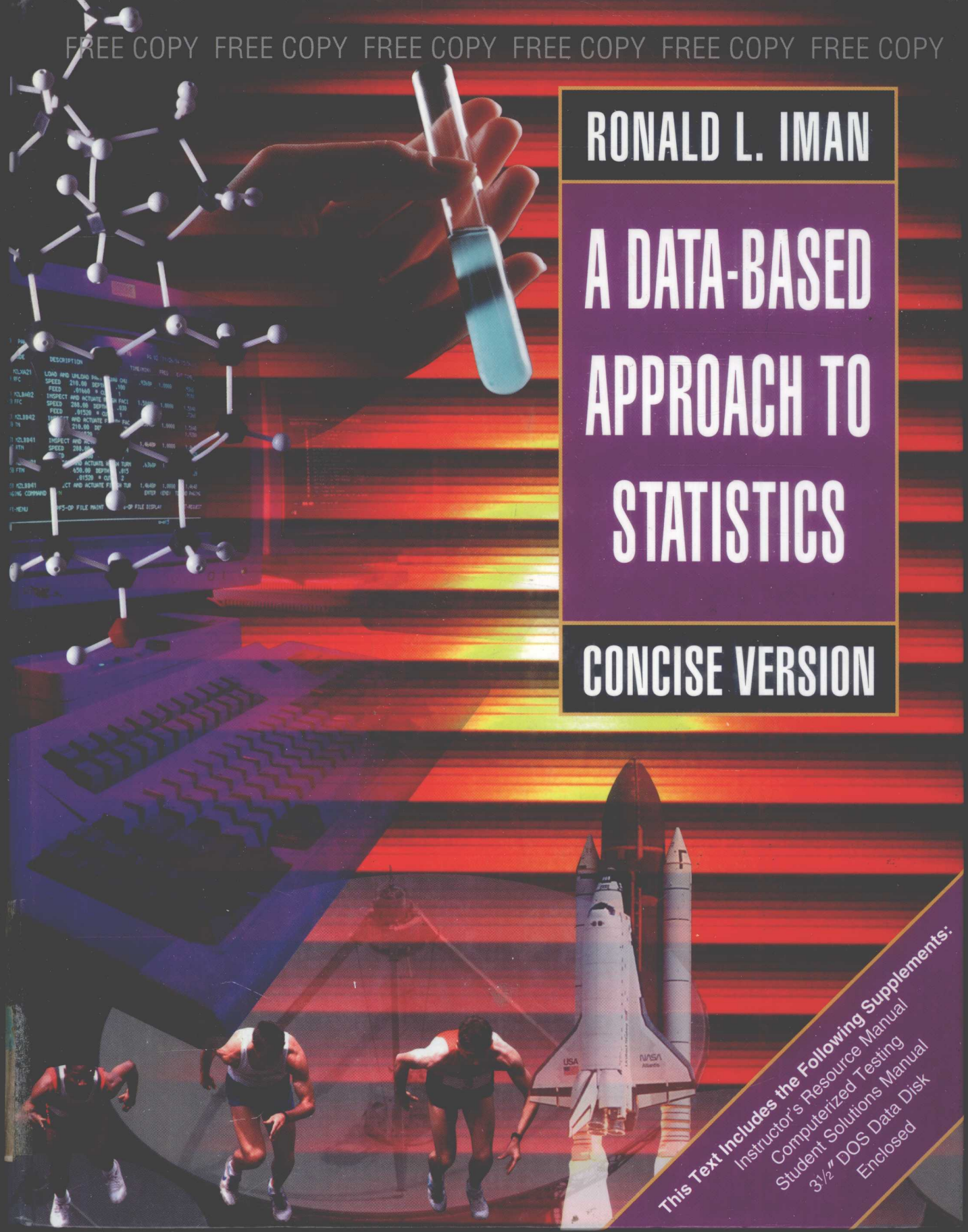
FREE COPY FREE COPY FREE COPY FREE COPY FREE COPY FREE COPY



RONALD L. IMAN

A DATA-BASED APPROACH TO STATISTICS

CONCISE VERSION

This Text Includes the Following Supplements:
Instructor's Resource Manual
Computerized Testing
Student Solutions Manual
3½″ DOS Data Disk
Enclosed

A DATA-BASED APPROACH TO STATISTICS
CONCISE VERSION

RONALD L. IMAN

Sandia National Laboratories

Duxbury Press
An Imprint of Wadsworth Publishing Company
I(T)P™ An International Thomson Publishing Company

Belmont • Albany • Bonn • Boston • Cincinnati • Detroit • London • Madrid • Melbourne
Mexico City • New York • Paris • San Francisco • Singapore • Tokyo • Toronto • Washington

Assistant Editor: Jennifer Burger
Editorial Assistant: Michelle O'Donnell
Production Editor: Sandra Craig
Print Buyer: Randy Hurst
Permissions Editor: Peggy Meehan
Cover and Text Design: Juan Vargas, Vargas/Williams/Design
Illustrations: Paloma Design & Production
Compositor: Interactive Composition Corporation
Printer: R.R. Donnelley & Sons/Crawfordsville

COPYRIGHT © 1995 by Wadsworth Publishing Company
A Division of International Thomson Publishing Inc.
I(T)P The ITP logo is a trademark under license.

Printed on acid-free recycled paper.

Printed in the United States of America
1 2 3 4 5 6 7 8 9 10—01 00 99 98 97 96 95

For more information, contact Wadsworth Publishing Company:

Wadsworth Publishing Company
10 Davis Drive
Belmont, California 94002, USA

International Thomson Publishing Europe
Berkshire House 168-173
High Holborn
London, WC1V 7AA, England

Thomas Nelson Australia
102 Dodds Street
South Melbourne 3205
Victoria, Australia

Nelson Canada
1120 Birchmount Road
Scarborough, Ontario
Canada M1K 5G4

International Thomson Editores
Campos Eliseos 385, Piso 7
Col. Polanco
11560 México D.F. México

International Thomson Publishing GmbH
Königswinterer Strasse 418
53227 Bonn, Germany

International Thomson Publishing Asia
221 Henderson Road
#05-10 Henderson Building
Singapore 0315

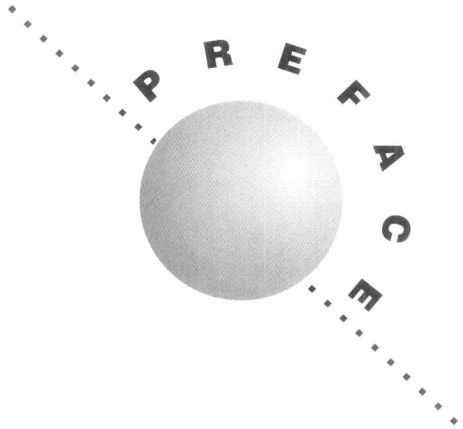
International Thomson Publishing Japan
Hirakawacho Kyowa Building, 3F
2-2-1 Hirakawacho
Chiyoda-ku, Tokyo 102, Japan

All rights reserved. No part of this work covered by the copyright hereon may be reproduced or used in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems—without the written permission of the publisher.

Library of Congress Cataloging-in-Publication Data

Iman, Ronald L.
A data-based approach to statistics : concise version / Ronald L. Iman.
p. cm.
Includes bibliographical references and index.
ISBN 0-534-23496-8 (acid-free)
1. Statistics. I. Title.
QA276.12.I48 1994b
001.4′22—dc20 94-19342

OR SALE INSTRUCTOR'S REVIEW COPY NOT FOR SALE INSTRUCTOR'S REVIEW COPY NOT FOR SALE



The Usefulness of Statistics

In this text, I have used a variety of practical situations to demonstrate the wide-ranging applications of statistics and to convey the value of statistics for students in their future employment. Statistics plays a very important role in almost every aspect of society. For example, medical studies are invariably based on statistical methods. Environmental determinations and risk assessments rely heavily on the use of statistics, as do economic forecasts. Japanese companies have honed data collection to an art and skillfully applied statistics to increase the quality of their manufactured products. This has been a major contributing factor in Japan's enviable position in the world marketplace. These are but a few examples of the usefulness of statistics and the benefits that it provides to virtually all professions.

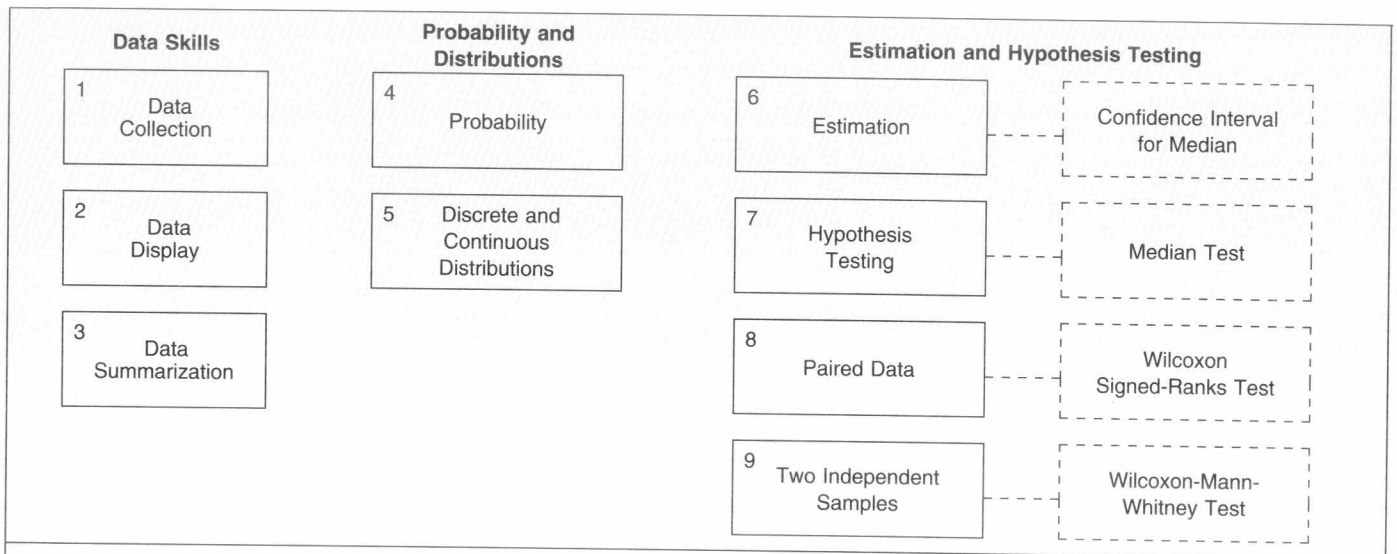
Areas of Applications

To illustrate the usefulness of statistics, I have used actual applications in such diverse fields as agriculture, biology, business, economics, education, engineering, environmentally conscious manufacturing, genetics, home economics, medicine, pharmacology, physical education, physiology, psychology, sports, and sociology. Many of these applications are based on my consulting experiences or influenced by them.

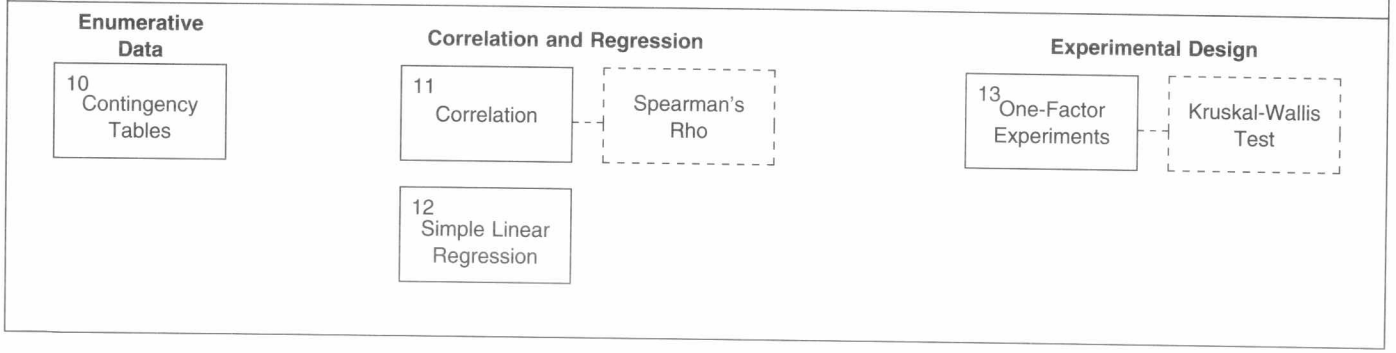
Structure of the Text

A Data-Based Approach to Statistics, Concise Version, offers a flexible one-term first course in statistics. The diagram that accompanies this preface illustrates the basic structure of the text. The first nine chapters are referred to as *basic core chapters,* which would usually be covered in all courses. These nine chapters are grouped into three areas: *data skills, probability and distributions,* and *estimation and hypothesis*

Basic Core Chapters



Enrichment Chapters



此为试读，需要完整PDF请访问：www.ertongbook.com

NOT FOR SALE INSTRUCTOR'S REVIEW COPY NOT FOR SALE INSTRUCTOR'S REVIEW COPY NOT FO

OR SALE INSTRUCTOR'S REVIEW COPY NOT FOR SALE INSTRUCTOR'S REVIEW COPY NOT FOR SALE

testing. These chapters can be supplemented by any combination of material from the last four chapters, which cover *analysis of enumerative data, correlation and regression,* and *experimental design.*

In the accompanying diagram, a dotted box is used to identify nonparametric techniques, which appear in several chapters. The parametric and nonparametric procedures are presented side-by-side to provide immediate access to the appropriate nonparametric analog when the parametric assumptions are not satisfied. *This format is one of convenience. The instructor need not feel obligated to take the time to cover these sections.* Any of the nonparametric topics can be omitted without loss of continuity, for a completely traditional course. The text is divided into 45 approximately lesson-sized sections.

The Approach

Each section starts with a realistic Data Setting that is used to illustrate the need for the methods that are introduced. The Data Setting and other examples are then used to demonstrate the methods presented in the section. Throughout the text, students are encouraged to use graphical displays to help understand the information contained in the sample data. These displays serve as a check on the reasonableness of the assumptions and as a basis for preliminary inferences, prior to the presentation of formal statistical methods.

Use of Computer Software

The development of statistical software such as Minitab®, Execustat®, SYSTAT®, and SAS® (among many others), and the availability of personal computers, has eliminated most of the computational burden (including graphical displays). Throughout this text, computer output has been heavily incorporated—primarily from Minitab, SYSTAT, and Execustat.

Data Disk

To facilitate use of the computer, a data disk accompanies this text. This disk contains 236 files that correspond to all data sets that appear in either figures or exercises. The files are easily accessed by figure or exercise number (see the README file on the disk) and are given in Minitab, SYSTAT, Execustat, and ASCII formats. The variable names are given in all but the ASCII files. Also provided on the disk are Minitab and SYSTAT macros, for ease in performing the Lilliefors test for normality.

Exercises

I have selected exercises that further emphasize actual applications. These exercises often lend themselves to classroom discussion (as do the Data Settings). Each section concludes with a set of exercises, and each chapter concludes with a set of review

OPY NOT FOR SALE INSTRUCTOR'S REVIEW COPY NOT FOR SALE INSTRUCTOR'S REVIEW COPY NOT FO

exercises. The exercises within a set are usually structured from easy to more sophisticated. Most exercise sets contain some exercises with computer output and some without. When the computer output is given, it serves two purposes. First, it provides the student with practice in interpreting computer output without actually using the computer. Second, it serves as a check for the student's own use of software. When the output is not given, the student is encouraged to use computer software to analyze the data. The latter exercises allow students to analyze the data on their own—as in actual applications.

Computer Lab

As the preceding discussion indicates, the format used in the exercises makes it easy to use the text in conjunction with a computer lab. The use of computer software is encouraged throughout. Exercises that specifically ask the student to use computer software have been marked with a special symbol: . Figures in the text that contain computer output often show the corresponding computer instructions. Each chapter concludes with a set of Minitab instructions related to the topics in the chapter. A brief introduction to Minitab is given in the instructor's manual and in the student answer book.

Acknowledgments

In writing this text, I have been guided by the helpful input of several reviewers:

William H. Beyer, University of Akron
W. J. Conover, Texas Tech University
Iris B. Fetta, Clemson University
Joseph Glaz, University of Connecticut
J. Stuart Hunter, Princeton University (emeritus)
Charles D. Kincaid, University of Florida
Linda C. Malone, University of Central Florida
Robert L. Mason, Southwest Research Institute
Paul I. Nelson, Kansas State University
Judah Rosenblatt, University of Texas Medical Branch, Galveston
Thomas P. Ryan, independent consultant
G. Morris Southward, New Mexico State University
Bruce E. Trumbo, California State University, Hayward
Joseph G. Walker, Georgia State University

I am very grateful for their help in shaping the final form of this text. I am also indebted to David Hoyt for excellent editing, to Juan Vargas for an outstanding design,

OR SALE INSTRUCTOR'S REVIEW COPY NOT FOR SALE INSTRUCTOR'S REVIEW COPY NOT FOR SALE

to the staff at Duxbury Press for their help in many different areas, and to Alex Kugushev for his unique understanding of student needs and for his unending support and encouragement. I hope that you find using this text to be a refreshing experience.

To the Student

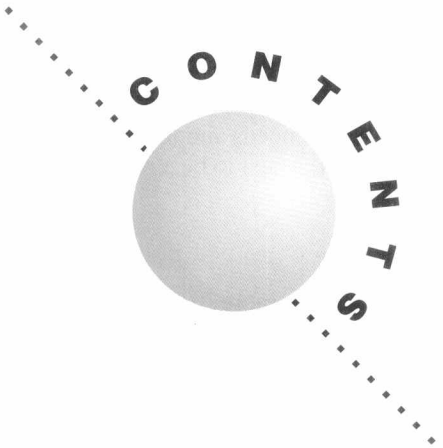
Have you ever wondered why some people are more successful than others, or perhaps why some long-established businesses suddenly fall on hard times? There are many possible answers to these questions, but success is often closely linked to the ability to obtain information, process it, and make decisions accordingly. Life is much like this. When you were young, you probably touched something hot and burned your finger (you obtained some information). Your finger hurt (you processed the information), and then you decided to avoid touching hot things (you made a decision). As we get older, situations become more complex, requiring us to gather and process more information before making decisions.

Information gathering requires careful planning and processing if the information is to be of assistance in intelligent decision making. This text was designed to help you learn how to gather information, process it, and make decisions. Numerous real-life situations are used to illustrate the entire process.

Information gathering often involves data in various forms. Statistics is a broad discipline that is concerned with methods for collecting and processing data and subsequently making defensible decisions. *A Data-Based Approach to Statistics, Concise Version,* contains insights from my experiences as a consultant in engineering applications, environmental issues, risk assessment, and pharmaceuticals (among other areas), to make clear that statistics can become a useful tool for you. I hope that this text will help you perceive the value of statistics in your life and career and that you will often take the opportunity to use it as a ready reference.

R.L.I.

PY NOT FOR SALE INSTRUCTOR'S REVIEW COPY NOT FOR SALE INSTRUCTOR'S REVIEW COPY NOT FO



I DATA SKILLS 1

1. Data Collection 2

Good Data Skills: A Key to Success in Your Career 3

1.1 Populations and Samples 6

1.2 The Importance of Random Samples 12

1.3 Experiments and Data Types 16

Review Exercises 20

2. Data Display 22

Graphical Display: The Art of Communication 23

2.1 Dotplots and Stem-and-Leaf Plots 28

2.2 Histograms 38

2.3 Cumulative Relative Frequency 47

2.4 Scatterplots 53

Review Exercises 59

3. Data Summarization 64

How Well Did I Do on My Test? 65

3.1 The Sample Mean 66

3.2 The Sample Standard Deviation 73

3.3 Modes, Quantiles, Proportions, and Boxplots 82

Review Exercises 94

FOR SALE INSTRUCTOR'S REVIEW COPY NOT FOR SALE INSTRUCTOR'S REVIEW COPY NOT FOR SALE

II PROBABILITY AND DISTRIBUTIONS 97

4. Probability and Discrete and Continuous Populations 98

An Experiment with a Point 99

4.1 Probability 100

4.2 Discrete Random Variables 116

4.3 Continuous Random Variables 128

4.4 Population Parameters 138

Review Exercises 148

5. Discrete and Continuous Distributions 153

Probability Models for Experimental Outcomes 154

5.1 The Binomial Distribution 155

5.2 The Normal Distribution 168

5.3 The Normal Approximation to the Binomial and Checking for Normality 181

5.4 The Distribution of the Sample Mean 193

Review Exercises 211

III ESTIMATION AND HYPOTHESIS TESTING 215

6. Estimation (One Sample) 216

Estimates Based on Random Samples 217

6.1 General Remarks about Estimation 218

6.2 Estimating the True Proportion in a Population 227

6.3 Estimating the Mean 236

6.4 Estimating the Population Median (Optional) 246

6.5 Estimating the Population Standard Deviation (Optional) 256

Review Exercises 265

7. Hypothesis Testing 268

Decisions Based on Casual Observations 269

7.1 Hypotheses, Test Statistics, and p-Values 271

7.2 The Decision Rule and Power 280

7.3 The Binomial Test and Acceptance Sampling 292

7.4 Testing Hypotheses about the Population Mean and Standard Deviation 304

COPY NOT FOR SALE INSTRUCTOR'S REVIEW COPY NOT FOR SALE INSTRUCTOR'S REVIEW COPY NOT F

7.5 Testing Hypotheses about the Population Median (Optional) 314
Review Exercises 324

8. Two Related Samples (Matched Pairs) 327

The Use of Paired Observations in Experiments 328
8.1 The Paired *t*-Test 329
8.2 The Nonparametric Approach: The Wilcoxon Signed-Ranks Test 341
Review Exercises 356

9. Estimation and Hypothesis Testing with Two Independent Samples 361

Experiments Based on Independent Samples 362
9.1 Large Samples: Inferences about the Difference between Two Means 363
9.2 Difference between Two Means: Normal Populations with Equal Variance (Two-Sample *t*-Test) 369
9.3 Difference between Two Means: Normal Populations with Unequal Variances (Satterthwaite's Test) (Optional) 377
9.4 The Nonparametric Approach: The Wilcoxon-Mann-Whitney Rank Sum Test 383
Review Exercises 393

IV ENUMERATIVE DATA 399

10. Analysis of Enumerative Data 400

Drug Screening of Employment Applicants 401
10.1 2 × 2 Contingency Tables 402
10.2 The $r \times c$ Contingency Table 412
10.3 The Chi-Square Goodness-of-Fit Test (Optional) 426
Review Exercises 433

V CORRELATION AND REGRESSION 441

11. Correlation 442

Correlation versus Cause and Effect 443
11.1 Introduction to Correlation 445

此为试读,需要完整PDF请访问: www.ertongbook.com

FOR SALE INSTRUCTOR'S REVIEW COPY NOT FOR SALE INSTRUCTOR'S REVIEW COPY NOT FOR SALE

11.2 The Nonparametric Approach: Spearman's Rho 458
Review Exercises 472

12. Simple Linear Regression 483

Predicting Your Starting Salary 484
12.1 An Introduction to Linear Regression 488
12.2 Inference for Linear Regression 500
Review Exercises 516

13. One-Factor Experiments 522

Can You Rely on the Results of Studies Reported in the Media? 523
13.1 Graphical Display for One-Factor Experiments 524
13.2 A Model for One-Factor Experiments 532
13.3 Comparing Means of Factor Levels 546
13.4 The Nonparametric Approach: The Kruskal-Wallis Test 559
Review Exercises 571

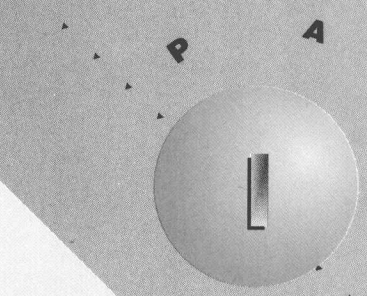
References A-1

Appendix A Tables A-5

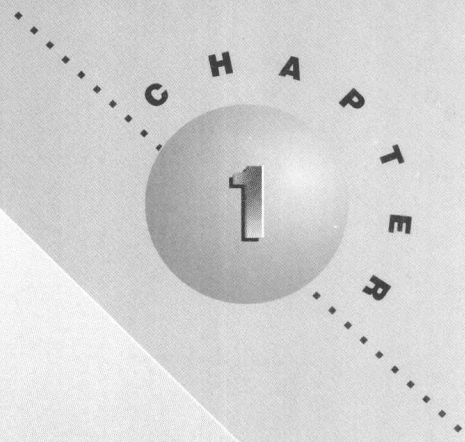
A1 Binomial Distribution A-6
A2 Normal Distribution for Finding Quantiles A-12
A2* Normal Distribution for Finding Probabilities A-14
A3 Student's *t* Distribution A-16
A4 Chi-Square Distribution A-18
A5 *F* Distribution A-20
A6 Exact Confidence Intervals for the Binomial Parameter *p* A-26

Appendix B Brief Answers to Odd-Numbered Exercises A-28

Index A-37



DATA SKILLS



DATA COLLECTION

GOOD DATA SKILLS: A Key to Success in Your Career

Many students taking statistics for the first time do not realize how *good data skills* will make them more effective in their careers. The need for such skills is often not recognized until situations arise on the job. As you proceed through this text, keep the following statement in mind.

> *Having good data skills—gathering, summarizing, displaying, and interpreting data—will serve you well throughout your career.*

Data, when properly collected, analyzed, displayed, and interpreted, help make you an *intelligent decision maker*. Such a person will succeed where others fail. This text will help you acquire the data skills you need.

Many people think of data simply as numbers. The dictionary definition of *data* provides a broader view:

> **data** (plural) things known or assumed; facts or figures from which conclusions can be inferred; information

The essence of this definition is that data contain *information* that assists decision making.

The following example illustrates how decisions are commonly made on the basis of casual observations.

EXAMPLE

DETERMINING THE PERCENTAGE OF JAPANESE CARS ON THE ROAD

While waiting for a traffic light to change you observed that 25 of the 40 cars passing through the intersection were Japanese-made. On the basis of this observation, you tell a friend that most of the cars on the road today are Japanese. Your statement is clearly based on data you have collected. However, do your data represent a passive observation on your part? Or were they collected as part of a specific plan designed to answer questions about the percentage of Japanese cars on the road? If not, your data may not contain sufficient *information* to answer specific questions, because of how the data were collected. Therefore, your data do not have much value in making a decision about the percentage of Japanese cars on the road.

The Subject of Statistics

The field of **statistics** is broad, but it often centers on the collection of *information-laden data* and the methods that can be used to analyze them. That is, statistics is the discipline that enables you to acquire good data skills.

The Subject of Statistics

Statistics is a broad discipline that includes data analysis. It is concerned with methods for collecting, organizing, summarizing, presenting, and analyzing data, as well as making valid conclusions about the characteristics of the sources from which the data were obtained.

Thus, statistics is intimately involved with *data analysis*. The types of analyses that can be performed are quite diverse, and they require statistical methods that are equally diverse. The statistical methods presented in this text have been selected on the basis of their proven usefulness in application after application.

Each section in this book begins with a *Data Setting*. These Data Settings illustrate a variety of actual situations in which data have a central role and show how you can use data to make intelligent decisions. This text will cover many of the statistical applications that you are most likely to encounter in your work, in your own projects, and in reports written by colleagues in your field.

The title of this text, *A Data-Based Approach to Statistics*, was chosen partly to emphasize the close link between data and the subject of statistics. It is written in a user-friendly manner, which will allow you to refer to it in the future when the need arises. The data settings are realistic and have been chosen to make the explanations of the methods easier to understand. Great effort has been taken to make the subject of statistics understandable without sacrificing the accuracy and correctness of the material.

Will You Ever Need to Use Statistics?

The following examples illustrate how some recent college graduates have discovered the need for good data skills.

A biology teacher: There was one aspect of grading that I hadn't given any thought. After I gave my first test, it occurred to me that the students in the class would expect some kind of summarization or display of the test scores to see how they had performed relative to the rest of the class. Although I had been on the receiving end of test scores for many years, I didn't know how to prepare a summary for my class.

A marketing analyst for a pharmaceutical firm: My company had just received approval for a new drug used in the treatment of patients with high blood pressure. I was given sales information on the market share held by our competitors. My instructions were to prepare a display of the information and present it at a briefing for regional sales managers, who were coming to the home office for a meeting the next week. I wasn't sure how to prepare a display of the information so it could be understood by the managers.

An accountant with an auditing firm: On my first assignment, I was a member of an auditing team that had been hired to perform an audit of a real estate firm. My supervisor

told me to select a sample of the sales of small businesses made by the company in the past year. I had gone into the audit prepared to examine whatever accounts were given to me. I didn't realize that I would be the one to select the accounts, and I didn't know how to go about selecting a sample that I could defend as reasonable.

A registered nurse: I was assigned to a post-operative ward for kidney transplant patients. The vital signs of these patients have to be monitored continuously after surgery. I recorded their temperature, blood pressure, and pulse rate every 15 minutes. The transplant surgeon wanted all this information displayed on a single chart that could be easily read to monitor the patient's condition. I thought that just writing down the readings would be enough, and I didn't know how to make a chart to display all the information the surgeon wanted to see.

An electrical engineer: I was assigned to work on a new, environmentally safe, no-clean soldering process used to produce printed wiring boards for personal computers. This project appealed to my environmental conscience, because the old process required the use of ozone-depleting chemicals to clean the boards after soldering. My job was to find the proper machine settings to produce boards with the fewest possible soldering defects. I was told that the preheat temperature and the time in the soldering pot were very important factors influencing the number of soldering defects found on the board. The preheat temperature could be either low or high, and the soldering time could be short or long. In addition, production had to be interrupted when anyone wanted to experiment with the soldering machine. My supervisor told me that I could use the machine on four different nights, with a technician available to adjust the machine one time for the night's experiment. I didn't know how to gather useful information that I would be able to explain and defend.

These people's experiences are typical of job-related situations that require *good data skills*—that is, skills in *gathering, summarizing, displaying,* and *interpreting data* in various forms. These individuals do not need to be trained statisticians; they just need basic data skills to help them perform their jobs effectively.

This is a good time for you to assess your present skill level. Before reading further, take time to prepare a short summary of ways in which these individuals could deal with the problems they encountered in their new jobs. It may be helpful to think of yourself as their immediate supervisor, to whom they have turned for advice. Be prepared to present your ideas in class. There are no right or wrong answers for this task; let your imagination be your guide.

此为试读,需要完整PDF请访问: www.ertongbook.com