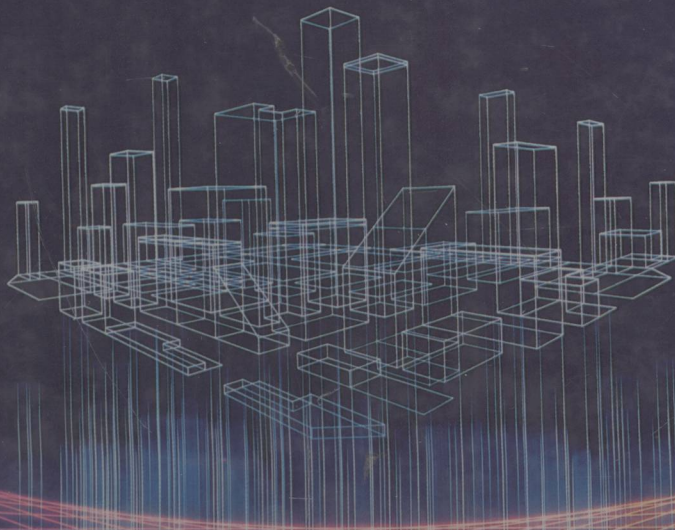


# Statistical Thinking



*for Managers*

Third Edition



David K. Hildebrand ♦ Lyman Ott

F270  
H642  
E-3



# STATISTICAL THINKING FOR MANAGERS

THIRD EDITION

DAVID K. HILDEBRAND

University of Pennsylvania

LYMAN OTT

Marion Merrell Dow



E200200621



PWS-KENT Publishing Company  
BOSTON



**PWS-KENT**  
Publishing Company

Copyright ©1991 by PWS-KENT Publishing Company. Copyright ©1987, 1983 by PWS Publishers. All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transcribed, in any form or by any means—electronic, mechanical, photocopying, recording, or otherwise—without the prior written permission of PWS-KENT Publishing Company, 20 Park Plaza, Boston, Massachusetts 02116.

PWS-KENT Publishing Company is a division of Wadsworth, Inc.

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

Hildebrand, David K., 1940—

Statistical thinking for managers/David K. Hildebrand, Lyman Ott.—3rd ed.

p. cm.

Includes bibliographical references and index.

ISBN 0-534-92561-8

1. Industrial management—Statistical methods. 2. Industrial management—Statistical methods—Data processing. 3. Economics—Statistical methods. 4. Statistics—Data processing. I. Ott, Lyman. II. Title.

HD30.215.H54 1991

519.5'024658—dc20

90-19988

CIP

International Student Edition ISBN: 0-534-98253-0

Printed in the United States of America

1 2 3 4 5 6 7 8 9—95 94 93 92 91

Sponsoring editor: Michael Payne

Assistant editor: Marcia Cole

Editorial assistant: Patricia Schumacher

Production editor: Pamela Rockwell

Manufacturing coordinator: Margaret Sullivan Higgins

Interior design: Chris Crochetière

Cover design: Jean Hammond

Cover illustration: K. Kitagawa/SuperStock, Inc.

Typesetting: Polyglot Pte Ltd

Cover printing: John P. Pow Company, Inc.

Printing and binding: R. R. Donelley & Sons Company

## Glossary of Symbols: Roman Letters

Symbol	Meaning	Chapter(s)
$S_{xx}$	Sum of squared $X$ deviations	13
$S_{xy}$	Sum of products of $X$ and $Y$ deviations	13
$S$	Sample space	3
$T$	Sum of ranks	10
$T_+$	Sum of positive ranks	10
$T_-$	Sum of negative ranks	11
$t_a$	Tabled $t$ value, right-tail area $\alpha$	8–15
$t$	A $t$ statistic	8–15
T.S.	Test statistic	9–15
$U$	Number of errors with unknown independent variable	11
$U$	Utility	17
$v(a_i, \theta_j)$	Payoff of action $a_i$ for state of nature $\theta_j$	17
$\text{Var}(X)$	Variance of random variable $X$	4ff
$X$	Random variable	4–6
$X, x$	Independent variable	13–15
$x_i$	$i$ th value of $X$	13–15
$x_{n+1}$	Value of $X$ for which prediction is made	13
$X$	Matrix of independent variables' values	14
$\bar{X}, \bar{x}$	Average of $x$ values	15, 18
$Y$	Random variable	4ff
$Y, y$	Dependent variable	13–15
$Y_{n+1}$	Actual $Y$ value at $x_{n+1}$	13–15
$\hat{Y}, \hat{y}$	Predicted $Y$ value	13–15
$\bar{Y}, \bar{y}$	Average of $Y$ values	all
$z$	Standardized normal statistic	5ff
$z_a$	Tabled standardized normal value, right-tail area $\alpha$	5ff

m 1230-2.30 S-63  
wed

## Glossary of Symbols: Greek Letters

Symbol	Meaning	Chapter(s)
$\alpha$	Probability of Type I error	9-15
$\alpha$	Smoothing constant	16
$\beta$	Probability of Type II error	9-15
$\beta_0$	True (population) intercept	13-15
$\beta_1$	True slope	13
$\beta_i$	True partial slope	14, 15
$\hat{\beta}_0$	Estimated intercept	13-15
$\hat{\beta}_1$	Estimated slope	13
$\hat{\beta}_i$	Estimated partial slope	14, 15
$\epsilon_i$	Random error term	12-16
$\lambda$	Error reduction, modal prediction	11
$\mu$	Population mean	all
$\mu$	Expected value	4-12
$\mu_i$	Mean of $i$ th population	12
$\mu$	Grand mean	12
$\mu_0$	Null hypothesis mean	9
$\mu_x, \mu_y$	Mean of random variable $X$ or $Y$	4, 5
$\pi$	True (population) proportion or probability for binomial distribution	5, 8-11
$\rho_{xy}$	True correlation of $X$ and $Y$	4, 13, 17
$\sigma$	Population standard deviation	2, 4-12
$\sigma^2$	Population variance	2, 4-12
$\sigma_\epsilon$	True residual standard deviation	13-16
$\sigma_x, \sigma_y$	Standard deviation of random variable $X$ or $Y$	4-6
$\sigma_{\hat{\theta}}$	True standard error of $\hat{\theta}$	7-9
$\theta$	General population parameter	7-10
$\theta$	Moving average parameter	16
$\hat{\theta}$	General estimator	7-10
$\theta_j$	$j$ th possible state of nature	17
$\chi^2$	Chi-square statistic	8, 9, 11, 12
$\chi_a^2$	Table $\chi^2$ value, right-tail area $\alpha$	8, 9, 11, 12





# **STATISTICAL THINKING FOR MANAGERS**





## THE DUXBURY SERIES IN STATISTICS AND DECISION SCIENCES

*Applications, Basics, and Computing of Exploratory Data Analysis*, Velleman and Hoaglin

*Applied Regression Analysis for Business and Economics*, Dielman

*Elementary Statistics, Fifth Edition*, Johnson

*Elementary Statistics for Business, Second Edition*, Johnson and Siskin

*Essential Business Statistics: A Minitab Framework*, Bond and Scott

*Fundamentals of Biostatistics, Third Edition*, Rosner

*Introduction to Mathematical Programming: Applications and Algorithms*, Winston

*Fundamental Statistics for the Behavioral Sciences, Second Edition*, Howell

*Introduction to Probability and Statistics, Eighth Edition*, Mendenhall and Beaver

*Making Hard Decisions: An Introduction to Decision Analysis*, Clemen

*An Introduction to Statistical Methods and Data Analysis, Third Edition*, Ott

*Introductory Business Statistics with Microcomputer Applications*, Shiffler and Adams

*Introductory Statistics and Management and Economics, Third Edition*, Kenkel

*Mathematical Statistics with Applications, Fourth Edition*, Mendenhall, Wackerly, and Scheaffer

*Minitab Handbook, Second Edition*, Ryan, Joiner, and Ryan

*Minitab Handbook for Business and Economics*, Miller

*Operations Research: Applications and Algorithms, Second Edition*, Winston

*Probability and Statistics for Engineers, Third Edition*, Scheaffer and McClave

*Probability and Statistics for Modern Engineering, Second Edition*, Lapin

*SAS Application Programming*, DiIorio

*Statistical Methods for Psychology, Second Edition*, Howell

*Statistical Thinking for Managers, Third Edition*, Hildebrand and Ott

*Statistics: A Tool for the Social Sciences, Fourth Edition*, Ott, Larson, and Mendenhall

*Statistics for Business and Economics*, Bechtold and Johnson

*Statistics for Business: Data Analysis and Modelling*, Cryer and Miller

*Statistics for Management and Economics, Sixth Edition*, Mendenhall, Reinmuth, and Beaver

*Student Edition of EXECUSTAT®*, The Developers of Statgraphics®

*Understanding Statistics, Fifth Edition*, Ott and Mendenhall



## THE DUXBURY ADVANCED SERIES IN STATISTICS AND DECISION SCIENCES

*Applied Nonparametric Statistics, Second Edition*, Daniel

*Applied Regression Analysis and Other Multivariable Methods, Second Edition*, Kleinbaum, Kupper, and Muller

*Classical and Modern Regression with Applications, Second Edition*, Myers

*Elementary Survey Sampling, Fourth Edition*, Scheaffer, Mendenhall, and Ott

*Introduction to Contemporary Statistical Methods, Second Edition*, Koopmans

*Introduction to Probability and Its Applications*, Scheaffer

*Introduction to Probability and Mathematical Statistics*, Bain and Englehardt

*An Introduction to the Theory of the Linear Statistical Model*, Myers and Milton

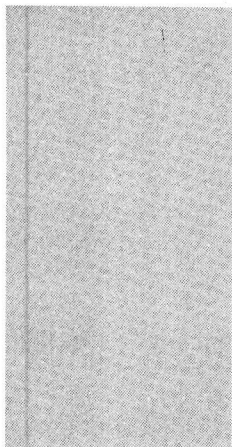
*Linear Statistical Models, Second Edition*, Bowerman and O'Connell

*Probability Modeling and Computer Simulation*, Matloff

*Quantitative Forecasting Methods*, Farnum and Stanton

*Time Series Analysis*, Cryer

*Time Series Forecasting: Unified Concepts and Computer Implementation, Second Edition*, Bowerman and O'Connell



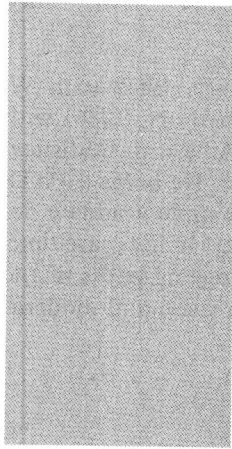
## ABOUT THE AUTHORS

**David K. Hildebrand** is Professor and former (happily for all concerned) Chair of Statistics at the Wharton School, University of Pennsylvania. Since arriving at Wharton in 1965, he has taught all known varieties of Wharton student (full-time undergraduate, part-time undergraduate, full-time MBA, part-time MBA, and Ph.D.). Besides numerous articles in various statistical journals, he has published two other books, *Prediction Analysis of Cross Classifications* (John Wiley and Sons, 1977) with J. D. Laing and H. Rosenthal, and *Statistical Thinking for Behavioral Scientists* (Duxbury, 1985), and a monograph *Analysis of Ordinal Data* (Sage University Papers, 1977), also with J. D. Laing and H. Rosenthal. He has served the American Statistical Association as an associate editor of its journal and by presenting a short course at its 1979 national meetings. He holds a Ph.D. in statistics from Carnegie-Mellon University and a B.A. in mathematics from Carleton College. In his spare time, he fishes for bass (the bass usually win) and enjoys watching major league baseball (and also the Philadelphia Phillies). He should be blamed for most of the changes in this edition.

**Lyman Ott** is Managing Director, Systems and Quality Improvement, at Marion Merrell Dow and an Adjunct Professor in the Division of Biostatistics at the University of Cincinnati. At Merrell Dow Research Institute he was responsible for biostatistics, computer services, clinical data processing, laboratory automation, and scientific information services. Prior to his career in the pharmaceutical industry, he was a faculty member in the Department of Statistics at the University of Florida, where he taught service courses and courses for statistics majors at both the undergraduate and graduate levels. He has published many research articles in various statistical journals and several textbooks, including *An Introduction to Statistical Methods and Data Analysis*, 3rd ed.



(Duxbury, 1988), *Understanding Statistics*, 5th ed. (Duxbury, 1990) with W. Mendenhall, and *Elementary Survey Sampling*, 3rd ed. (Duxbury, 1986) with R. Schaeffer and W. Mendenhall. He is a Fellow of the American Statistical Association and has served on the ASA Board of Directors. He holds a Ph.D. in statistics from the Virginia Polytechnic Institute and an undergraduate degree in mathematics and education from Bucknell University. In his spare time, he enjoys tennis and golf and has explored many of the jogging paths of the North American continent. He has served as the voice of reason and moderation for this revision.



# PREFACE

In the preface to the second edition of this book, we wrote: "...statistics texts must reflect the wide availability of computers. Now that there are literally dozens of statistical packages available ... the emphasis in managerial statistics courses obviously must shift away from computation and toward thoughtful selection of methods and critical interpretation of results."

We stand by those statements, most emphatically. Now, at almost all colleges and universities, students have easy access to statistical software on a mainframe or personal computer to do the important plotting and computation involved in making sense of data. Endless repetition of routine hand calculations may once have been a justifiable focus for a course; now routine computations are a distraction. The interesting questions—how should one gather the data, what methods of analysis should be used, what assumptions must be made and how can they be checked, what do the results mean—should be the primary focus of a managerial statistics course. We have tried to write a text that emphasizes the interesting questions.

We have made several changes and additions to the text to increase its usefulness as a computer-era text.

- **Quality and Productivity.** The role of statistical thinking in quality control and improvement has been made more explicit. Control charts and Pareto diagrams have been introduced in Chapter 2, accompanied by a brief new section discussing management's role in quality. A non-technical discussion of off-line quality improvement by experimentation, including the ideas of orthogonal array and fractional factorial design, has been included in Chapter 12. These changes should increase the value of the text as a reference.

- **Case Studies.** A case study has been included at the end of each chapter. These cases aren't long (they were called "cassettes" and "caselings" at various stages of the project). The presentation of each case asks students to integrate concepts from the chapter and to communicate the results of their analysis in nontechnical language.
- **New Exercises.** We have added numerous new exercises and eliminated some older, less-interesting exercises. A few of the new exercises are conventional pencil-and-paper exercises to fix ideas; the rest are intended to be thought-provoking. Many incorporate computer output of the tedious calculations, allowing students to focus on more important questions and concepts.
- **Computer Exercises.** Many special computer-based exercises have been created. For these exercises, marked by a computer disk icon, the data are stored on a disk available (at no cost to adopters) from the publisher. An instructor can assign these exercises as a means of familiarizing students with whichever computer package is in use. The data sets are formatted as blank-delimited ASCII files without variable-name headers, so they should be readable by almost any package.
- **PC-Based Software.** There is increased emphasis on PC-based computer packages, particularly Minitab, Statgraphics, and Systat. These packages are easier to use than mainframe packages, and their output is usually easier to comprehend.
- **Streamlined Decision Theory.** Formerly separate chapters on basic decision theory and on information in decision theory have been combined, to emphasize the unity of the topics.
- **Influence Concepts in Regression.** High-influence points in regression are explained in Chapters 13 and 14.
- **Earlier Introduction of Bayes' Theorem.** Explicit presentation of Bayes' Theorem has been moved forward to Chapter 4.
- **Solutions Manuals.** A partial solutions manual for students is available, containing extended discussion of the solutions to about 30% of the exercises in the text. The exercises discussed in the solutions manual are numbered with boldface numbers. A complete solutions manual is available to instructors.

Many of the features of the second edition have been retained:

- Review exercises follow sets of several chapters, giving students practice without the artificial clue of exercise placement.
- Several computer simulations of the performance of various statistical methods, under many different conditions, are presented in the text. These simulations provide concrete illustrations of a number of technical concepts. Transparency masters of these and many other simulations are available from the publisher.
- There is heavy emphasis on graphical and exploratory methods, including stem-and-leaf displays, box plots, and normal probability plots. These displays can be produced by any good computer package and

are immensely useful in thoughtful understanding of the data and choice of method.

- There are additional exercises calling for use of calculus methods. These exercises help students to see that calculus methods are natural extensions of algebraic methods. However, calculus isn't a prerequisite to use of the book. Sections and exercises requiring use of differential and integral calculus have been marked by derivative and integral symbols for the convenience of those who wish to skip calculus-based material.
- We have retained a brief chapter describing some of the data management chores necessary before an analysis, as well as guidelines for a statistical analysis and report.

We thank the many individuals who have suggested changes and improvements. We hope that the features we've described are tangible indications that we're listening.

## ACKNOWLEDGMENTS

We didn't do this revision all by ourselves. We're most appreciative of the help of several key people. Patricia Hildebrand, our chief computer guru, reprinted clearer tables, suggested improvements in content, helped with word processing, and tolerated one of the authors with perhaps more grace than was warranted. Michael Payne and Marcia Cole were editors, cheerleaders, and taskmasters in most reasonable proportion.

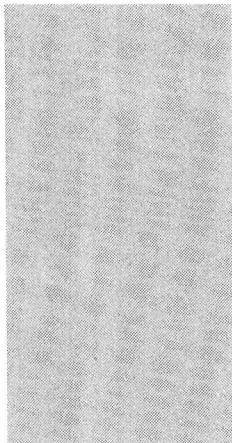
We benefitted greatly from the comments of several thoughtful, and not excessively nasty, reviewers, including

Bruce K. Cooil  
Vanderbilt University  
C. H. Hesse  
University of California, Berkeley  
Burt Holland  
Temple University  
Kevin McCardle  
Duke University  
John W. Mamer  
University of California,  
Los Angeles  
Philip J. Mizzi  
Arizona State University  
Mark Reiser  
Arizona State University

Lawrence A. Sherr  
University of Kansas  
Stanley A. Taylor  
California State University,  
Sacramento  
Patrick A. Thompson  
University of Florida  
Nancy Weida  
Bucknell University  
William B. Widhelm  
University of Maryland  
John W. Wilkinson  
Rensselaer Polytechnic Institute  
R. Worthley  
University of Hawaii

Of course, we have not always agreed with their suggestions; even those ideas that we didn't buy forced us to rethink and clarify. Our sincere thanks to them all.

To Pat, Marty, and Jeff  
Sally, Curtis, and Kathy  
who all helped



# CONTENTS

About the Authors ■ xiv  
Preface ■ xvi

## 1

### MAKING SENSE OF DATA

■ 1

- 1.1 Gathering Data ■ 2
- 1.2 Summarizing Data ■ 3
- 1.3 The Role of Probability ■ 4
- 1.4 Making Inferences from Data ■ 5
- 1.5 The Role of the Computer ■ 5
- SUMMARY ■ 6

## 2

### SUMMARIZING DATA ABOUT ONE VARIABLE

■ 7

- 2.1 The Distribution of Values of a Variable ■ 7
- 2.2 On the Average: Typical Values ■ 16
- 2.3 Measuring Variability ■ 24
- 2.4 Other Summary Measures of Data ■ 41
- 2.5 Calculators and Computer Software Systems ■ 45
- 2.6 Statistical Methods and Quality Improvement ■ 46
- SUMMARY ■ 48
- CASE: Summarizing Data ■ 65
- APPENDIX: Summation Notation ■ 67



---

### 3 A FIRST LOOK AT PROBABILITY ■ 70

---

- 3.1 Interpretations of Probability ■ 70
- 3.2 Basic Concepts and Axioms of Probability Theory ■ 80
- 3.3 Probability Laws ■ 83
- 3.4 Statistical Independence ■ 92
- 3.5 Probability Tables and Probability Trees ■ 96
- 3.6 Bayes' Theorem, Prior and Posterior Probabilities ■ 105
- SUMMARY ■ 111
- CASE: Probability Principles ■ 115
- Review Exercises Chapters 2–3 ■ 117

### 4 RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS ■ 121

---

- 4.1 Random Variable: Basic Ideas ■ 121
- 4.2 Probability Distributions for Discrete Random Variables ■ 124
- 4.3 Probability Distributions for Continuous Random Variables ( $\partial$ ,  $f$ ) ■ 129
- 4.4 Expected Value, Variance, and Standard Deviation:  
Discrete Random Variables ■ 136
- 4.5 Expected Value, Variance, and Standard Deviation:  
Continuous Random Variables ( $f$ ) ■ 141
- 4.6 Joint Probability Distributions and Independence ■ 145
- 4.7 Covariance and Correlation of Random Variables ■ 150
- 4.8 Joint Probability Densities for Continuous Random Variables ( $f$ ) ■ 154
- SUMMARY ■ 160
- CASE: Probability Distributions ■ 166
- APPENDIX 4A: Properties of Expected Values and Variances ■ 167
- APPENDIX 4B: Some Reminders About Calculus ■ 170

### 5 SOME SPECIAL PROBABILITY DISTRIBUTIONS ■ 175

---

- 5.1 Counting Possible Outcomes ■ 175
- 5.2 Bernoulli Trials and the Binomial Distribution ■ 178
- 5.3 Hypergeometric Distribution ■ 184
- 5.4 Geometric and Negative Binomial Distributions ■ 187
- 5.5 Poisson Distribution ■ 190
- 5.6 The Uniform Distribution ■ 193
- 5.7 Exponential Distribution ( $f$ ) ■ 196
- 5.8 Normal Distribution ■ 199

- 5.9 Normal Approximations to Binomial and Poisson Probabilities ■ 204  
SUMMARY ■ 208  
CASE: Special Probability Distributions ■ 216

## 6 RANDOM SAMPLING AND SAMPLING DISTRIBUTIONS ■ 218

---

- 6.1 Random Sampling ■ 219  
6.2 Sample Statistics and Sampling Distributions ■ 222  
6.3 Expected Values and Standard Errors of Sample Sums and Sample Means ■ 227  
6.4 Sampling Distributions for Means and Sums ■ 230  
6.5 Uses and Misuses of the Central Limit Theorem ■ 238  
6.6 Computer Simulations ■ 240  
SUMMARY ■ 250  
CASE: Sampling and Sampling Distributions ■ 255  
APPENDIX: Standard Error of a Mean ■ 255  
Review Exercises Chapters 4–6 ■ 256

## 7 POINT ESTIMATION ■ 260

---

- 7.1 Point Estimators ■ 260  
7.2 Sampling With and Without Replacement ■ 269  
7.3 Maximum Likelihood Estimation ( $\partial$ ) ■ 272  
SUMMARY ■ 277  
CASE: Estimation ■ 284

## 8 INTERVAL ESTIMATION ■ 286

---

- 8.1 Interval Estimation of a Population Mean with Known Standard Deviation ■ 287  
8.2 Confidence Intervals for a Proportion ■ 291  
8.3 How Large a Sample Is Needed? ■ 293  
8.4 The  $t$  Distribution ■ 297  
8.5 Confidence Intervals with the  $t$  Distribution ■ 302  
8.6 Assumptions for Interval Estimation ■ 305  
8.7 Confidence Intervals for a Median ■ 308  
8.8 Confidence Intervals for a Standard Deviation ■ 312  
SUMMARY ■ 318  
CASE: Interval Estimation ■ 332  
APPENDIX: The Chi-Square and  $t$  Distributions ■ 334

---

## 9 HYPOTHESIS TESTING ■ 336

---

- 9.1 A Test for a Binomial Proportion ■ 337
- 9.2 Type II Error,  $\beta$  Probability, and Power of a Test ■ 342
- 9.3 A Test for a Population Mean with Known Standard Deviation ■ 345
- 9.4 The Probability for  $z$  Tests ■ 353
- 9.5 The  $p$ -Value for a Hypothesis Test ■ 358
- 9.6 Hypothesis Testing with the  $t$  Distribution ■ 362
- 9.7 The Effect of Population Nonnormality ■ 367
- 9.8 Tests About a Population Median ■ 368
- 9.9 Testing a Population Proportion Using a Normal Approximation ■ 370
- 9.10 The Relation Between Hypothesis Tests and Confidence Intervals ■ 372
- 9.11 Testing a Standard Deviation ■ 376
- 9.12 Hypothesis Testing as a Decision Method ■ 378
- SUMMARY ■ 380
- CASE: Hypothesis Testing ■ 389
- Review Exercises Chapters 7–9 ■ 389

---

## 10 COMPARING TWO SAMPLES ■ 393

---

- 10.1 Comparing the Means of Two Populations with Known Standard Deviations ■ 393
- 10.2 Comparing the Means of Two Populations with Unknown Standard Deviations ■ 399
- 10.3 A Nonparametric Alternative: The Wilcoxon Rank Sum Test ■ 409
- 10.4 Paired-Sample Methods ■ 416
- 10.5 The Signed-Rank Method ■ 420
- 10.6 Two-Sample Procedures for Proportions ■ 428
- 10.7 Two-Sample Procedures for Standard Deviations ■ 432
- SUMMARY ■ 438
- CASE: Comparing Two Samples ■ 450
- APPENDIX: The Mathematics of Pooled-Variance  $t$  Methods ■ 451

---

## 11 METHODS FOR PROPORTIONS AND QUALITATIVE VARIABLES ■ 453

---

- 11.1 Tests for Several Proportions ■ 453
- 11.2 Chi-Square Tests of Independence ■ 457
- 11.3 Measuring Strength of Relation ■ 464
- SUMMARY ■ 469
- CASE: Testing a Relation ■ 473