edition and has helped us to ensure a highly accurate, clean text. The Prentice Hall staff of Ann Heath, Millicent Treloar, Mindy Ince, Evan Girard, Bill Paquin, Linda Behrens, and Alan Fischer and Elm Street Publishing Services' Martha Beyerlein, Barb Lange, Sue Langguth, and Cathy Ferguson helped greatly with all phases of the text development, production, and marketing effort. Emily Thompson did an outstanding job as copy editor. Finally, we owe special thanks to Faith Sincich, whose efforts in preparing the manuscript for production and proofreading all stages of the book deserve special recognition.

## SUPPLEMENTS FOR THE INSTRUCTOR

The supplements for the sixth edition have been completely revised to reflect the extensive revisions of the text. Each element in the package has been accuracy checked to ensure adherence to the approaches presented in the main text, clarity, and freedom from computational, typographical, and statistical errors.

### *Instructor's Solutions Manual* by Nancy S. Boudreau (ISBN 0-13-579285-1)

Solutions to all of the even-numbered exercises are given in this manual. Careful attention has been paid to ensure that all methods of solution and notation are consistent with those used in the core text. Solutions to the odd-numbered exercises are found in the *Student's Solutions Manual.*

### *Test Bank* by Mark Dummeldinger (ISBN 0-13-579273-2)

Entirely rewritten, the *Test Bank* now includes more than 1,000 problems that correlate to problems presented in the text.

### Windows PH Custom Test (ISBN 0-13-595547-5)
### Mac PH Custom Test (0-13-472044-X): prepared for *Statistics, 7e;* may also be used with this text

Incorporates three levels of test creation: (1) selection of questions from a test bank; (2) addition of new questions with the ability to import test and graphics files from WordPerfect, Microsoft Word, and Wordstar; and (3) algorithmic generation of multiple questions from a single question template. PH Custom Test has a full-featured graphics editor supporting the complex formulas and graphics required by the statistics discipline.

### Data Disk (ISBN 0-13-579301-7): prepared for *A First Course in Statistics,* 6e

The data sets described in "Appendix B" and the data for all exercises containing twenty or more observations are available on a 3 ½-inch diskette in ASCII format. A list of the exercise data on the disk, with file names, is provided on pages vii–viii.

### NEW! *Annotated Instructor's Edition* (AIE) (ISBN 0-13-471657-4): prepared for *Statistics, 7e;* may also be used with this text

Marginal notes placed next to discussions of essential teaching concepts include:

- Teaching Tips—suggest alternative presentations or point out common student errors
- Exercises—reference specific section and chapter exercises that reinforce the concept
-  —identify accompanying PowerPoint slides
- Short Answers—section and chapter exercise answers are provided next to the selected exercises

### NEW! *Instructor's Notes* by Mark Dummeldinger (ISBN 0-13-494931-5): prepared for *Statistics, 7e;* may also be used with this text

This new printed resource contains suggestions for using the questions at the end of the cases as the basis for class discussion, a complete short answer book with letter of permission to duplicate for student use, and many of the exercises and solutions that were removed from the sixth edition of *Statistics.*

### NEW! PowerPoint Presentation Demo Disk (ISBN 0-13-472151-9): prepared for *Statistics, 7e;* may also be used with this text

This versatile Windows-based tool may be used by professors in a number of different ways.

- Slide show in an electronic classroom
- Printed and used as transparency masters

- Printed copies may be distributed to students as a convenient note-taking device.

Included are: learning objectives, thinking challenges, concept presentation slides, and examples with worked out solutions. The full product may be downloaded from a Prentice Hall FTP site. Details are available from the publisher.

### NEW! *New York Times Supplement* (ISBN 0-13-261637-8)

Copies of this supplement may be requested from Prentice Hall by instructors for distribution in their classes. This supplement contains high interest articles published recently in *The New York Times* that relate to topics covered in the text.

### NEW! *Computer Software Tutorials: SAS, SPSS, Minitab, ASP* by Terry Sincich (ISBN 0-13-531609-X)

This self-contained manual provides a brief introduction to each of these statistical packages. Keystroke commands and an extensive use of output instructs the student in the use of the chosen statistical software package.

## SUPPLEMENTS AVAILABLE FOR PURCHASE BY STUDENTS

### *Student's Solutions Manual* by Nancy S. Boudreau (ISBN 0-13-595539-4)

Fully worked-out solutions to all of the odd-numbered exercises are provided in this manual. Careful attention has been paid to ensure that all methods of solution and notation are consistent with those used in the core text.

### Student Versions of SPSS

Student versions of SPSS, the award-winning and market-leading commercial data analysis package, are available for student purchase. They are designed specifically for hands-on classroom teaching and learning of data analysis, statistics, and research methods. Windows, Windows 95, and Power Mac versions of the software allow the user to take full advantage of the easy-to-use graphical user interface combined with the traditional power of SPSS. Details on all current products are available from the publisher.

### ConStatS by Tufts University (ISBN 0-13-502600-8)

ConStatS is a set of Microsoft Windows-based programs designed to help college students understand concepts taught in a first semester course on probability and statistics. Under development at Tufts University for over eight years, ConStatS helps improve students' conceptual understanding of statistics by engaging them in an active, experimental style of learning. ConStatS is available for individual purchase or to schools on a site license basis. A companion ConStatS workbook (ISBN 0-13-522848-4) that guides students through the labs and ensures they gain the maximum benefit is also available.

**For additional information about texts and other materials available from Prentice Hall, visit us on-line at http//.www.prenhall.com.**

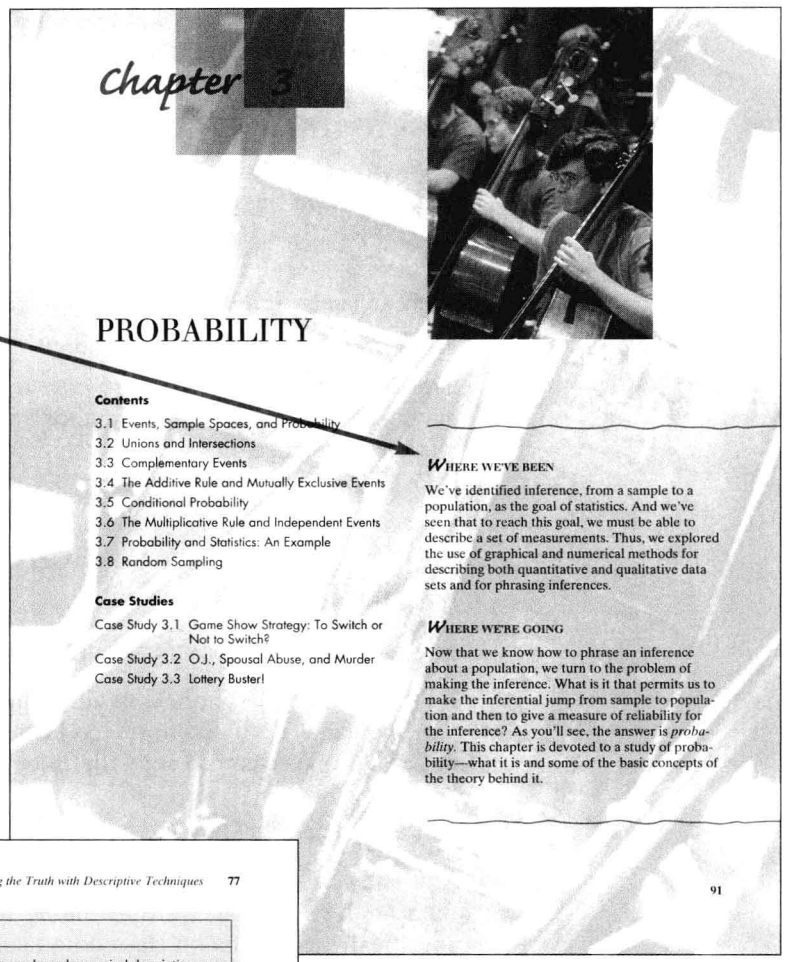
## How to Use This Book

### To the Student

The following four pages will demonstrate how to use this text in the most effective way to make studying easier and to understand the connection between statistics and your world.

### Chapter Openers Provide a Road Map

- **Where We've Been** quickly reviews how information learned previously applies to the chapter at hand.
- **Where We're Going** highlights how the chapter topics fit into your growing understanding of statistical inference.


Chapter 3

PROBABILITY

**Contents**



**Case Studies**



*WHERE WE'VE BEEN*

We've identified inference, from a sample to a population, as the goal of statistics. And we've seen that to reach this goal, we must be able to describe a set of measurements. Thus, we explored the use of graphical and numerical methods for describing both quantitative and qualitative data sets and for phrasing inferences.

*WHERE WE'RE GOING*

Now that we know how to phrase an inference about a population, we turn to the problem of making the inference. What is it that permits us to make the inferential jump from sample to population and then to give a measure of reliability for the inference? As you'll see, the answer is *probability*. This chapter is devoted to a study of probability—what it is and some of the basic concepts of the theory behind it.

91



**Suicide in Urban Jails**

CASE STUDY 2.3

Suicide is the leading cause of death of Americans incarcerated in correctional facilities. Moreover, the rate of completed suicide among jailed inmates who have made previous attempts is more than 100 times the rate in the general population. What factors increase the risk of suicide in urban jails?

To answer this question, a group of researchers (with backgrounds in political science, psychology, psychiatry, and correctional facilitation) collected data on all suicides that occurred from 1967 to 1992 in the Wayne County Jail, Detroit, Michigan (*American Journal of Psychiatry*, July 1995). A total of 37 suicides occurred during this period, all by hanging. For each suicide victim, the following variables were measured:

Number of days in jail before suicide
Marital status (married, single, widowed, or divorced)
Race (white or nonwhite)
Charge (murder/manslaughter or other)
Shift on which suicide occurred: day (7 AM–3 PM), afternoon (3 PM–11 PM), or night (11 PM–7 AM)

The complete data set is provided in Table 2.14.

Use graphs and numerical descriptive measures to summarize and describe the data in Table 2.14. Next, use these descriptive techniques to answer the following questions raised by the researchers. Then use your answers to the questions to make inferences about the general population of suicidal inmates. (We'll show you how to attach measures of reliability to these inferences in Chapters 7 and 8.)

**Focus**

a. Are suicides at the jail more likely to be committed by inmates charged with murder/manslaughter or with lesser crimes?
b. Are suicides at the jail more likely to be committed at night?
c. What is the typical length of time an inmate is in jail before committing suicide?
d. Are more suicides committed by white or nonwhite inmates?
e. Have suicides at the jail declined over the years?
f. Which are more likely to commit suicide earlier in their length of stay at the jail, white or nonwhite inmates? Inmates charged with murder/manslaughter or other inmates? Married or nonmarried inmates?

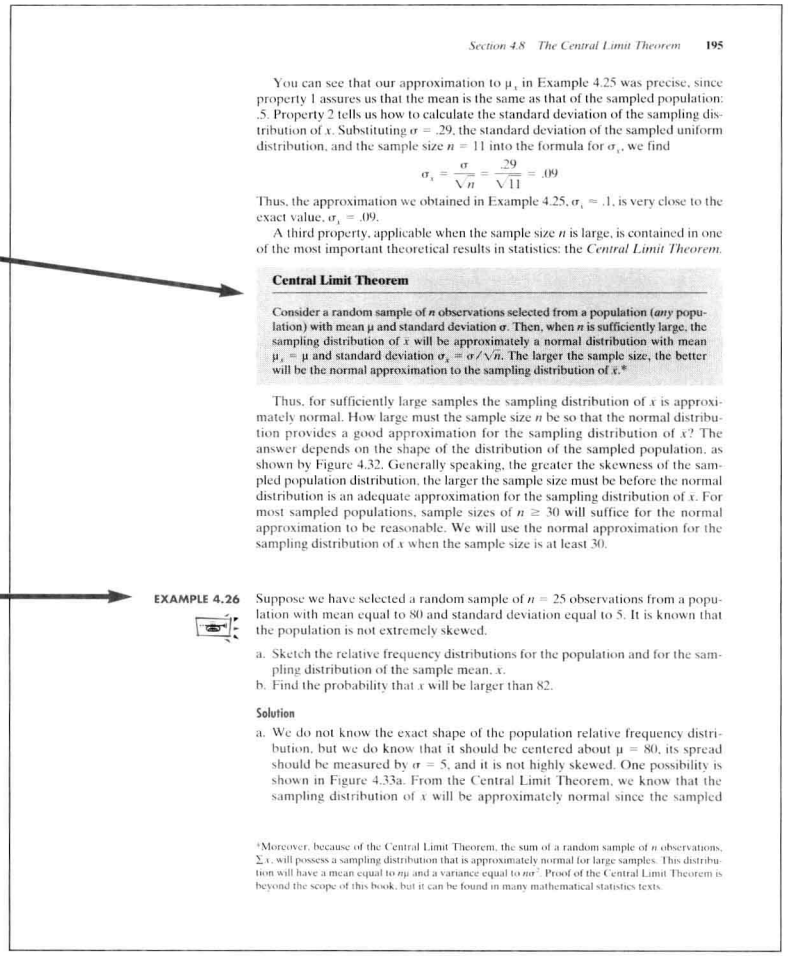
## 2.9 DISTORTING THE TRUTH WITH DESCRIPTIVE TECHNIQUES

A picture may be "worth a thousand words," but pictures can also color messages or distort them. In fact, the pictures in statistics—relative frequency histograms, charts, and other graphical descriptions—are susceptible to distortion, so we have to examine each of them with care.

We will mention a few of the pitfalls to watch for when interpreting a chart or graph. But first we should mention the **time series graph**, which is often the object of distortion. This type of [illegible] records the behavior of some variable of interest

### Case Studies Explore High Interest Issues

- One to three cases per chapter showcase controversial, contemporary issues.
- Work through the **Focus** questions to help you evaluate the findings.



*Section 4.8 The Central Limit Theorem* 195

You can see that our approximation to $\mu_{\bar{x}}$ in Example 4.25 was precise, since property 1 assures us that the mean is the same as that of the sampled population: .5. Property 2 tells us how to calculate the standard deviation of the sampling distribution of $\bar{x}$. Substituting $\sigma = .29$, the standard deviation of the sampled uniform distribution, and the sample size $n = 11$ into the formula for $\sigma_{\bar{x}}$, we find

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{.29}{\sqrt{11}} = .09$$

Thus, the approximation we obtained in Example 4.25, $\sigma_{\bar{x}} \approx .1$, is very close to the exact value, $\sigma_{\bar{x}} = .09$.

A third property, applicable when the sample size $n$ is large, is contained in one of the most important theoretical results in statistics: the *Central Limit Theorem*.

**Central Limit Theorem**

Consider a random sample of $n$ observations selected from a population (*any* population) with mean $\mu$ and standard deviation $\sigma$. Then, when $n$ is sufficiently large, the sampling distribution of $\bar{x}$ will be approximately a normal distribution with mean $\mu_{\bar{x}} = \mu$ and standard deviation $\sigma_{\bar{x}} = \sigma/\sqrt{n}$. The larger the sample size, the better will be the normal approximation to the sampling distribution of $\bar{x}$.*

Thus, for sufficiently large samples the sampling distribution of $\bar{x}$ is approximately normal. How large must the sample size $n$ be so that the normal distribution provides a good approximation for the sampling distribution of $\bar{x}$? The answer depends on the shape of the distribution of the sampled population, as shown by Figure 4.32. Generally speaking, the greater the skewness of the sampled population distribution, the larger the sample size must be before the normal distribution is an adequate approximation for the sampling distribution of $\bar{x}$. For most sampled populations, sample sizes of $n \geq 30$ will suffice for the normal approximation to be reasonable. We will use the normal approximation for the sampling distribution of $\bar{x}$ when the sample size is at least 30.

**EXAMPLE 4.26** Suppose we have selected a random sample of $n = 25$ observations from a population with mean equal to 80 and standard deviation equal to 5. It is known that the population is not extremely skewed.

a. Sketch the relative frequency distributions for the population and for the sampling distribution of the sample mean, $\bar{x}$.
b. Find the probability that $\bar{x}$ will be larger than 82.

**Solution**

a. We do not know the exact shape of the population relative frequency distribution, but we do know that it should be centered about $\mu = 80$, its spread should be measured by $\sigma = 5$, and it is not highly skewed. One possibility is shown in Figure 4.33a. From the Central Limit Theorem, we know that the sampling distribution of $\bar{x}$ will be approximately normal since the sampled

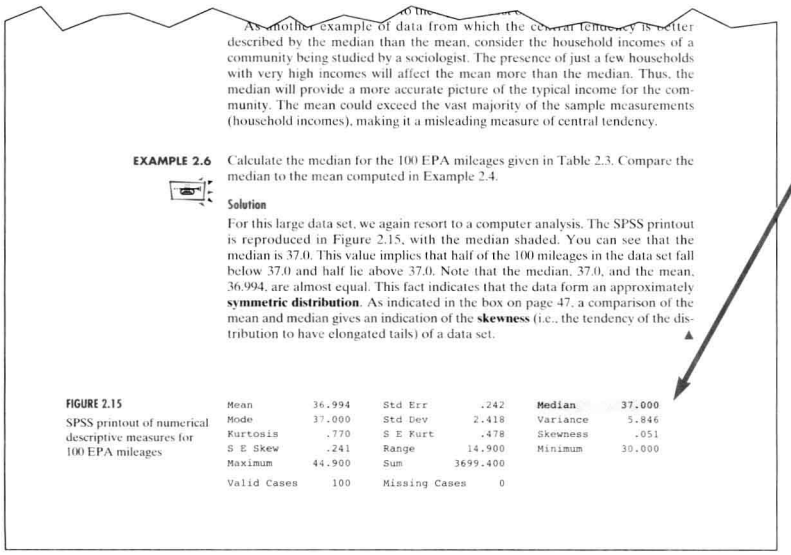
*Moreover, because of the Central Limit Theorem, the sum of a random sample of $n$ observations, $\Sigma x$, will possess a sampling distribution that is approximately normal for large samples. This distribution will have a mean equal to $n\mu$ and a variance equal to $n\sigma^2$. Proof of the Central Limit Theorem is beyond the scope of this book, but it can be found in many mathematical statistics texts.

## Colored Boxes Highlight Important Information

- Definitions, Strategies, Key Formulas, and other important information are highlighted.
- Prepare for quizzes and tests by reviewing the highlighted information.

## Interesting Examples with Solutions

- Examples, with complete solutions and explanations, illustrate every concept. Work through the solution carefully to prepare for the section exercise set.
- All examples are numbered for easy reference.
- The end of the solution is marked with a ▲ symbol.



As another example of data from which the central tendency is better described by the median than the mean, consider the household incomes of a community being studied by a sociologist. The presence of just a few households with very high incomes will affect the mean more than the median. Thus, the median will provide a more accurate picture of the typical income for the community. The mean could exceed the vast majority of the sample measurements (household incomes), making it a misleading measure of central tendency.

**EXAMPLE 2.6** Calculate the median for the 100 EPA mileages given in Table 2.3. Compare the median to the mean computed in Example 2.4.

**Solution**

For this large data set, we again resort to a computer analysis. The SPSS printout is reproduced in Figure 2.15, with the median shaded. You can see that the median is 37.0. This value implies that half of the 100 mileages in the data set fall below 37.0 and half lie above 37.0. Note that the median, 37.0, and the mean, 36.994, are almost equal. This fact indicates that the data form an approximately **symmetric distribution**. As indicated in the box on page 47, a comparison of the mean and median gives an indication of the **skewness** (i.e., the tendency of the distribution to have elongated tails) of a data set. ▲

**FIGURE 2.15**
SPSS printout of numerical descriptive measures for 100 EPA mileages

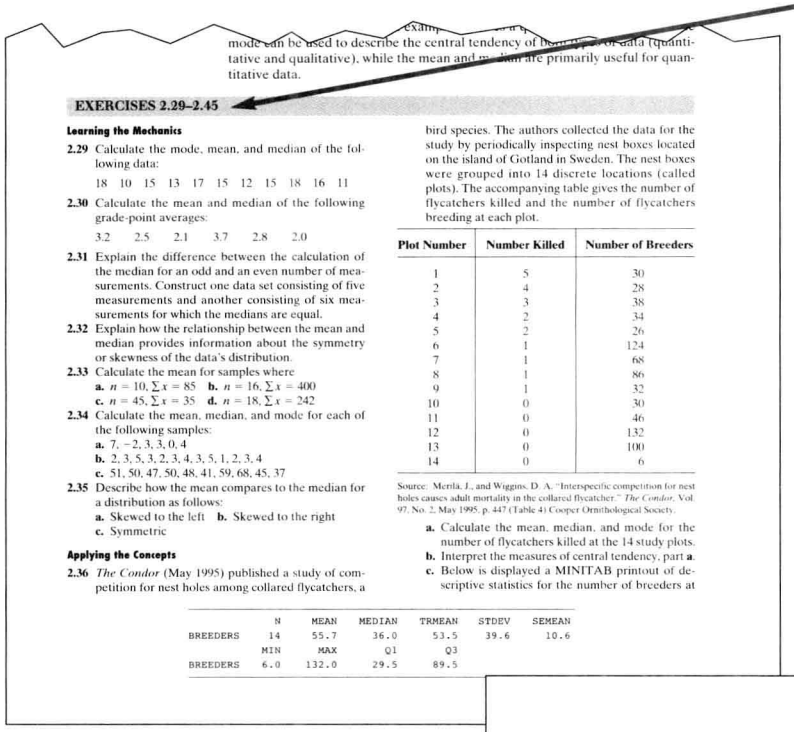
| | | | | | |
|---|---|---|---|---|---|
| Mean | 36.994 | Std Err | .242 | Median | 37.000 |
| Mode | 37.000 | Std Dev | 2.418 | Variance | 5.846 |
| Kurtosis | .770 | S E Kurt | .478 | Skewness | .051 |
| S E Skew | .241 | Range | 14.900 | Minimum | 30.000 |
| Maximum | 44.900 | Sum | 3699.400 | | |
| Valid Cases | 100 | Missing Cases | 0 | | |

## Computer Output Integrated Throughout

- Statistical software packages, such as SPSS, Minitab, SAS, or ASP crunch data quickly so you can spend time analyzing the results. Learning how to interpret statistical output will prove helpful in future classes or on the job.
- When computer output appears in examples, the solution explains how to read and interpret the output.

## Lots of Exercises for Practice

- Every section in the book is followed by an Exercise Set divided into two parts.
- **Learning the Mechanics** has straightforward applications of new concepts. Test your mastery of definitions, concepts, and basic computation. Make sure you can answer all of these questions before moving on.
- **Applying the Concepts** tests your understanding of concepts and requires you to apply statistical techniques in solving real world problems. Spending time on these problems will help you develop good problem-solving skills.



mode can be used to describe the central tendency of [illegible] data (quantitative and qualitative), while the mean and median are primarily useful for quantitative data.

**EXERCISES 2.29–2.45**

**Learning the Mechanics**

**2.29** Calculate the mode, mean, and median of the following data:

18 10 15 13 17 15 12 15 18 16 11

**2.30** Calculate the mean and median of the following grade-point averages:

3.2 2.5 2.1 3.7 2.8 2.0

**2.31** Explain the difference between the calculation of the median for an odd and an even number of measurements. Construct one data set consisting of five measurements and another consisting of six measurements for which the medians are equal.

**2.32** Explain how the relationship between the mean and median provides information about the symmetry or skewness of the data's distribution.

**2.33** Calculate the mean for samples where
**a.** $n = 10, \sum x = 85$ **b.** $n = 16, \sum x = 400$
**c.** $n = 45, \sum x = 35$ **d.** $n = 18, \sum x = 242$

**2.34** Calculate the mean, median, and mode for each of the following samples:
**a.** 7, −2, 3, 3, 0, 4
**b.** 2, 3, 5, 3, 2, 3, 4, 3, 5, 1, 2, 3, 4
**c.** 51, 50, 47, 50, 48, 41, 59, 68, 45, 37

**2.35** Describe how the mean compares to the median for a distribution as follows:
**a.** Skewed to the left **b.** Skewed to the right
**c.** Symmetric

**Applying the Concepts**

**2.36** *The Condor* (May 1995) published a study of competition for nest holes among collared flycatchers, a bird species. The authors collected the data for the study by periodically inspecting nest boxes located on the island of Gotland in Sweden. The nest boxes were grouped into 14 discrete locations (called plots). The accompanying table gives the number of flycatchers killed and the number of flycatchers breeding at each plot.

| Plot Number | Number Killed | Number of Breeders |
|---|---|---|
| 1 | 5 | 30 |
| 2 | 4 | 28 |
| 3 | 3 | 38 |
| 4 | 2 | 34 |
| 5 | 2 | 26 |
| 6 | 1 | 124 |
| 7 | 1 | 68 |
| 8 | 1 | 86 |
| 9 | 1 | 32 |
| 10 | 0 | 30 |
| 11 | 0 | 46 |
| 12 | 0 | 132 |
| 13 | 0 | 100 |
| 14 | 0 | 6 |

Source: Merila, J., and Wiggins, D. A. "Interspecific competition for nest holes causes adult mortality in the collared flycatcher." *The Condor*, Vol. 97, No. 2, May 1995, p. 447 (Table 4) Cooper Ornithological Society.

**a.** Calculate the mean, median, and mode for the number of flycatchers killed at the 14 study plots.
**b.** Interpret the measures of central tendency, part **a**.
**c.** Below is displayed a MINITAB printout of descriptive statistics for the number of breeders at each plot. Locate the measures of central tendency on the printout and interpret these values.

| | N | MEAN | MEDIAN | TRMEAN | STDEV | SEMEAN |
|---|---|---|---|---|---|---|
| BREEDERS | 14 | 55.7 | 36.0 | 53.5 | 39.6 | 10.6 |
| | MIN | MAX | Q1 | Q3 | | |
| BREEDERS | 6.0 | 132.0 | 29.5 | 89.5 | | |



**2.37** The conventional method of measuring the refractive status of an eye involves three quantities: (1) sphere power, (2) cylinder power, and (3) axis. Optometric researchers at a Johannesburg (South Africa) university studied the variation in these three measures of refraction (*Optometry and Vision Science*, June 1995). Twenty-five successive refractive measurements (using a single Topcon RM-A6000 autorefractor) were obtained on the eyes of more than 100 university students. The cylinder power measurements for the left eye of one particular student (ID #11) are listed in the table. [*Note*: All measurements are negative values.] Numerical descriptive measures for the data set are provided in the accompanying SAS printout.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| .08 | .08 | 1.07 | .09 | .16 | .04 | .07 | .17 | .11 |
| .06 | .12 | .17 | .20 | .12 | .17 | .09 | .07 | .16 |
| .15 | .16 | .09 | .06 | .10 | .21 | .06 | | |

Source: Rubin, A., and Harris, W. F. "Refractive variation during autorefraction: Multivariate distribution of refractive status." *Optometry and Vision Science*, Vol. 72, No. 6, June 1995, p. 409 (Table 4).

UNIVARIATE PROCEDURE
Variable=CYLPOWER

Moments

| | | | |
|---|---|---|---|
| N | 25 | Sum Wgts | 25 |
| Mean | −0.1544 | Sum | −3.86 |
| Std Dev | 0.196767 | Variance | 0.038717 |
| Skewness | −4.52208 | Kurtosis | 21.70196 |
| USS | 1.5252 | CSS | 0.929216 |
| CV | −127.44 | Std Mean | 0.039353 |
| T:Mean=0 | −3.92342 | Prob>\|T\| | 0.0006 |
| Sgn Rank | −162.5 | Prob>\|S\| | 0.0001 |
| Num ^= 0 | 25 | | |

Quantiles(Def=5)

| | | | |
|---|---|---|---|
| 100% Max | −0.04 | 99% | −0.04 |
| 75% Q3 | −0.08 | 95% | −0.06 |
| 50% Med | −0.11 | 90% | −0.06 |
| 25% Q1 | −0.16 | 10% | −0.2 |
| 0% Min | −1.07 | 5% | −0.21 |
| | | 1% | −1.07 |
| Range | 1.03 | | |
| Q3−Q1 | 0.08 | | |
| Mode | −0.17 | | |

Extremes

| Lowest | Obs | Highest | Obs |
|---|---|---|---|
| −1.07( | 3) | −0.07( | 17) |
| −0.21( | 24) | −0.06( | 10) |
| −0.2( | 13) | −0.06( | 22) |
| −0.17( | 15) | −0.06( | 25) |
| −0.17( | 12) | −0.04( | 6) |

**a.** Locate the measures of central tendency on the printout and interpret their values.
**b.** Note that the data contain one unusually large (negative) cylinder power measurement relative to the other measurements in the data set. Find this measurement, called an **outlier**.
**c.** Delete the outlier, part **b**, from the data set and recalculate the measures of central tendency. Which measure is most affected by the elimination of the outlier?

**2.38** Demographics plays a key role in the recreation industry. According to the *Journal of Leisure Research* (Vol. 23, 1991), difficult times lay ahead for the industry. The article reports that the median age of the population in the United States was 30 in 1980, but will be about 36 by the year 2000.
**a.** Interpret the value of the median for both 1980 and 2000 and explain the trend.
**b.** If the recreation industry relies on the 18–30 age group for much of its business, what effect will this shift in the median age have on the recreation industry? Explain.

**2.39** Applicants for an academic position (e.g., assistant professor) at a college or university are usually required to submit at least three letters of recommendation. A recent study of 148 applicants for an entry-level position in experimental psychology at the University of Alaska Anchorage revealed that many did not meet the three-letter requirement (*American Psychologist*, July 1995). Summary statistics for the number of recommendation letters in each application are given below. Interpret these summary measures.

Mean = 2.28
Median = 3
Mode = 3

**2.40** Platelet-activating factor (PAF) is a potent chemical that occurs in patients suffering from shock, inflammation, hypotension, and allergic responses as well as respiratory and cardiovascular disorders. Consequently, drugs that effectively inhibit PAF, keeping it from binding to human cells, may be successful in treating these disorders. A bioassay was undertaken to investigate the potential of 17 traditional Chinese herbal drugs in PAF inhibition (*Progress in Natural Science*, June 1995). The prevention of the PAF binding process, measured as a percentage, for each drug is provided in the accompanying table.
**a.** Construct a stem-and-leaf display for the data.
**b.** Compute the median inhibition percentage for the 17 herbal drugs. Interpret the result.
**c.** Compute the mean inhibition percentage for the 17 herbal drugs. Interpret the result.
**d.** Compute the mode of the 17 inhibition percentages. Interpret the result.

## Real Data

- Most of the exercises contain data or information taken from newspaper articles, magazines, and journals published since 1990. Statistics are all around you.

## Computer Output

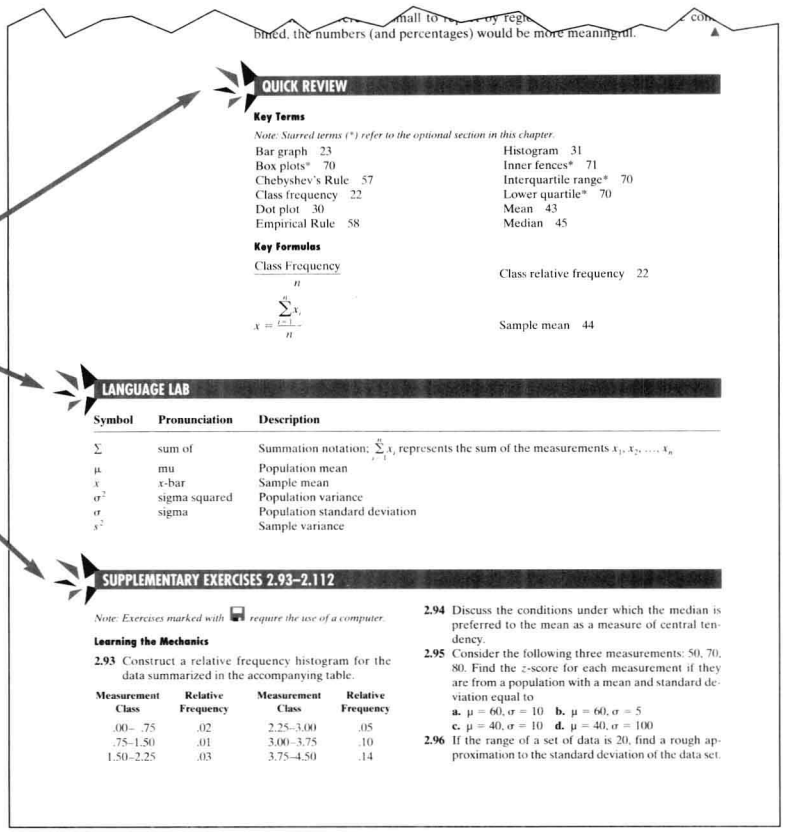
- Computer output screens appear in the exercise sets to give you practice in interpretation.

## End of Chapter Review

- Each chapter ends with information designed to help you check your understanding of the material, study for tests, and expand your knowledge of statistics.
- **Quick Review** provides a list of key terms and formulas with page number references.
- **Language Lab** helps you learn the language of statistics through pronunciation guides, descriptions of symbols, names, etc.
- **Supplementary Exercises** review all of the important topics covered in the chapter.

Exercises marked with 💾 require a computer for solution.

Data sets for use with the 💾 problems are available on disk.



...small to ... by regi... con...
bined, the numbers (and percentages) would be more meaningful. ▲

**QUICK REVIEW**

**Key Terms**

*Note: Starred terms (*) refer to the optional section in this chapter.*

Bar graph 23
Box plots* 70
Chebyshev's Rule 57
Class frequency 22
Dot plot 30
Empirical Rule 58
Histogram 31
Inner fences* 71
Interquartile range* 70
Lower quartile* 70
Mean 43
Median 45

**Key Formulas**

$\frac{\text{Class Frequency}}{n}$ Class relative frequency 22

$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ Sample mean 44

**LANGUAGE LAB**

| Symbol | Pronunciation | Description |
|---|---|---|
| $\Sigma$ | sum of | Summation notation; $\sum_{i=1}^{n} x_i$ represents the sum of the measurements $x_1, x_2, \ldots, x_n$ |
| $\mu$ | mu | Population mean |
| $\bar{x}$ | x-bar | Sample mean |
| $\sigma^2$ | sigma squared | Population variance |
| $\sigma$ | sigma | Population standard deviation |
| $s^2$ | | Sample variance |

**SUPPLEMENTARY EXERCISES 2.93–2.112**

*Note: Exercises marked with 💾 require the use of a computer.*

**Learning the Mechanics**

**2.93** Construct a relative frequency histogram for the data summarized in the accompanying table.

| Measurement Class | Relative Frequency | Measurement Class | Relative Frequency |
|---|---|---|---|
| .00–.75 | .02 | 2.25–3.00 | .05 |
| .75–1.50 | .01 | 3.00–3.75 | .10 |
| 1.50–2.25 | .03 | 3.75–4.50 | .14 |

**2.94** Discuss the conditions under which the median is preferred to the mean as a measure of central tendency.

**2.95** Consider the following three measurements: 50, 70, 80. Find the $z$-score for each measurement if they are from a population with a mean and standard deviation equal to

**a.** $\mu = 60, \sigma = 10$ **b.** $\mu = 60, \sigma = 5$
**c.** $\mu = 40, \sigma = 10$ **d.** $\mu = 40, \sigma = 100$

**2.96** If the range of a set of data is 20, find a rough approximation to the standard deviation of the data set.

**STUDENT PROJECTS**

Choose a population pertinent to your major area of interest—a population that has an unknown mean or, if the population is binomial, that has an unknown probability of success. For example, a marketing major may be interested in the proportion of consumers who prefer a certain product. A sociology major may be interested in estimating the proportion of people in a particular socioeconomic group or the mean income of people living in a particular part of a city. A political science major may wish to estimate the proportion of an electorate in favor of a certain candidate, a certain amendment, or a certain presidential policy. A pre-med student might want to find the average length of time patients stay in the hospital or the average number of people treated daily in the emergency room. We could continue with examples, but the point should be clear—choose something of interest to you.

Define the parameter you want to estimate and conduct a *pilot study* to obtain an initial estimate of the parameter of interest and, more importantly, an estimate of the variability associated with the estimator. A pilot study is a small experiment (perhaps 20 to 30 observations) used to gain some information about the population of interest. The purpose of the study is to help plan more elaborate future experiments. Using the results of your pilot study, determine the sample size necessary to estimate the parameter to within a reasonable bound (of your choice) with a 95% confidence interval.

**EXPLORING DATA WITH A COMPUTER**

Refer to Exploring Data with a Computer in Chapter 4. Recall the values of the "population" mean $\mu$ and standard deviation $\sigma$ for the 962 FTC measurements on tar, nicotine, or carbon monoxide in cigarette smoke. Suppose our objective is to sample from this population and to estimate the mean $\mu$ using a 95% confidence interval.

a. Determine the sample size $n_1$ necessary to estimate $\mu$ to within 1 milligram with 95% confidence. Then generate one hundred 95% confidence intervals by repeatedly drawing samples of size $n_1$ (with replacement) from the 962 measurements and using the sample statistics to form a confidence interval. Treat $\sigma$ as unknown when forming the confidence intervals. What percentage of confidence intervals will contain $\mu$?
b. Determine the sample size $n_2$ necessary to estimate $\mu$ to within .5 milligram with 95% confidence. Then generate one hundred 95% confidence intervals by repeatedly drawing samples of size $n_2$ (with replacement) from the 962 measurements and using the sample statistics to form a confidence interval. Treat $\sigma$ as unknown when forming the confidence intervals. What percentage of confidence intervals will contain $\mu$?
c. Repeat part a, but this time use an 80% confidence interval.

**Student Projects** provide challenging projects for further exploration by yourself or with a group of students. These projects give you good practice in gathering and analyzing data and report writing—skills that will be important in future classes and in the workplace.

**Exploring Data with a Computer** Five large databases available on disk and referenced in Appendix B are the basis for statistical explorations using a statistical software package.

# CONTENTS