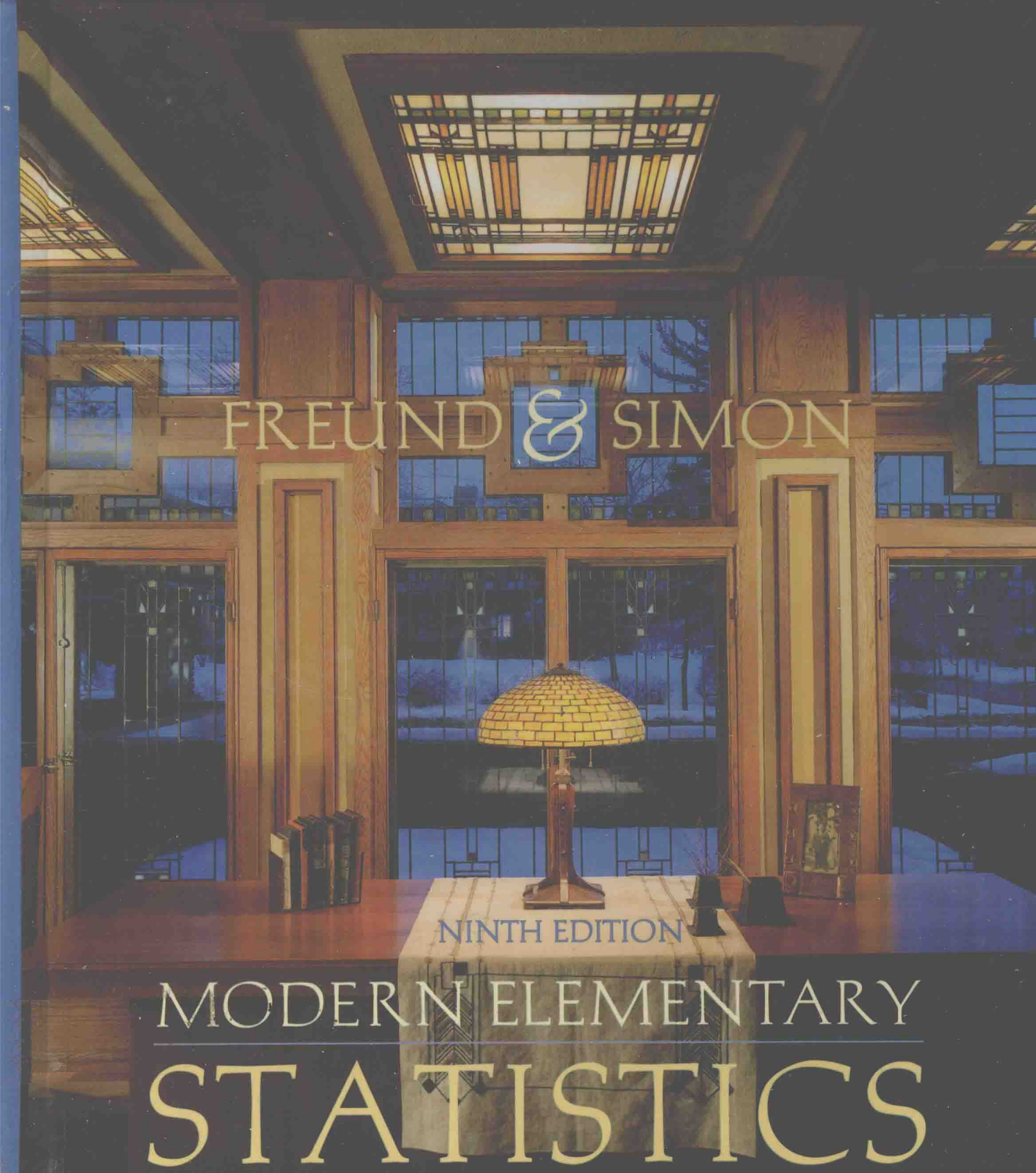

FREUND & SIMON
NINTH EDITION
MODERN ELEMENTARY
STATISTICS

NINTH EDITION
REVISED

MODERN ELEMENTARY STATISTICS

JOHN E. FREUND
Arizona State University

GARY A. SIMON
New York University

PRENTICE HALL
Upper Saddle River, New Jersey 07458

Library of Congress Cataloging-in-Publication Data

Freund, John E.
Modern elementary statistics / John E. Freund, Gary A. Simon — rev. 9th ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-13-858291-2 (hard)
1. Statistics. I. Simon, Gary A. II. Title.
QA276. 12.F738 1997
519.5—dc20 96-22742
CIP

Editorial director: TIM BOZIK
Editor-in-chief: JEROME GRANT
Acquisition editor: ANN HEATH
Development editor: MILLICENT TRELOAR
Assistant VP production and manufacturing: DAVID RICCARDI
Executive managing editor: KATHLEEN SCHIAPARELLI
Managing editor: LINDA MIHATOV BEHRENS
Editorial assistants: NANCY BAUER / MINDY INCE
Creative director: PAULA MAYLAHN
Art director: AMY ROSEN
Interior designer: KEVIN KALL
Cover designer: JOSEPH SENGOTTA
Assistant to art director: ROD HERNANDEZ
Art manager: GUS VIBAL
Marketing manager: EVAN GIRARD
Marketing assistant: BILL PAQUIN
Manufacturing manager: TRUDY PISCIOTTI
Manufacturing buyer: ALAN FISCHER

Cover photo: Frank Lloyd Wright's May House in Grand Rapids Michigan (1908)

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

ISBN 0-13-858291-2

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pte. Limited, *Sydney*
Prentice-Hall Canada, Inc., *Toronto*
Prentice-Hall Hispanoamericana, S.A., *Mexico*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Simon & Schuster Asia Pte. Ltd., *Singapore*
Editora Prentice-Hall do Brasil Ltda., *Rio de Janeiro*

PREFACE

The principal objective for this, the ninth edition of *Modern Elementary Statistics,* is to introduce beginning students to the fundamental ideas of modern statistics. The organization of the text is unchanged, but there are modifications in content, emphasis, and visual presentation.

This ninth edition contains some 1400 exercises that involve both conceptual and computational challenges. Our purpose here is to make the reader familiar with the mechanics of calculations but also to think about the relevance of techniques introduced.

This edition offers greater opportunity for the student to use the computer in the study of statistics. The computer is spectacularly effective for producing graphic displays such as scatter plots and box-and-whisker plots, and can be used to great advantage to simulate values of random variables in connection with the law of large numbers or the sampling distribution of the mean; to determine tail probabilities for the values of n and p not found in tables and thus ascertain the proximity of the Poisson approximation to the binomial. Moreover, the computer is indispensable for multiple-regression computations that are intricate and not amenable to manual calculation.

We do not state specifically the subjects, sections, or chapters to study, but we do indicate which topics can be omitted without loss of continuity. Such "optional" sections are identified by an icon, ⏢, preceding the topic heading. For example, some instructors will not cover the material on grouped data, yet those who must deal with published government data may consider it essential.

NEW TO THIS EDITION

Supplementing the innovations highlighted as follows, this edition is enhanced by an "open" style of presentation, meaning that important equations or statements and lists of data are displayed rather than embedded in the text.

topic selection

The selection of topics for this text has evolved during its eight preceding editions, and the experience and classroom use implied by the wide, implicit acceptance of its presentation. Additionally, for this edition we have included, in Chapter 14, material concerning multiple comparisons. As a logical follow-up to comparative experiments, frequently involving consumer goods, decisions concerning the relative merits of individual products must be made; this section provides guidance for that process.

computer printouts

Illustrated copiously with computer output, this edition uses MINITAB 10 for Windows, for which most of the commands are executed by using menus. As the menu item is selected, this program creates automatically the command code familiar to previous releases of MINITAB. In this edition the command code illustrates the applications. Instructors should have no difficulty in using this text with other software packages.

reduced complexity of exercises

In the early chapters the student will make computations of the form Σx, Σx^2, and $\Sigma (x - \bar{x})^2$ during the process of learning about the mean and the standard deviation. For the more advanced topics, later in the book, the exercises simply present the student with values of $\bar{x}$ and s, rather than requiring that these be computed from raw data. Thus, students can devote their concentration to the new subject and avoid calculations related to topics studied previously.

icons to classify exercises

Three icons are used for the reader's convenience: The icon  precedes exercises requiring computer software. The more challenging exercises are denoted by the symbol . Finally, the icon  is used to flag exercises related to the material covered in "optional" sections. Occasionally, more than one icon may apply to a given exercise.

data disk

Exercise data sets are now provided on disk. These exercises pertain to topics requiring extensive data entries.

acknowledgments

Our appreciation goes to our colleagues and students for their helpful suggestions and reviews of the present as well as preceding editions of this text. For this revision we are grateful to Leslie Ribera of the University of Central Oklahoma, Dee Oyler of Utah Valley State College, Louis Bianco of the University of Massachusetts at Dartmouth, David Lund of the University of Wisconsin at Eau Claire, and Bernard Okun of Brooklyn College for their helpful comments. We also thank Zhenhong Fan for his assistance in preparing the exercise solutions. Finally, our gratitude to our editors at Prentice Hall, Ann Heath and Millicent Treloar, and to Nicholas Romanelli for his help during the production process.

The revisions in the original version of the 9th edition of this book are entirely the work of Gary Simon.

GARY SIMON

CONTENTS

PREFACE x

1 INTRODUCTION 1

1.1 The Growth of Modern Statistics 2
1.2 The Study of Statistics 3
1.3 Descriptive Statistics and Statistical Inference 5
♢ 1.4 The Nature of Statistical Data 7
1.5 Checklist of Key Terms 11
1.6 Review Exercises 11
1.7 References 12

2 SUMMARIZING DATA: FREQUENCY DISTRIBUTIONS 14

2.1 Displaying Numeric Values 15
2.2 Frequency Distributions 20
2.3 Graphical Presentations 31
2.4 Checklist of Key Terms 39
2.5 Review Exercises 39
2.6 References 41

3 SUMMARIZING DATA: MEASURES OF LOCATION 43

3.1 Populations and Samples 44
3.2 The Mean 45
3.3 The Weighted Mean 50
3.4 The Median 55
♢ 3.5 Other Fractiles 59
3.6 The Mode 62
♢ 3.7 The Description of Grouped Data 67

3.8 Technical Note (Summations) 73
3.9 Checklist of Key Terms 76
3.10 Review Exercises 77
3.11 References 79

4 SUMMARIZING DATA: MEASURES OF VARIATION 80

4.1 The Range 81
4.2 The Variance and the Standard Deviation 82
4.3 Applications of the Standard Deviation 86
♢ 4.4 The Description of Grouped Data 95
♢ 4.5 Some Further Descriptions 96
4.6 Checklist of Key Terms 100
4.7 Review Exercises 101
4.8 References 103

5 POSSIBILITIES AND PROBABILITIES 104

5.1 Counting 105
5.2 Permutations 109
5.3 Combinations 112
5.4 Probability 121
5.5 Checklist of Key Terms 128
5.6 Review Exercises 129
5.7 References 131

6 SOME RULES OF PROBABILITY 132

6.1 Sample Spaces and Events 133
6.2 The Postulates of Probability 143
6.3 Probabilities and Odds 146
6.4 Addition Rules 153
6.5 Conditional Probability 161
6.6 Multiplication Rules 165
♢ 6.7 Bayes' Theorem 172
6.8 Checklist of Key Terms 178
6.9 Review Exercises 179
6.10 References 184

7 EXPECTATIONS AND DECISIONS 185

7.1 Mathematical Expectation 186
♢ 7.2 Decision Making 192
♢ 7.3 Statistical Decision Problems 194
7.4 Checklist of Key Terms 199
7.5 Review Exercises 200
7.6 References 202

8 PROBABILITY DISTRIBUTIONS 203

8.1 Random Variables 204
8.2 Probability Distributions 205
8.3 The Binomial Distribution 208
8.4 The Hypergeometric Distribution 218
8.5 The Poisson Distribution 223
♢ 8.6 The Multinomial Distribution 228
8.7 The Mean of a Probability Distribution 231
8.8 The Standard Deviation of a Probability Distribution 233
♢ 8.9 Technical Note (Simulation) 240
8.10 Checklist of Key Terms 245
8.11 Review Exercises 245
8.12 References 250

9 THE NORMAL DISTRIBUTION 251

9.1 Continuous Distribution 252
9.2 The Normal Distribution 255
♢ 9.3 Check for "Normality" 265
9.4 Applications of the Normal Distribution 269
9.5 The Normal Approximation to the Binomial Distribution 272
♢ 9.6 Technical Note (Simulation) 278
9.7 Checklist of Key Terms 279
9.8 Review Exercises 279
9.9 References 282

10 SAMPLING AND SAMPLING DISTRIBUTIONS 283

10.1 Random Sampling 284
♢ 10.2 Sample Designs 290
♢ 10.3 Systematic Sampling 290
♢ 10.4 Stratified Sampling 291
♢ 10.5 Cluster Sampling 293
10.6 Sampling Distributions 297
10.7 The Standard Error of the Mean 300
10.8 The Central Limit Theorem 303
10.9 Some Further Considerations 306
♢ 10.10 Technical Note (Simulation) 310
10.11 Checklist of Key Terms 312
10.12 Review Exercises 312
10.13 References 314

11 INFERENCES ABOUT MEANS 316

11.1 The Estimation of Means 317
11.2 The Estimation of Means (Small Samples) 323
♢ 11.3 The Estimation of Means (A Bayesian Method) 326
11.4 Tests of Hypotheses 335
11.5 Significance Tests 343
11.6 Tests Concerning Means 349
11.7 Tests Concerning Means (Small Samples) 354
11.8 Differences Between Means 360
11.9 Differences Between Means (Small Samples) 363
11.10 Differences Between Means (Paired Data) 364
11.11 Checklist of Key Terms 370
11.12 Review Exercises 371
11.13 References 374

12 INFERENCES ABOUT STANDARD DEVIATIONS 375

12.1 The Estimation of σ 376
12.2 Tests Concerning Standard Deviations 381
12.3 Tests Concerning Two Standard Deviations 384
12.4 Checklist of Key Terms 389
12.5 Review Exercises 389
12.6 References 391

13 INFERENCES ABOUT PROPORTIONS 392

13.1 The Estimate of Proportions 393
♢ 13.2 The Estimation of Proportions (A Bayesian Method) 398
13.3 Tests Concerning Proportions 404
13.4 Differences Between Proportions 408
13.5 The Analysis of an $r \times c$ Table 416
13.6 Goodness of Fit 429
13.7 Checklist of Key Terms 436
13.8 Review Exercises 436
13.9 References 441

14 ANALYSIS OF VARIANCE 443

14.1 Differences Among k Means: An Example 444
14.2 The Design of Experiments: Randomization 449
14.3 One-Way Analysis of Variance 450
14.4 Multiple Comparisons 457
14.5 The Design of Experiments: Blocking 463
14.6 Two-Way Analysis of Variance 465
14.7 The Design of Experiments: Replication 472
14.8 Latin Squares 472
14.9 The Design of Experiments: Some Further Considerations 476
14.10 Checklist of Key Terms 480
14.11 Review Exercises 480
14.12 References 483

15 REGRESSION 484

15.1 Curve Fitting 485
15.2 The Method of Least Squares 487
15.3 Regression Analysis 500
♢ 15.4 Multiple Regression 509
♢ 15.5 Nonlinear Regression 514
15.6 Checklist of Key Terms 521
15.7 Review Exercises 521
15.8 References 524

16 CORRELATION 525

16.1 The Coefficient of Correlation 526
16.2 The Interpretation of r 531
16.3 Correlation Analysis 537
16.4 Correlations of Count Data 541
16.5 Multiple and Partial Correlation 543
16.6 Checklist of Key Terms 547
16.7 Review Exercises 548
16.8 References 549

17 NONPARAMETRIC TESTS 550

17.1 The Sign Test 551
17.2 The Sign Test (Large Samples) 554
17.3 Rank Sums: The Signed-Rank Test 557
17.4 Rank Sums: The Signed-Rank Test (Large Samples) 560
17.5 Rank Sums: The U Test 564
17.6 Rank Sums: The U Test (Large Samples) 568
17.7 Rank Sums: The H Test 570
17.8 Tests of Randomness: Runs 574
17.9 Tests of Randomness: Runs (Large Samples) 576
17.10 Tests of Randomness: Runs Above and Below the Median 577
17.11 Rank Correlation 579
17.12 Some Further Considerations 583
17.13 Summary 584
17.14 Checklist of Key Terms 584
17.15 Review Exercises 585
17.16 References 588

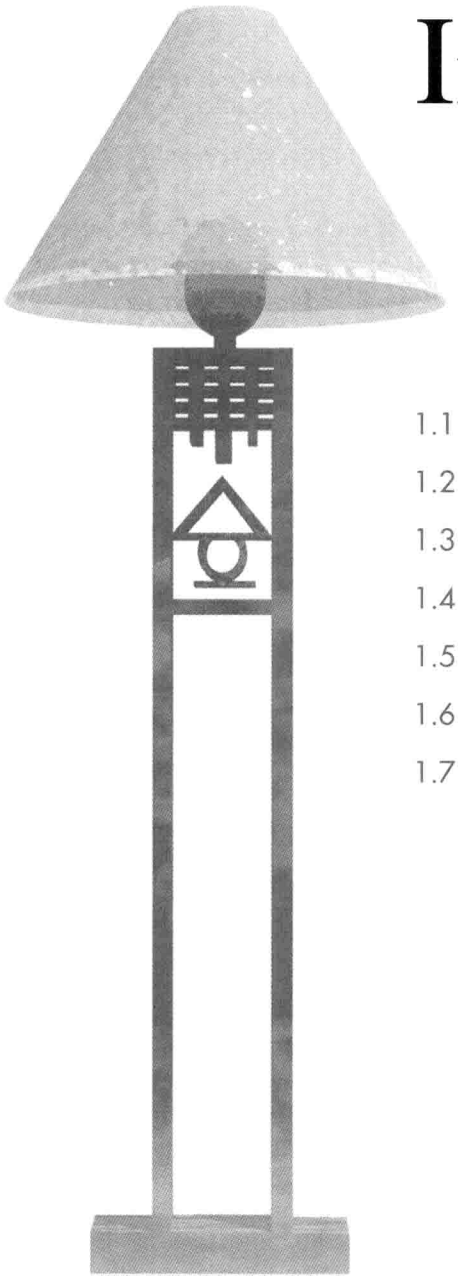
STATISTICAL TABLES 589

I Normal-Curve Areas 589
II Critical Values of t 590
III Critical Values of χ^2 591
IV Critical Values of F 592
V Binomial Probabilities 596
VI Critical Values of T 601
VII Critical Values of U 602
VIII Critical Values of u 604
IX Values of $Z = \frac{1}{2} \cdot \ln \frac{1+r}{1-r}$ 606
X Binomial Coefficients 607
XI Values of e^{-x} 608
XII Critical Values of q 609

ANSWERS TO SELECTED EXERCISES A-1

INDEX I-1

CHAPTER 1



INTRODUCTION

1.1 The Growth of Modern Statistics 2

1.2 The Study of Statistics 3

1.3 Descriptive Statistics and Statistical Inference 5

1.4 The Nature of Statistical Data 7

1.5 Checklist of Key Terms 11

1.6 Review Exercises 11

1.7 References 12

The collection, processing, interpretation, and presentation of numerical data all belong to the domain of statistics. These tasks include the calculation of baseball batting averages, collecting data on births and deaths, evaluating the effectiveness of commercial products, and forecasting the weather. Statistical information is presented to us constantly on radio and television. Our enthusiasm for statistical facts is encouraged by national newspapers such as The Wall Street Journal *and* USA Today.

The word "statistics" is used in several ways. It can refer not only to the mere tabulation of numeric information, as in reports of stock market transactions, but also to the body of techniques used in processing or analyzing data.

The word "statistician" is also used in several ways. The term can be applied to those who simply collect information, as well as to those who prepare analyses or interpretations, and it is also applied to scholars who develop the mathematical theory on which statistics is based.

In Sections 1.1 and 1.2 we discuss the recent growth of statistics and its ever-widening range of applications. In Section 1.3 we explain the distinction between the two major branches of statistics, descriptive statistics and statistical inference, and in the optional Section 1.4 we discuss the nature of various kinds of data and in connection with this warn the reader against the indiscriminate mathematical treatment of statistical data. ◩

1.1 THE GROWTH OF MODERN STATISTICS

There are several reasons why the scope of statistics and the need to study statistics have grown enormously in the last fifty or so years. One reason is the increasingly quantitative approach employed in all the sciences, as well as in business and many other activities which directly affect our lives. This includes the use of mathematical techniques in the evaluation of antipollution controls, in inventory planning, in the analysis of traffic patterns, in the study of the effects of various kinds of medications, in the evaluation of teaching techniques, in the analysis of competitive behavior of business managers and governments, in the study of diet and longevity, and so forth. The availability of powerful computers has greatly increased our ability to deal with numerical information. Many types of computers are also inexpensive, so that sophisticated statistical work can be done by small businesses, college students, and even high school students.

The other reason is that the amount of data that is collected, processed, and disseminated to the public for one reason or another has increased almost beyond comprehension, and what part is good statistics and what part is bad statistics is anybody's guess. To act as watchdogs, more and more persons with some knowledge of statistics are needed to take an active part in the collection of the data, in the analysis of the data, and, what is equally important, in all of the preliminary planning. Without the latter, it is frightening to think of all the

things that can go wrong in the compilation of statistical data. The results of costly surveys can be useless if questions are ambiguous or asked in the wrong way, if they are asked of the wrong persons, in the wrong place, or at the wrong time. Much of this is just common sense, as is illustrated by the following examples:

Example

To determine public sentiment about the continuation of a certain government program, an interviewer asks: "Do you feel that this wasteful program should be stopped?" Explain why this question will probably not yield the desired information.

solution

The interviewer is begging the question by suggesting, in fact, that the program is wasteful. ◩ ◩ ◩

Example

To study consumer reaction to a new convenience food, a house-to-house survey is conducted during weekday mornings, with no provisions for return visits in case no one is home. Explain why this approach may well yield misleading information.

solution

This survey will fail to reach those who are most likely to use the product: single persons and married couples with both spouses employed. ◩ ◩ ◩

Although much of the above-mentioned growth of statistics began prior to the computer revolution, the widespread availability and use of computers have greatly accelerated the process. In particular, computers enable us to handle, analyze, and dissect large masses of data, and they enable us to perform calculations which previously had been too cumbersome even to contemplate. Our objective in this book is your gaining an understanding of the ideas of statistics. Access to a computer is not critical for this objective. Computer uses are occasionally illustrated in this textbook, but nearly all the exercises can be done with nothing more than a four-function calculator.

1.2 THE STUDY OF STATISTICS

The subject of statistics can be presented at various levels of mathematical difficulty, and it may be directed toward applications in various fields of inquiry. Accordingly, many textbooks have been written on business statistics, educational statistics, medical statistics, psychological statistics, . . . , and even on statistics for historians. Although problems arising in these various disciplines will sometimes require special statistical techniques, none of the basic methods discussed in this book is restricted to any particular field of application. In the same way in which $2 + 2 = 4$ regardless of whether we are adding dollar amounts, horses, or trees, the methods we shall present provide **statistical models** which

apply regardless of whether the data are IQs, tax payments, reaction times, humidity readings, test scores, and so on. To illustrate this further, consider an exercise from Chapter 13, specifically exercise 13.104

13.104 In a random sample of 200 retired persons, 137 stated that they prefer living in an apartment to living in a one-family home. At the 0.05 level of significance does this refute the claim that 60 percent of all retired persons prefer living in an apartment to living in a one-family home?

The question asked here should be clear, and it should also be clear that the answer would be of interest mainly to social scientists or to persons in the construction industry. However, if we wanted to cater to the special interests of students of biology, engineering, education, or ecology, we might rephrase the exercise as follows:

13.104 In a random sample of 200 citrus trees exposed to a 20° frost, 137 showed some damage to their fruit. At the 0.05 level of significance does this refute the claim that 60 percent of all citrus trees exposed to a 20° frost will show some damage to their fruit?

13.104 In a random sample of 200 transistors made by a given manufacturer, 137 passed an accelerated performance test. At the 0.05 level of significance does this refute the claim that 60 percent of all the transistors made by the manufacturer will pass the test?

13.104 In a random sample of 200 high school seniors in a large city, 137 said that they will go on to college. At the 0.05 level of significance does this refute the claim that 60 percent of all the high school seniors in this city will go on to college?

13.104 In a random sample of 200 cars tested for the emission of pollutants, 137 failed to meet a state's legal standards. At the 0.05 level of significance does this refute the claim that 60 percent of all cars tested in this state will fail to meet its legal emission standards?

Insofar as the work in this book is concerned, the statistical treatment of all these versions of Exercise 13.104 is the same, and with some imagination the reader should be able to rephrase it for virtually any field of specialization. As some authors do, we could present, and so designate, special problems for readers with special interests, but this would defeat our goal of impressing upon the reader the importance of statistics in all of science, business, and everyday life. To attain this goal, we have included in this text exercises covering a wide spectrum of interests.

To avoid the possibility of misleading anyone with our various versions of Exercise 13.104, let us make it clear that we cannot squeeze all statistical problems into the same mold. Although the methods we study in this book are all widely applicable, it is always important to make sure that the statistical model we are using is the right one.

EXERCISES

1.1 Rephrase Exercise 13.104 referred to above, so that it will be of special interest to

(a) a cosmetics salesperson; (b) a musician; (c) a traffic engineer.

1.2 Bad statistics may well result from asking questions in the wrong way or of the wrong persons. Explain why the following may lead to useless data:

(a) To determine public sentiment about a certain foreign trade restriction, an interviewer asks voters: "Do you feel that this unfair practice should be stopped?"

(b) To predict a municipal election, a public opinion poll telephones persons selected haphazardly from the city's telephone directory.