STATISTICS

FOR ENGINEERING AND THE SCIENCES

FIFTH EDITION

WILLIAM MENDENHALL TERRY SINCICH

FIFTH EDITION

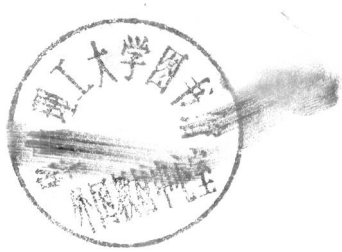
0212
M537
E.5

STATISTICS

for Engineering and the Sciences

William Mendenhall
University of Florida

Terry Sincich
University of South Florida



E200604282

PEARSON
Prentice
Hall

Upper Saddle River, New Jersey

Library of Congress Cataloging-in-Publication Data
Mendenhall, William.
Statistics for engineering and the sciences / William Mendenhall, Terry Sincich. 5th ed.
Includes index.
ISBN 0-13-187706-2
1. Mathematical statistics. 2. Engineering—Statistical methods.
I. Sincich, Terry. II. Title.
QA276.M429 2007
519.5 dc20 2007008927

Editor-in-Chief: *Sally Yagan*
Executive Acquisition Editor: *Petra Recter*
Project Manager: *Michael Bell*
Production Management/Composition: *GGS Book Services*
Assistant Managing Editor: *Bayani Mendoza de Leon*
Senior Managing Editor: *Linda Mihatov Behrens*
Executive Managing Editor: *Kathleen Schiaparelli*
Manufacturing Manager: *Alexis Heydt-Long*
Manufacturing Buyer: *Maura Zaldivar*
Marketing Manager: *Wayne Parkins*
Director of Creative Services: *Paul Belfanti*
Creative Director: *Juan R. López*
Art Director: *Heather Scott*
Art Editor: *Thomas Benfatti*
Interior Designer: *Judith Matz-Coniglio*
Cover Designer: *Tamara Newnam*
Manager, Cover Visual Research & Permissions: *Karen Sanatar*
Permissions Researcher: *Kathy Weisbrod*
Art Studio: *LaserWords*
Editorial Assistant/Supplement Editor: *Joanne Wendelken*
Cover photos: *Pennsylvania, Harrisburg, Pennsylvania, USA—railroad bridge. ©Jeff Lepore/Panoramic Images*

©2007, 1995 by Prentice Hall, Inc.
Pearson Prentice Hall
Pearson Education, Inc.
Upper Saddle River, New Jersey 07458

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

Pearson Prentice Hall™ is a trademark of Pearson Education, Inc.

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

ISBN 0-13-187706-2

Pearson Education LTD., *London*
Pearson Education Australia PTY, Limited, *Sydney*
Pearson Education Singapore, Pte. Ltd
Pearson Education North Asia Ltd, *Hong Kong*
Pearson Education Canada, Ltd., *Toronto*
Pearson Educación de Mexico, S.A. de C.V.
Pearson Education—Japan, *Tokyo*
Pearson Education Malaysia, Pte. Ltd

Preface

Overview

This text is designed for a two-semester introductory course in statistics for students majoring in engineering or any of the physical sciences. Inevitably, once these students graduate and are employed, they will be involved in the collection and analysis of data and will be required to think critically about the results. Consequently, they need to acquire knowledge of the basic concepts of data description and statistical inference and familiarity with statistical methods they are required to use on the job.

Pedagogy

Chapters 1–6 identify the objectives of statistics, explain how we can describe data sets, and present the basic concepts of probability. Chapters 7 and 8 introduce the two methods for making inferences about population parameters: estimation and testing hypotheses. These notions are extended in the remaining chapters to cover other topics that are useful in analyzing engineering and scientific data, including the analysis of categorical data (Chapter 9), regression analysis and model building (Chapters 10–12), the analysis of variance for designed experiments (Chapters 13–14), nonparametric statistics (Chapter 15), statistical quality control (Chapter 16), and product and system reliability (Chapter 17).

The assumed mathematical background is a two-semester sequence in calculus—that is, the course could be taught to students of average mathematical talent and with a basic understanding of the principles of differential and integral calculus. Presentation requires the ability to perform one-variable differentiation and integration, but examples involving topics from multivariable calculus are designated as optional. Thus, the theoretical concepts are sketched and presented in a one-variable context, but it is easy for the instructor to delve deeper into the theoretical and mathematical aspects of statistics using the optional topics, examples, and exercises.

Features

Specific features of the text are as follows:

1. **Blend of theory and applications.** The basic theoretical concepts of mathematical statistics are integrated with a two-semester presentation of statistical methodology. Thus, the instructor has the option of presenting a course with either of two characteristics—a course stressing basic concepts and applied statistics, or a course that, while still tilted toward application, presents a modes introduction to the theory underlying statistical inference.
2. **Statistical software applications with tutorials.** The instructor and student have the option of using statistical software to perform the statistical calculations. Printouts from three popular statistical software packages—SAS, SPSS, and MINITAB—as well as Microsoft Excel output are fully integrated into the text. Tutorials with menu screens and dialog boxes are provided in Appendices C, D, and E. These tutorials are designed for the novice user; no prior experience with the software in needed.

3. **Blended coverage of topics and applications.** To meet the diverse needs of future engineers and scientists, the text provides coverage of a wide range of data analysis topics. The material on multiple regression and model building (Chapters 11–12), principles of experimental design (Chapter 13), quality control (Chapter 16), and reliability (Chapter 17) sets the text apart from the typical introductory statistics text. Although the material often refers to theoretical concepts, the presentation is oriented toward applications.
4. **Numerous real data-based exercises.** The text contains a large number of applied exercises designed to motivate students and suggest future uses of the methodology. Nearly every exercise is based on data or experimental results extracted from professional journals or obtained from organizations in the engineering and physical sciences. Exercises are located at the ends of key sections and at the ends of chapters.
5. **"Statistics in Action" case studies.** The Fifth Edition of the text now includes a contemporary scientific study ("Statistics in Action") and the accompanying data and analysis at the end of each chapter. Our goal is to show students the importance of applying sound statistical techniques in order to evaluate the findings and to think through the statistical issues involved.
6. **Data sets provided online.** All of the data sets associated with examples, exercises, and cases are provided online at www.prenhall.com (search for this textbook). Each data set is marked with a CD icon and file name in the text. The data files are saved in five different formats: MINITAB, SAS, SPSS, Excel, and ASCII (for easy importing into other statistical software packages). By analyzing these data using statistical software, calculations are minimized, allowing students to concentrate on the interpretation of the results.

New to the Fifth Edition

Although the scope and coverage remain the same, the Fifth Edition of the text contains several substantial changes, additions, and enhancements:

1. **"Statistics in Action" cases at the end of each chapter.** As mentioned above, the text now includes a discussion of an actual, recent scientific study at the end of each chapter. The data and analysis are discussed in detail, demonstrating how the statistical methods of the chapter are used to answer relevant questions in the field.
2. **Updated statistical software printouts.** Throughout the text we have greatly increased the number of printouts, and now include the newest versions of SPSS and EXCEL output as well as updated SAS and MINITAB printouts. A printout accompanies every statistical technique presented, allowing the instructor to emphasize interpretations of the statistical results rather than the calculations required to obtain the results.
3. **End-of-chapter summary material.** At the end of each chapter, we provide a summary of the topics presented via a Quick Review (key words and key formulas), Language Lab (a listing of key symbols and pronunciation guide), and Chapter Summary Notes. These features help the student summarize and reinforce the important points from the chapter and are useful study tools.
4. **More exercises with real data** Many new "real-life" scientific exercises have been added throughout the text. All of these are extracted from news articles, scientific magazines, and professional journals.
5. **Chapter 1: Statistics and critical thinking.** A new section on the role of statistics in the assessment of the credibility and value of the inferences made from data—i.e., critical thinking (Section 1.4)—has been added.

此为试读,需要完整PDF请访问: www.ertongbook.com

6. **Chapter 2: Distorting the truth with descriptive statistics.** Several examples on how graphs and numerical descriptive measures can be used to give a distorted view of the data, and how to recognize when this occurs, are now included in a new section (Section 2.8).
7. **Chapters 4–6: Standard mathematical notation for a random variable.** Throughout the chapters on random variables, we now use the standard mathematical notation for representing a random variable. Uppercase letters are used to represent the random variable, and lowcase letters represent the values that the random variable can assume.
8. **Chapter 6: Monte Carlo simulation.** We emphasize the use of Monte Carlo simulation for approximating a sampling distribution through examples in Section 6.8.
9. **Chapters 7 and 8: Bootstrapping and Bayesian methods.** New optional sections now present two alternative estimation methods (Section 7.12) and hypothesis-testing methods (Section 8.13)—bootstrapping and Bayesian methods.
10. **Chapter 9: Exact tests for a contingency table.** A new optional section (Section 9.6) presents Fisher's exact test for independence in a 2×2 contingency table.
11. **Chapter 11: Multiple regression models.** Examples of several different multiple regression models are given in separate sections in Chapter 11–a first-order model (Section 11.7), an interaction model (Section 11.8), and a second-order model (Section 11.9).
12. **Chapter 12: External model validation.** A new optional section (Section 12.9) on validating a multiple regression model externally has been added to the model-building chapter, Chapter 12. Cross-validation and jackknifing methods are discussed.
13. **Chapters 13 and 14: Experimental design and ANOVA.** The previous material on analysis of variance is now presented in two chapters: "Principles of Experimental Design" (Chapter 13) and "ANOVA for Designed Experiments" (Chapter 14).
14. **Chapter 16: Capability analysis.** The chapter on statistical process and quality control now includes a new optional section (Section 16.9) on the importance of a capability analysis.

Numerous, less obvious changes in details have been made throughout the text in response to suggestions by current users and reviewers of the text.

Supplements

The text is also accompanied by the following supplementary material:

1. **Student's Solutions Manual** (by Mark Dummeldinger) (ISBN: 0-13-187708-9). The manual contains the full solutions for all the odd-numbered exercises contained in the text.
2. **Instructor's Solutions Manual** (by Mark Dummeldinger) (ISBN: 0-13-187707-0). The manual contains the full solutions to all the even-numbered exercises contained in the text.

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 Fifth Edition:

Reviewers Involved with the Fifth Edition of *Statistics for Engineering and the Sciences*:

Dharam Chopra (Wichita University)
Edward Danial (Morgan State University)
Frank Guess (University of Tennessee)
Carol O'Connor Holloman (University of Louisville)
Nancy Matthews (University of Oklahoma)
Alan Rabideau (University of Bufffalo)
Charles Reilly (University of Central Florida)
Larry Ringer (Texas A&M University)
Arnold Sweet (Purdue University)
Paul Switzer (Stanford University)

Reviewers of Previous Editions

Carl Bodenschatz (United States Air Force Academy)
George C. Derringer (Battelle Columbus, Ohio, Division)
Danny Dyer (University of Texas-Arlington)
Herberg Eisenberg (West Virginia College of Graduate Studies)
Christopher Ennis (Normandale Community College)
Nasrollah Etemadi (University of Illinois-Chicago)
Linda Gans (California State Polytechnic University)
Carol Gattis (University of Arkansas)
K. G. Janardan (Eastern Michigan University)
H. Lennon (Coventry Polytechnic, Coventry, England)
Jeffery Maxey (University of Central Florida)
Curtis McKnight (University of Oklahoma)
Chand Midha (University of Akron)
Balgobin Nandram (Worcester Polytechnic Institute)
Paul Nelson (Kansas State University)
Norbert Oppenheim (City College of New York)
Giovanni Parmigiani (Duke University)
David Powers (Clarkson University)
David Robinson (St. Cloud State University)
Shiva Saksena (University of North Carolina-Wilmington)
Dennis Wackerly (University of Florida)
Donald Woods (Texas A&M Universtity)

Other Contributors

Special thanks are due to our ancillary author, Mark Dummeldinger. Dr. Ed Danial and Weiwei Chen have done an excellent job of accuracy checking the Fifth Edition and have helped us to ensure a highly accurate, clean text. Finally, the Prentice Hall staff of Sally Yagan, Jennifer Urban, and Joanne Wendelken helped greatly with all phases of the text development and production.

Contents

Preface ix

Chapter 1 Introduction 1

1.1 Statistics: The Science of Data 2
1.2 Fundamental Elements of Statistics 2
1.3 Types of Data 5
1.4 The Role of Statistics in Critical Thinking 7
1.5 A Guide to Statistical Methods Presented in This Text 9
STATISTICS IN ACTION Contamination of Fish in the Tennessee River: Collecting the Data 10

Chapter 2 Descriptive Statistics 12

2.1 Graphical and Numerical Methods for Describing Qualitative Data 13
2.2 Graphical Methods for Describing Quantitative Data 19
2.3 Numerical Methods for Describing Quantitative Data 27
2.4 Measures of Central Tendency 28
2.5 Measures of Variation 32
2.6 Measures of Relative Standing 38
2.7 Methods for Detecting Outliers 41
2.8 Distorting the Truth with Descriptive Statistics 45
STATISTICS IN ACTION Characteristics of Contaminated Fish in the Tennessee River, Alabama 56

Chapter 3 Probability 59

3.1 The Role of Probability in Statistics 60
3.2 Events, Sample Spaces, and Probability 60
3.3 Compound Events 70
3.4 Complementary Events 72
3.5 Conditional Probability 76
3.6 Probability Rules for Unions and Intersections 80
3.7 Bayes' Rule (*Optional*) 90
3.8 Some Counting Rules 93
3.9 Probability and Statistics: An Example 103
3.10 Random Sampling 105
STATISTICS IN ACTION Assessing Predictors of Software Defects in NASA Spacecraft Instrument Code 114

Chapter 4 Discrete Random Variables 117

4.1 Discrete Random Variables 118
4.2 The Probability Distribution for a Discrete Random Variable 118
4.3 Expected Values for Random Variables 123
4.4 Some Useful Expectation Theorems 127

4.5 Bernoulli Trials 129
4.6 The Binomial Probability Distribution 130
4.7 The Multinomial Probability Distribution 137
4.8 The Negative Binomial and the Geometric Probability Distributions 142
4.9 The Hypergeometric Probability Distribution 146
4.10 The Poisson Probability Distribution 151
4.11 Moments and Moment Generating Functions (*Optional*) 157
STATISTICS IN ACTION The Reliability of a "One-Shot" Device 165

Chapter 5 Continuous Random Variables 168

5.1 Continuous Random Variables 169
5.2 The Density Function for a Continuous Random Variable 170
5.3 Expected Values for Continuous Random Variables 172
5.4 The Uniform Probability Distribution 177
5.5 The Normal Probability Distribution 180
5.6 Descriptive Methods for Assessing Normality 184
5.7 Gamma-Type Probability Distributions 190
5.8 The Weibull Probability Distribution 194
5.9 Beta-Type Probability Distributions 197
5.10 Moments and Moment Generating Functions (*Optional*) 200
STATISTICS IN ACTION Super Weapons Development—Optimizing the Hit Ratio 206

Chapter 6 Bivariate Probability Distributions and Sampling Distributions 211

6.1 Bivariate Probability Distributions for Discrete Random Variables 212
6.2 Bivariate Probability Distributions for Continuous Random Variables 217
6.3 The Expected Value of Functions of Two Random Variables 221
6.4 Independence 223
6.5 The Covariance and Correlation of Two Random Variables 225
6.6 Probability Distributions and Expected Values of Functions of Random Variables (*Optional*) 228
6.7 Sampling Distributions 236
6.8 Approximating a Sampling Distribution by Monte Carlo Simulation 237
6.9 The Sampling Distributions of Means and Sums 240
6.10 Normal Approximation to the Binomial Distribution 245
6.11 Sampling Distributions Related to the Normal Distribution 248
STATISTICS IN ACTION Availability of an Up/Down Maintained System 259

Chapter 7 Estimation Using Confidence Intervals 262

7.1 Point Estimators and their Properties 263
7.2 Finding Point Estimators: Classical Methods of Estimation 267
7.3 Finding Interval Estimators: The Pivotal Method 274
7.4 Estimation of a Population Mean 281
7.5 Estimation of the Difference Between Two Population Means: Independent Samples 287

7.6 Estimation of the Difference Between Two Population Means: Matched Pairs 294
7.7 Estimation of a Population Proportion 299
7.8 Estimation of the Difference Between Two Population Proportions 302
7.9 Estimation of a Population Variance 305
7.10 Estimation of the Ratio of Two Population Variances 309
7.11 Choosing the Sample Size 315
7.12 Alternative Interval Estimation Methods: Bootstrapping and Bayesian Methods (*Optional*) 318
STATISTICS IN ACTION Bursting Strength of PET Beverage Bottles 332

Chapter 8 Tests of Hypotheses 335

8.1 The Relationship Between Statistical Tests of Hypotheses and Confidence Intervals 336
8.2 Elements and Properties of a Statistical Test 336
8.3 Finding Statistical Tests: Classical Methods 342
8.4 Choosing the Null and Alternative Hypotheses 346
8.5 Testing a Population Mean 348
8.6 The Observed Significance Level for a Test 354
8.7 Testing the Difference Between Two Population Means: Independent Samples 357
8.8 Testing the Difference Between Two Population Means: Matched Pairs 364
8.9 Testing a Population Proportion 369
8.10 Testing the Difference Between Two Population Proportions 372
8.11 Testing a Population Variance 376
8.12 Testing the Ratio of Two Population Variances 379
8.13 Alternative Testing Procedures: Bootstrapping and Bayesian Methods (*Optional*) 383
STATISTICS IN ACTION Comparing Methods for Dissolving Drug Tablets—Dissolution Method Equivalence Testing 395

Chapter 9 Categorical Data Analysis 400

9.1 Categorical Data and Multinomial Probabilities 401
9.2 Estimating Category Probabilities in a One-Way Table 401
9.3 Testing Category Probabilities in a One-Way Table 405
9.4 Inferences About Category Probabilities in a Two-Way (Contingency) Table 409
9.5 Contingency Tables with Fixed Marginal Totals 417
9.6 Exact Tests for Independence in a Contingency Table Analysis (*Optional*) 422
STATISTICS IN ACTION The Public's Perception of Engineers and Engineering 432

Chapter 10 Simple Linear Regression 439

10.1 Regression Models 440
10.2 Model Assumptions 441
10.3 Estimating β_0 and β_1: The Method of Least Squares 443
10.4 Properties of the Least Squares Estimators 454

10.5 An Estimator of σ^2 456
10.6 Assessing the Utility of the Model: Making Inferences About the Slope β_1 460
10.7 The Coefficient of Correlation 465
10.8 The Coefficient of Determination 469
10.9 Using the Model for Estimation and Prediction 473
10.10 A Complete Example 480
10.11 A Summary of the Steps to Follow in Simple Linear Regression 483
STATISTICS IN ACTION Can Dowsers Really Detect Water? 491

Chapter 11 Multiple Regression Analysis 495

11.1 General Form of a Multiple Regression Model 496
11.2 Model Assumptions 497
11.3 Fitting the Model: The Method of Least Squares 498
11.4 Computations Using Matrix Algebra: Estimating and Making Inferences About the Individual β Parameters 499
11.5 Assessing Overall Model Adequacy 506
11.6 A Confidence Interval for $E(y)$ and a Prediction Interval for a Future Value of y 510
11.7 A First-Order Model with Quantitative Predictors 519
11.8 An Interaction Model with Quantitative Predictors 528
11.9 A Quadratic (Second-Order) Model with a Quantitative Predictor 533
11.10 Checking Assumptions: Residual Analysis 540
11.11 Some Pitfalls: Estimability, Multicollinearity, and Extrapolation 559
11.12 A Summary of the Steps to Follow in a Multiple Regression Analysis 568
STATISTICS IN ACTION Bid-Rigging in the Highway Construction Industry 575

Chapter 12 Model Building 583

12.1 Introduction: Why Model Building Is Important 584
12.2 The Two Types of Independent Variables: Quantitative and Qualitative 585
12.3 Models with a Single Quantitative Independent Variable 586
12.4 Models with Two Quantitative Independent Variables 593
12.5 Coding Quantitative Independent Variables (*Optional*) 600
12.6 Models with One Qualitative Independent Variable 605
12.7 Models with Both Quantitative and Qualitative Independent Variables 611
12.8 Tests for Comparing Nested Models 621
12.9 External Model Validation (*Optional*) 628
12.10 Stepwise Regression 630
STATISTICS IN ACTION Deregulation of the Intrastate Trucking Industry 640

Chapter 13 Principles of Experimental Design 648

13.1 Introduction 649
13.2 Experimental Design Terminology 649
13.3 Controlling the Information in an Experiment 651
13.4 Noise-Reducing Designs 652
13.5 Volume-Increasing Designs 658
13.6 Selecting the Sample Size 663
13.7 The Importance of Randomization 665
STATISTICS IN ACTION Anticorrosive Behavior of Epoxy Coatings Augmented with Zinc 668

此为试读,需要完整PDF请访问: www.ertongbook.com

Chapter 14 The Analysis of Variance for Designed Experiments 671

14.1 Introduction 672
14.2 The Logic Behind an Analysis of Variance 672
14.3 One-Factor Completely Randomized Designs 674
14.4 Randomized Block Designs 685
14.5 Two-Factor Factorial Experiments 698
14.6 More Complex Factorial Designs (*Optional*) 714
14.7 Nested Sampling Designs (*Optional*) 721
14.8 Multiple Comparisons of Treatment Means 732
14.9 Checking ANOVA Assumptions 739
STATISTICS IN ACTION On the Trail of the Cockroach 751

Chapter 15 Nonparametric Statistics 755

15.1 Introduction: Distribution-Free Tests 756
15.2 Testing for Location of a Single Population 757
15.3 Comparing Two Populations: Independent Random Samples 762
15.4 Comparing Two Populations: Matched-Pairs Design 769
15.5 Comparing Three or More Populations: Completely Randomized Design 775
15.6 Comparing Three or More Populations: Randomized Block Design 780
15.7 Nonparametric Regression 784
STATISTICS IN ACTION Deadly Exposure: Agent Orange and Vietnam Vets 796

Chapter 16 Statistical Process and Quality Control 800

16.1 Total Quality Management 801
16.2 Variable Control Charts 801
16.3 Control Chart for Means: $\bar{x}$-Chart 806
16.4 Control Chart for Process Variation: R-Chart 814
16.5 Detecting Trends in a Control Chart: Runs Analysis 819
16.6 Control Chart for Percent Defectives: p-Chart 821
16.7 Control Chart for the Number of Defectives per Item: c-Chart 825
16.8 Tolerance Limits 829
16.9 Capability Analysis (*Optional*) 832
16.10 Acceptance Sampling for Defectives 839
16.11 Other Sampling Plans (*Optional*) 844
16.12 Evolutionary Operations (*Optional*) 845
STATISTICS IN ACTION Testing Jet Fuel Additive for Safety 851

Chapter 17 Product and System Reliability 857

17.1 Introduction 858
17.2 Failure Time Distributions 858
17.3 Hazard Rates 859
17.4 Life Testing: Censored Sampling 863
17.5 Estimating the Parameters of an Exponential Failure Time Distribution 864
17.6 Estimating the Parameters of a Weibull Failure Time Distribution 867

17.7 System Reliability 872
STATISTICS IN ACTION Modeling the Hazard Rate of Reinforced Concrete Bridge Deck Deterioration 879

Appendix A Matrix Algebra 882

A.1 Matrices and Matrix Multiplication 882
A.2 Identity Matrices and Matrix Inversion 886
A.3 Solving Systems of Simultaneous Linear Equations 889
A.4 A Procedure for Inverting a Matrix 891

Appendix B Useful Statistical Tables 896

Table 1 Random Numbers 897
Table 2 Cumulative Binomial Probabilities 901
Table 3 Exponentials 905
Table 4 Cumulative Poisson Probabilities 906
Table 5 Normal Curve Areas 908
Table 6 Gamma Function 909
Table 7 Critical Values for Student's T 910
Table 8 Critical Values of χ^2 911
Table 9 Percentage Points of the F Distribution, $\alpha = .10$ 913
Table 10 Percentage Points of the F Distribution, $\alpha = .05$ 915
Table 11 Percentage Points of the F Distribution, $\alpha = .025$ 917
Table 12 Percentage Points of the F Distribution, $\alpha = .01$ 919
Table 13 Percentage Points of the Studentized Range $q(p, \nu)$, $\alpha = .05$ 921
Table 14 Percentage Points of the Studentized Range $q(p, \nu)$, $\alpha = .01$ 923
Table 15 Critical Values of T_L and T_U for the Wilcoxon Rank Sum Test: Independent Samples 925
Table 16 Critical Values of T_0 for the Wilcoxon Matched-Pairs Signed Rank Test 926
Table 17 Critical Values of Spearman's Rank Correlation Coefficient 927
Table 18 Critical Values of C for the Theil Zero-Slope Test 928
Table 19 Factors Used When Constructing Control Charts 932
Table 20 Values of K for Tolerance Limits for Normal Distributions 933
Table 21 Sample Size n for Nonparametric Tolerance Limits 934
Table 22 Sample Size Code Letters: MIL-STD-105D 934
Table 23 A Portion of the Master Table for Normal Inspection (Single Sampling): MIL-STD-105D 935

Appendix C SAS for Windows Tutorial 936

Appendix D MINITAB for Windows Tutorial 965

Appendix E SPSS for Windows Tutorial 998

References 1028

Selected Short Answers 1036

Credits 1049

Index 1054

CHAPTER 1

Introduction

OBJECTIVE

To identify the role of statistics in the analysis of data from engineering and the sciences

CONTENTS

1.1 Statistics: The Science of Data

1.2 Fundamental Elements of Statistics

1.3 Types of Data

1.4 The Role of Statistics in Critical Thinking

1.5 A Guide to Statistical Methods Presented in This Text

- *STATISTICS IN ACTION*
  Contamination of Fish in the Tennessee River: Collecting the Data

1.1 Statistics: The Science of Data

According to *The Random House College Dictionary*, statistics is "the science that deals with the collection, classification, analysis, and interpretation of numerical facts or data." In short, **statistics** is the **science of data.**

Definition 1.1

Statistics is the science of data. This involves collecting, classifying, summarizing, organizing, analyzing, and interpreting data.

The science of statistics is commonly applied to two types of problems:

1. Summarizing, describing, and exploring data
2. Using sample data to infer the nature of the data set from which the sample was selected

As an illustration of the descriptive applications of statistics, consider the United States census, which involves the collection of a data set that purports to characterize the socioeconomic characteristics of the approximately 295 million people living in the United States. Managing this enormous mass of data is a problem for the computer software engineer, and describing the data utilizes the methods of statistics. Similarly, an environmental engineer uses statistics to describe the data set consisting of the daily emissions of sulfur oxides of an industrial plant recorded for 365 days last year. The branch of statistics devoted to these applications is called **descriptive statistics.**

Definition 1.2

The branch of statistics devoted to the organization, summarization, and description of data sets is called **descriptive statistics.**

Sometimes the phenomenon of interest is characterized by a data set that is either physically unobtainable or too costly or time-consuming to obtain. In such situations, we obtain a subset of the data—called a *sample*—and use the sample information to infer its nature. To illustrate, suppose the phenomenon of interest is the drinking-water quality on an inhabited, but remote, Pacific island. You might expect water quality to depend on such factors as temperature of the water, the level of the most recent rainfall, etc. In fact, if you were to measure the water quality repeatedly within the same hour at the same location, the quality measurements would vary, even for the same water temperature. Thus, the phenomenon "drinking-water quality" is characterized by a large data set that consists of many (actually, an infinite number of) water quality measurements—a data set that exists only conceptually. To determine the nature of this data set, we *sample* it—i.e., we record quality for n water specimens collected at specified times and locations, and then use this sample of n quality measurements to infer the nature of the large conceptual data set of interest. The branch of statistics used to solve this problem is called **inferential statistics**.

Definition 1.3

The branch of statistics concerned with using sample data to make an inference about a large set of data is called **inferential statistics.**

1.2 Fundamental Elements of Statistics

In statistical terminology, the data set that we want to describe, the one that characterizes a phenomenon of interest to us, is called a **population.** Then, we can define a **sample** as a subset of data selected from a population. Sometimes, the words *population* and *sample* are used to represent the objects upon which the measurements

are taken (i.e., the *experimental units*). In a particular study, the meaning attached to these terms will be clear by the context in which they are used.

Definition 1.4

A statistical **population** is a data set (usually large, sometimes conceptual) that is our target of interest.

Definition 1.5

A **sample** is a subset of data selected from the target population.

Definition 1.6

The object (e.g., person, thing, transaction, specimen, or event) upon which measurements are collected is called the **experimental unit.** (*Note:* A population consists of data collected on many experimental units.)

In studying populations and samples, we focus on one or more characteristics or properties of the experimental units in the population. The science of statistics refers to these characteristics as **variables.** For example, in the drinking-water quality study, two variables of interest to engineers are the chlorine-residual (measured in parts per million) and the number of fecal coliforms in a 100-milliliter water specimen.

Definition 1.7

A **variable** is a characteristic or property of an individual experimental unit.

Example 1.1

Engineers with the University of Kentucky Transportation Research Program have collected data on accidents occurring at intersections in Lexington, Kentucky. One of the goals of the study was to estimate the rate at which left-turn accidents occur at intersections without left-turn-only lanes. This estimate will be used to develop numerical warrants (or guidelines) for the installation of left-turn lanes at all major Lexington intersections. The engineers collected data at each of 50 intersections without left-turn-only lanes over a 1-year period. At each intersection, they monitored traffic and recorded the total number of cars turning left that were involved in an accident.

a. Identify the variable and experimental unit for this study.
b. Describe the target population and the sample.
c. What inference do the transportation engineers want to make?

Solution

a. Since the engineers collected data at each of 50 intersections, the experimental unit is an intersection without a left-turn-only lane. The variable measured is the total number of cars turning left that were involved in an accident.
b. The goal of the study is to develop guidelines for the installation of left-turn lanes at all major Lexington intersections; consequently, the target population consists of all major intersections in the city. The sample consists of the subset of 25 intersections monitored by the engineers.
c. The engineers will use the sample data to estimate the rate at which left-turn accidents occur at all major Lexington intersections. (We learn, in Chapter 7, that this estimate is the number of left-turn accidents in the sample divided by the total number of cars making left turns in the sample.)

The preceding definitions and example identify four of the five elements of an inferential statistical problem: a population, one or more variables of interest, a sample, and an inference. The fifth element pertains to knowing how good the inference is—that is, the **reliability** of the inference. The measure of reliability that accompanies an inference separates the science of statistics from the art of fortune-telling. A palm

reader, like a statistician, may examine a sample (your hand) and make inferences about the population (your future life). However, unlike statistical inferences, the palm reader's inferences include no measure of how likely the inference is to be true.

To illustrate, consider the transportation engineers' estimate of the left-turn accident rate at Lexington, Kentucky, intersections in Example 1.1. The engineers are interested in the *error of estimation* (i.e., the difference between the sample accident rate and the accident rate for the target population). Using statistical methods, we can determine a *bound on the estimation error*. This bound is simply a number (e.g., 10%) that our estimation error is not likely to exceed. In later chapters, we learn that this bound is used to help measure our "confidence" in the inference. The reliability of statistical inferences is discussed throughout this text. For now, simply realize that an inference is incomplete without a measure of reliability.

Definition 1.8

A **measure of reliability** is a statement (usually quantified) about the degree of uncertainty associated with a statistical inference.

A summary of the elements of both descriptive and inferential statistical problems is given in the following boxes.

Four Elements of Descriptive Statistical Problems

1. The population or sample of interest
2. One or more variables (characteristics of the population or sample units) that are to be investigated
3. Tables, graphs, or numerical summary tools
4. Identification of patterns in the data

Five Elements of Inferential Statistical Problems

1. The population of interest
2. One or more variables (characteristics of the experimental units) that are to be investigated
3. The sample of experimental units
4. The inference about the population based on information contained in the sample
5. A measure of reliability for the inference

Applied Exercises

1.1 *Steel anticorrosion study.* Researchers at the Department of Materials Science and Engineering, National Technical University (Athens, Greece), examined the anticorrosive behavior of different epoxy coatings on steel. (*Pigment & Resin Technology*, Vol. 32, 2003.) Flat panels cut from steel sheets were coated with one of four different types of epoxy (S1, S2, S3, and S4). After exposing the panels to water for one day, the corrosion rate (nanoamperes per square centimeter) was determined for each panel.

a. What are the experimental units for the study?
b. Suppose you are interested in describing only the corrosion rates of steel panels coated with epoxy type S1. Define the target population and relevant sample.

SWREUSE

1.2 *Success/failure of software reuse.* The PROMISE Software Engineering Repository, hosted by the University of Ottawa, is a collection of publicly available data sets to serve researchers in building prediction software models. A PROMISE data set on software reuse, saved in the **SWREUSE** file, provides information on the success or failure of reusing previously developed software for each project in a sample of 24 new software development projects. (Data source: *IEEE Transactions on Software Engineering*, Vol. 28, 2002.) Of the 24 projects, 9 were judged failures and 15 were successfully implemented.

a. Identify the experimental units for this study.
b. Describe the population from which the sample is selected.
c. Use the sample information to make an inference about the population.

1.3 *Orchard contamination from insecticides.* Pesticides applied to an extensively grown crop can result in inadvertent

此为试读,需要完整PDF请访问: www.ertongbook.com