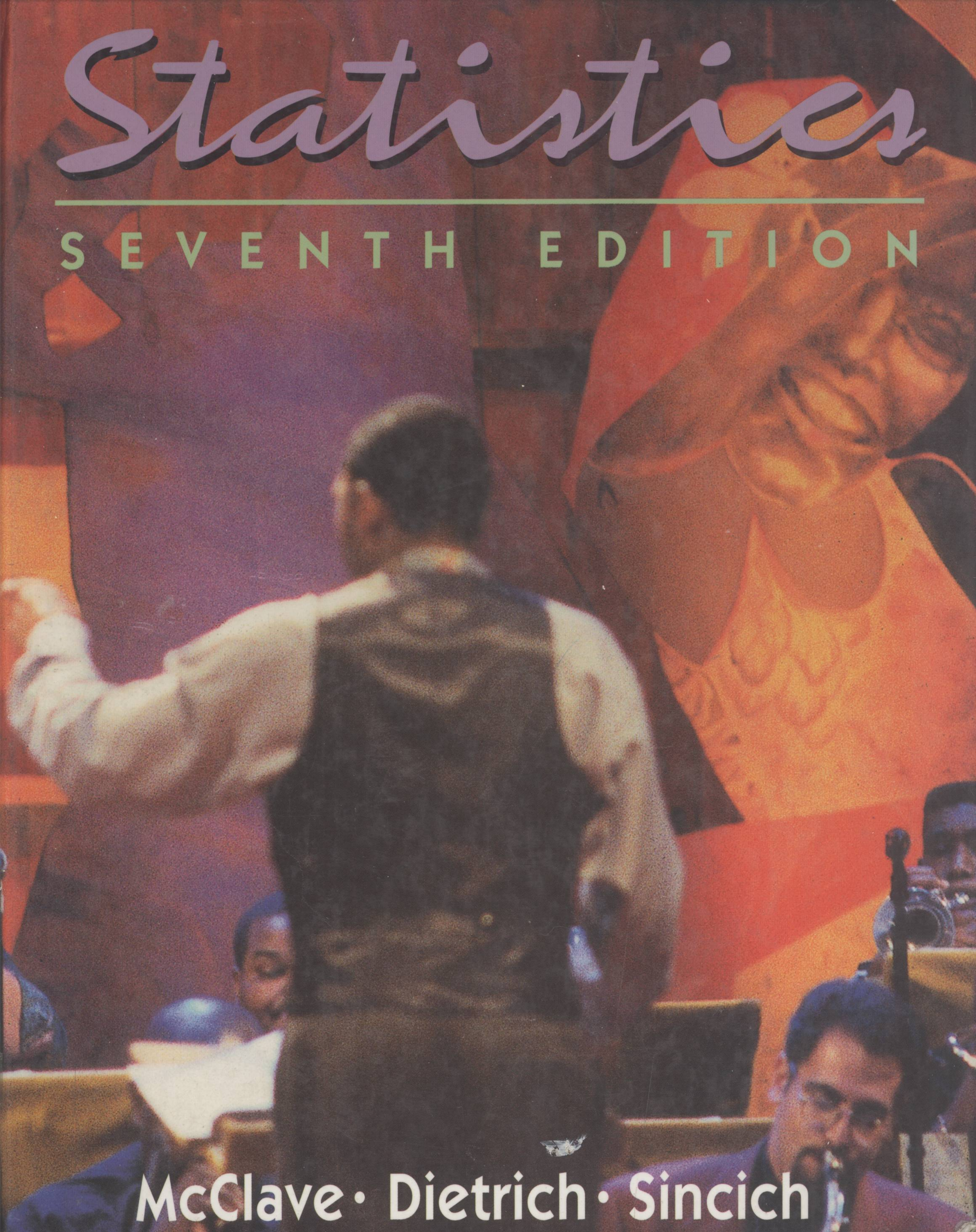
Statistics

SEVENTH EDITION



McClave · Dietrich · Sincich

0212
M478-2
E-7
贈閱

Statistics

Seventh Edition

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

Frank H. Dietrich, II
Northern Kentucky University

Terry Sincich
University of South Florida

华南理工大学图书馆藏书
E200200620

PRENTICE HALL
Upper Saddle River, New Jersey 07458

Library of Congress Cataloging-in-Publication Data

McClave, James T.

Statistics / James T. McClave, Frank H. Dietrich II, Terry Sincich. — 7th ed.

p. cm.

Includes index.

ISBN 0-13-471542-X

ISBN 0-13-471657-4 (Annotated Instructor's Edition)

1. Statistics. I. Dietrich, Frank H. II. Sincich, Terry.

III. Title.

QA276.12.M4 1997

519.5—dc20 96-12002

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: Katheen 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: Wynton Marsalis in performance at Tanglewood. Photo by Frank Stewart
Interior Photos:
Pages ix, 1, 21: Paul Rees/Tony Stone Images
Pages 91, 155, 193, 227, 685, 715: Frank Stewart
Pages 253, 289, 339: Mark E. Gibson/Visuals Unlimited
Pages 401, 473, 535, 605: Mitch Wojnarowicz/The Image Works

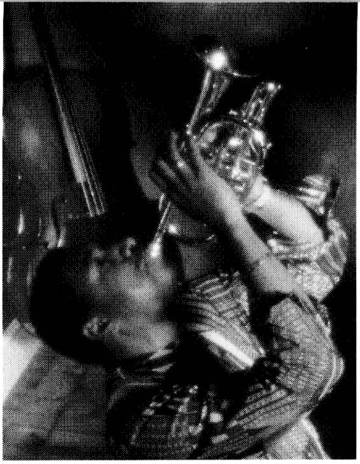
© 1997 by Prentice-Hall, Inc.
Simon & Schuster / A Viacom Company
Upper Saddle River, NJ 07458

All rights reserved. No part of this book may be reproduced, in any form or by any means, without permission in writing from the publisher.

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

ISBN 0-13-471542-X

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 seventh edition of *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 an introductory text emphasizing inference, with extensive coverage of data collection and analysis as needed to evaluate the reported results of statistical studies. It assumes a mathematical background of basic algebra.

NEW IN THE SEVENTH 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.
- Chapter 4 now includes an optional section on the hypergeometric random variable.
- Chapters 7 through 9 more heavily stress confidence intervals and rely more heavily on computer output than on formulas.
- Chapter 8 now provides an equal balance between applications relying on p-values and those relying on critical values in their interpretation.

- Chapter 10 contains a separate section on multiple comparisons with emphasis on computer output rather than formulas. Examples of both Bonferroni's and Tukey's method are presented.
- Chapter 11 now incorporates computer printouts throughout the discussion of simple linear regression.

Pedagogy

- All new, restructured cases, approximately two per chapter, summarize statistical studies on contemporary, controversial issues and include questions to the student 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 *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 1,200 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.

此为试读，需要完整PDF请访问：www.ertongbook.com

A Choice in Level of Coverage of Probability

One of the most troublesome aspects of an introductory statistics course is the study of probability. Probability poses a challenge for instructors because they must decide on the level of presentation, and students find it a difficult subject to comprehend. We believe that one cause for these problems is the mixture of probability and counting rules that occurs in most introductory texts. We have included the counting rules in a separate and optional section at the end of the chapter on probability. In addition, all exercises that require the use of counting rules are marked with an asterisk (*). Thus, the instructor can control the level of coverage of probability.

Extensive Coverage of Multiple Regression Analysis and Model Building

This topic represents one of the most useful statistical tools for the solution of applied problems. Although an entire text could be devoted to regression modeling, we feel that we have presented coverage that is understandable, usable, and much more comprehensive than the presentations in other introductory statistics texts. We devote three chapters to discussing the major types of inferences that can be derived from a regression analysis, showing how these results appear in computer printouts and, most important, selecting multiple regression models to be used in an analysis. Thus, the instructor has the choice of a one-chapter coverage of simple regression, a two-chapter treatment of simple and multiple regression, or a complete three-chapter coverage of simple regression, multiple regression, and model building. This extensive coverage of such useful statistical tools will provide added evidence to the student of the relevance of statistics to the solution of applied problems.

Footnotes

Although the text is designed for students with a non-calculus background, footnotes explain the role of calculus in various derivations. Footnotes are also used to inform the student about some of the theory underlying certain results. The footnotes allow additional flexibility in the mathematical and theoretical level at which the material is presented.

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 have contributed to the seventh edition:

Reviewers of Previous Editions of *Statistics*

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 the seventh 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 seventh 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.

NEW! *Annotated Instructor's Edition* (AIE) (ISBN 0-13-471657-4)

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)

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 this text.

Instructor's Solutions Manual by Nancy S. Boudreau (ISBN 0-13-471616-7)

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 Solution Manual.*

Test Bank by Mark Dummeldinger (ISBN 0-13-472201-9)

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 0-13-471673-6)
Mac PH Custom Test (0-13-472044-X)

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.

NEW! PowerPoint Presentation Disk (ISBN 0-13-472151-9)

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 on the software disk are: learning objectives, thinking challenges, concept presentation slides, and examples with worked out solutions.

Data Disk (ISBN 0-13-472036-9)

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 viii–x.

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 0-13-471666-3)

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 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.

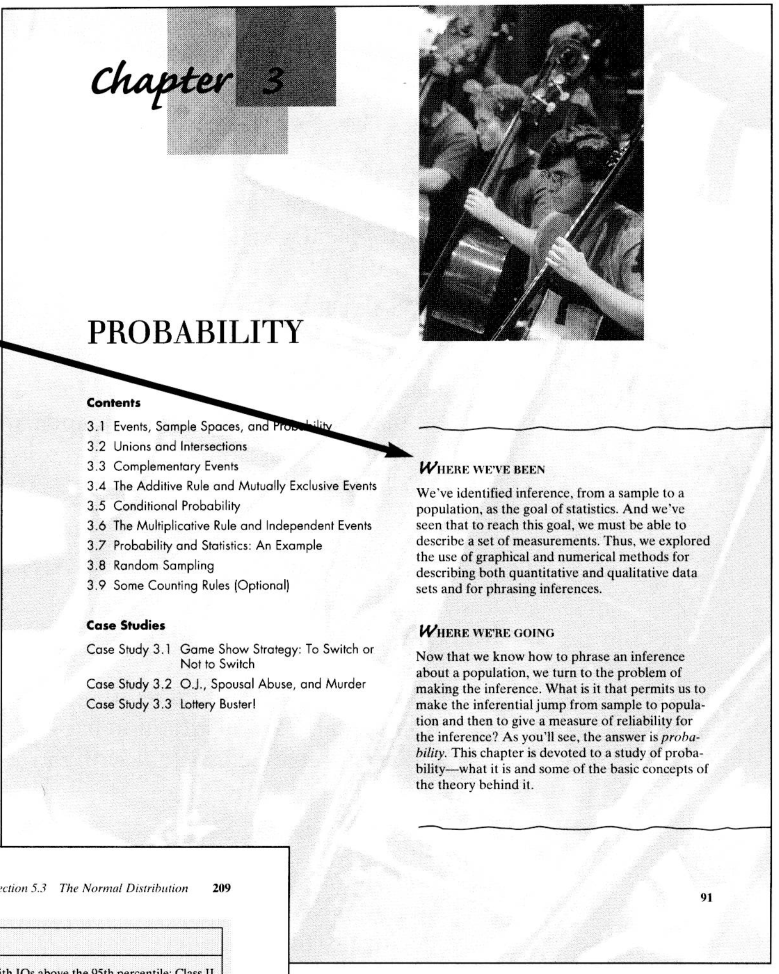
How to Use This Book

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

Section 5.3 The Normal Distribution 209

IQ and the Bell Curve

CASE STUDY 5.1

In their controversial book *The Bell Curve* (Free Press, 1994), Professors Richard J. Herrnstein (a Harvard psychologist) and Charles Murray (a political scientist at MIT) explore, as the subtitle states, "intelligence and class structure in American life." *The Bell Curve* heavily employs statistical analyses in an attempt to support the authors' positions. One of the biggest controversies sparked by the book is the authors' tenet that intelligence levels differ across ethnic groups and that intelligence (or lack thereof) is a cause of a wide range of intractable social problems. Herrnstein and Murray warn, however, that "these group differences do not justify prejudicial assumptions about any member of a given group since his or her intelligence and potential may, in fact, be anywhere under the bell curve." For this case study, we focus on the distribution of intelligence, that is, the **bell curve**.

The measure of intelligence chosen by the authors is the well known Intelligent Quotient (IQ). Numerous tests have been developed to measure IQ; Herrnstein and Murray use the Armed Forces Qualification Test (AFQT), originally designed to measure the cognitive ability of military recruits. Psychologists traditionally treat IQ as a random variable having a normal distribution with mean $\mu = 100$ and standard deviation $\sigma = 15$. This distribution, or *bell curve*, is shown in Figure 5.17.

In their book, Herrnstein and Murray refer to five cognitive classes of people defined by percentiles of the normal distribution. Class I ("very bright") consists of those with IQs above the 95th percentile; Class II ("bright") are those with IQs between the 75th and 95th percentiles; Class III ("normal") includes IQs between the 25th and 75th percentiles; Class IV ("dull") are those with IQs between the 5th and 25th percentiles; and Class V ("very dull") are IQs below the 5th percentile. These classes are also illustrated in Figure 5.17.

Focus

a. Assuming that the distribution of IQ is accurately represented by the bell curve in Figure 5.17, determine the proportion of people with IQs in each of the five cognitive classes defined by Herrnstein and Murray.
b. Although the cognitive classes above are defined in terms of percentiles, the authors stress that IQ scores should be compared with z-scores, not percentiles. In other words, it is more informative to give the difference in z-scores for two IQ scores than it is to give the difference in percentiles. To demonstrate this point, calculate the difference in z-scores for IQs at the 50th and 55th percentiles. Do the same for IQs at the 94th and 99th percentiles. What do you observe?
c. Researchers have found that scores on many intelligence tests are decidedly nonnormal. Some distributions are skewed toward higher scores, others toward lower scores. How would the proportions in the five cognitive classes differ for an IQ distribution that is skewed right? Skewed left?

5.23 The random variable x has a normal distribution with $\mu = 300$ and $\sigma = 30$.
a. Find the probability that x assumes a value more than 2 standard deviations from its mean. More than 3 standard deviations from μ.
b. Find the probability that x assumes a value within 1 standard deviation of its mean. Within 2 standard deviations of μ.
c. Find the value of x that represents the 80th percentile of this distribution. The 10th percentile.

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.

此为试读，需要完整PDF请访问：www.ertongbook.com

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.

Section 5.3 The Normal Distribution **205**

Steps for Finding a Probability Corresponding to a Normal Random Variable

1. Sketch the normal distribution and indicate the mean of the random variable x. Then shade the area corresponding to the probability you want to find.
2. Convert the boundaries of the shaded area from x values to standard normal random variable z values using the formula

$$z = \frac{x - \mu}{\sigma}$$

Show the z values under the corresponding x values on your sketch.

3. Use Table IV in Appendix A (and inside the front cover) to find the areas corresponding to the z values. If necessary, use the symmetry of the normal distribution to find areas corresponding to negative z values and the fact that the total area on each side of the mean equals .5 to convert the areas from Table IV to the probabilities of the event you have shaded.

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.

EXAMPLE 5.7 Suppose an automobile manufacturer introduces a new model that has an advertised mean in-city mileage of 27 miles per gallon. Although such advertisements seldom report any measure of variability, suppose you write the manufacturer for the details of the tests, and you find that the standard deviation is 3 miles per gallon. This information leads you to formulate a probability model for the random variable x, the in-city mileage for this car model. You believe that the probability distribution of x can be approximated by a normal distribution with a mean of 27 and a standard deviation of 3.

a. If you were to buy this model of automobile, what is the probability that you would purchase one that averages less than 20 miles per gallon for in-city driving? In other words, find $P(x < 20)$.

b. Suppose you purchase one of these new models and it does get less than 20 miles per gallon for in-city driving. Should you conclude that your probability model is incorrect?

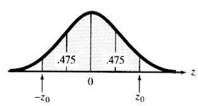
Solution

a. The probability model proposed for x, the in-city mileage, is shown in Figure 5.13. We are interested in finding the area A to the left of 20 since this area corresponds to the probability that a measurement chosen from this distribution falls below 20. In other words, if this model is correct, the area A represents the fraction of cars that can be expected to get less than 20 miles per gallon for in-city driving. To find A, we first calculate the z value corresponding to $x = 20$. That is,

$$z = \frac{x - \mu}{\sigma} = \frac{20 - 27}{3} = -\frac{7}{3} = -2.33$$

FIGURE 5.13

Area under the normal curve for Example 5.7



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 data 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 applicatons of new concepts. Test your mastery of definitons, 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 both types of 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. $n = 10, \Sigma x = 85$ **b.** $n = 16, \Sigma x = 400$
c. $n = 45, \Sigma x = 35$ **d.** $n = 18, \Sigma x = 242$

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: Merilä, 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
```

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.

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.

End of Chapter Review

- Each chapter ends with information designed to help you check you 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.

where all of the choices are correct. For each person, two scores (ranging from 0 to 20) are obtained, an attitude score (KAE-A) and a general knowledge score (KAE-GK). Based on a large-scale study of college students, the distribution of KAE-A scores

less than 6.

e. Suppose you observe a sample mean KAE score of 6.5. Is this result more likely to be the value of $\bar{x}_A$ or $\bar{x}_{GK}$? Explain.

QUICK REVIEW

Key Terms

Biased estimate 235
Central Limit Theorem 240
Error of estimation 235
Parameter 228
Point estimator 234
Sample statistic 228
Sampling distribution 229
Standard error of the mean 240
Unbiased estimate 235

Key Formulas

	Mean	Standard Deviation	z-score
Sampling distribution of $\bar{x}$	$\mu_{\bar{x}} = \mu$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$	$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

LANGUAGE LAB

Symbol	Pronunciation	Description
θ	theta	Population parameter (general)
$\mu_{\bar{x}}$	mu of *x*-bar	True mean of sampling distribution of $\bar{x}$
$\sigma_{\bar{x}}$	sigma of *x*-bar	True standard deviation of sampling distribution of $\bar{x}$

SUPPLEMENTARY EXERCISES 6.31–6.50

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

Learning the Mechanics

6.31 Consider a sample statistic *A*. As with all sample statistics, *A* is computed by utilizing a specified function (formula) of the sample measurements. (For example, if *A* were the sample mean, the specified formula would be to sum the measurements and divide by the number of measurements.)

a. Describe what we mean by the phrase "the sampling distribution of the sample statistic *A*."

STUDENT PROJECTS

To understand the Central Limit Theorem and sampling distribution, consider the following experiment: Toss four identical coins, and record the number of heads observed. Then repeat this experiment four more times, so that you end up with a total of five observations for the random variable *x*, the number of heads when four coins are tossed.

Now derive and graph the probability distribution for *x*, assuming the coins are balanced. Note that the mean of this distribution is $\mu = 2$ and the standard deviation is $\sigma = 1$. This probability distribution represents the one from which you are drawing a random sample of five measurements.

Next, calculate the mean $\bar{x}$ of the five measurements—that is, calculate the mean number of heads you observed in five repetitions of the experiment. Although you have repeated the basic experiment five times, you have only one observed value of $\bar{x}$. To derive the probability distribution or sampling distribution of $\bar{x}$ empirically, you have to repeat the entire process (of tossing four coins five times) many times. Do it 100 times.

The approximate sampling distribution of $\bar{x}$ can be derived theoretically by making use of the Central Limit Theorem. We expect at least an approximate normal probability distribution with a mean $\mu = 2$ and a standard deviation

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{5}} = .45$$

Count the number of your 100 $\bar{x}$'s that fall in each of the intervals in the figure on page 251. Use the normal prob-

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.

EXPLORING DATA WITH A COMPUTER

The Federal Trade Commission (FTC) annually ranks American cigarette brands in terms of the amount of three hazardous substances—tar, nicotine, and carbon monoxide. The test results are obtained as follows: A sequential smoking machine is used to "smoke" cigarettes to a 23-millimeter butt length. Based on tests of 100 cigarettes per brand, the carbon monoxide, tar, and nicotine concentrations (rounded to the nearest milligram) in the residual "dry" particulate matter are determined.

Appendix B contains the results of the FTC's 1995 tests of 962 domestic cigarette brands. Select one of the three hazardous substances (e.g., tar). Use a statistical software package to obtain the mean and standard deviation of the 962 tar amounts. Assume that these quantities represent the population mean μ and the population standard deviation σ for all domestic cigarette brands.

a. Draw 100 random samples of $n = 20$ observations from the 962 tar amounts. Select the samples with replacement—i.e., replace each measurement before selecting the next.* Calculate the 100 sample means. Generate a stem-and-leaf display or a relative frequency histogram for the 100 means. Then count the number of the 100 sample means that fall in the intervals $\mu \pm \sigma/\sqrt{n}$, $\mu \pm 2\sigma\sqrt{n}$, and $\mu \pm 3\sigma\sqrt{n}$. How do the graphical description and the percentage of means falling in the intervals agree with a normal distribution having mean μ and standard deviation $\sigma/\sqrt{n}$?

b. Repeat part a using a sample size of $n = 50$. Is the sampling distribution of the sample means closer to normal for the larger sample size?

CONTENTS

此为试读，需要完整PDF请访问：www.ertongbook.com