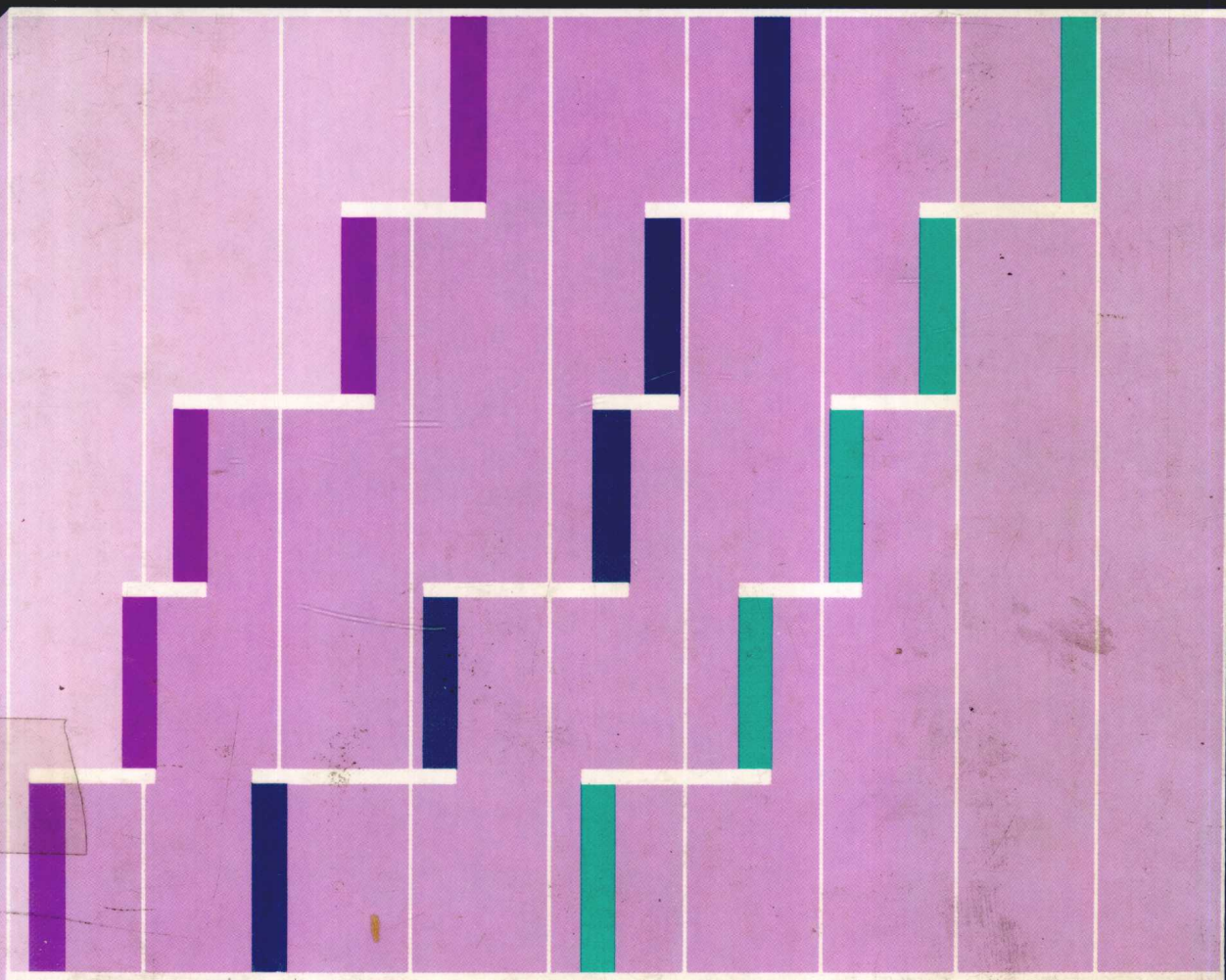


RONALD L. IMAN
W.J. CONOVER

A MODERN APPROACH TO STATISTICS



A Modern Approach To Statistics

RONALD L. IMAN

Sandia National Laboratories

W. J. CONOVER

Texas Tech University

JOHN WILEY & SONS

New York Chichester Brisbane
Toronto Singapore

Copyright © 1983, by John Wiley & Sons, Inc.
All rights reserved. Published simultaneously in Canada.
Reproduction or translation of any part of
this work beyond that permitted by Sections
107 and 108 of the 1976 United States Copyright
Act without the permission of the copyright
owner is unlawful. Requests for permission
or further information should be addressed to
the Permissions Department, John Wiley & Sons.

Library of Congress Cataloging in Publication Data:

Iman, Ronald L.

A modern approach to statistics.

Includes index.

I. Conover, W. J. II. Title.

QA276.15.I43 1983 519.5 82-8344

ISBN 0-471-09667-9 AACR2

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

WILEY SERIES IN PROBABILITY AND MATHEMATICAL STATISTICS

ESTABLISHED BY WALTER A. SHEWHART AND SAMUEL S. WILKS
Editors

*Ralph A. Bradley, J. Stuart Hunter, David G. Kendall,
Rupert G. Miller, Jr., Geoffrey S. Watson*

Probability and Mathematical Statistics

ADLER • The Geometry of Random Fields
ANDERSON • The Statistical Analysis of Time Series
ANDERSON • An Introduction to Multivariate Statistical Analysis
ARAUJO and GINE • The Central Limit Theorem for Real and Banach
Valued Random Variables
ARNOLD • The Theory of Linear Models and Multivariate Analysis
BARLOW, BARTHOLOMEW, BREMNER, and BRUNK • Statistical
Inference Under Order Restrictions
BARNETT • Comparative Statistical Inference, *Second Edition*
BHATTACHARYYA and JOHNSON • Statistical Concepts and Methods
BILLINGSLEY • Probability and Measure
CASSEL, SARNDAL, and WRETMAN • Foundations of Inference in
Survey Sampling
COCHRAN • Contributions to Statistics
DE FINETTI • Theory of Probability, Volumes I and II
DOOB • Stochastic Processes
FELLER • An Introduction to Probability Theory and Its Applications,
Volume I, *Third Edition*, Revised; Volume II, *Second Edition*
FULLER • Introduction to Statistical Time Series
GRENANDER • Abstract Inference
GUTTMAN • Linear Models: An Introduction
HANNAN • Multiple Time Series
HANSEN, HURWITZ, and MADOW • Sample Survey Methods and
Theory, Volumes I and II
HARDING and KENDALL • Stochastic Geometry
HOEL • Introduction to Mathematical Statistics, *Fourth Edition*
HUBER • Robust Statistics
IMAN and CONOVER • A Modern Approach to Statistics
IOSIFESCU • Finite Markov Processes and Applications
ISAACSON and MADSEN • Markov Chains
KAGAN, LINNIK, and RAO • Characterization Problems in Mathematical
Statistics
KENDALL and HARDING • Stochastic Analysis
LAHA and ROHATGI • Probability Theory
LARSON • Introduction to Probability Theory and Statistical Inference,
Third Edition
LEHMANN • Testing Statistical Hypotheses
MATTHES, KERSTAN, and MECKE • Infinitely Divisible Point Processes
MUIRHEAD • Aspects of Multivariate Statistical Analysis
PARZEN • Modern Probability Theory and Its Applications
PURI and SEN • Nonparametric Methods in Multivariate Analysis
RANGLES and WOLFE • Introduction to the Theory of Nonparametric
Statistics
RAO • Linear Statistical Inference and Its Applications, *Second Edition*
ROHATGI • An Introduction to Probability Theory and Mathematical
Statistics
ROSS • Stochastic Processes
RUBINSTEIN • Simulation and The Monte Carlo Method
SCHEFFE • The Analysis of Variance
SEBER • Linear Regression Analysis
SEN • Sequential Nonparametrics: Invariance Principles and Statistical
Inference

SERFLING • Approximation Theorems of Mathematical Statistics
 TJUR • Probability Based on Radon Measures
 WILLIAMS • Diffusions, Markov Processes, and Martingales, Volume I:
 Foundations
 ZACKS • Theory of Statistical Inference

Applied Probability and Statistics

ANDERSON, AUQUIER, HAUCK, OAKES, VANDAELE, and
 WEISBERG • Statistical Methods for Comparative Studies
 ARTHANARI and DODGE • Mathematical Programming in Statistics
 BAILEY • The Elements of Stochastic Processes with Applications to the
 Natural Sciences
 BAILEY • Mathematics, Statistics and Systems for Health
 BARNETT • Interpreting Multivariate Data
 BARNETT and LEWIS • Outliers in Statistical Data
 BARTHOLOMEW • Stochastic Models for Social Processes, *Third
 Edition*
 BARTHOLOMEW and FORBES • Statistical Techniques for Manpower
 Planning
 BECK and ARNOLD • Parameter Estimation in Engineering and Science
 BELSLEY, KUH, and WELSCH • Regression Diagnostics: Identifying
 Influential Data and Sources of Collinearity
 BENNETT and FRANKLIN • Statistical Analysis in Chemistry and the
 Chemical Industry
 BHAT • Elements of Applied Stochastic Processes
 BLOOMFIELD • Fourier Analysis of Time Series: An Introduction
 BOX • R. A. Fisher, The Life of a Scientist
 BOX and DRAPER • Evolutionary Operation: A Statistical Method for
 Process Improvement
 BOX, HUNTER, and HUNTER • Statistics for Experimenters: An
 Introduction to Design, Data Analysis, and Model Building
 BROWN and HOLLANDER • Statistics: A Biomedical Introduction
 BROWNE • Statistical Theory and Methodology in Science and
 Engineering, *Second Edition*
 BURY • Statistical Models in Applied Science
 CHAMBERS • Computational Methods for Data Analysis
 CHATTERJEE and PRICE • Regression Analysis by Example
 CHERNOFF and MOSES • Elementary Decision Theory
 CHOW • Analysis and Control of Dynamic Economic Systems
 CHOW • Econometric Analysis by Control Methods
 CLELLAND, BROWN, and deCANI • Basic Statistics with Business
 Applications, *Second Edition*
 COCHRAN • Sampling Techniques, *Third Edition*
 COCHRAN and COX • Experimental Designs, *Second Edition*
 CONOVER • Practical Nonparametric Statistics, *Second Edition*
 CONOVER and IMAN • Introduction to Modern Business Statistics
 CORNELL • Experiments with Mixtures: Designs, Models and The Analysis
 of Mixture Data
 COX • Planning of Experiments
 DANIEL • Biostatistics: A Foundation for Analysis in the Health Sciences,
Third Edition
 DANIEL • Applications of Statistics to Industrial Experimentation
 DANIEL and WOOD • Fitting Equations to Data: Computer Analysis of
 Multifactor Data, *Second Edition*
 DAVID • Order Statistics, *Second Edition*
 DAVISON • Multidimensional Scaling
 DEMING • Sample Design in Business Research
 DODGE and ROMIG • Sampling Inspection Tables, *Second Edition*
 DOWDY and WEARDEN • Statistics for Research
 DRAPER and SMITH • Applied Regression Analysis, *Second Edition*
 DUNN • Basic Statistics: A Primer for the Biomedical Sciences, *Second
 Edition*

DUNN and CLARK • Applied Statistics: Analysis of Variance and Regression
 ELANDT-JOHNSON • Probability Models and Statistical Methods in Genetics
 ELANDT-JOHNSON and JOHNSON • Survival Models and Data Analysis
 FLEISS • Statistical Methods for Rates and Proportions, *Second Edition*
 FRANKEN, KÖNIG, ARNDT, and SCHMIDT • Queues and Point Processes
 GALAMBOS • The Asymptotic Theory of Extreme Order Statistics
 GIBBONS, OLKIN, and SOBEL • Selecting and Ordering Populations: A New Statistical Methodology
 GNANADESIKAN • Methods for Statistical Data Analysis of Multivariate Observations
 GOLDBERGER • Econometric Theory
 GOLDSTEIN and DILLON • Discrete Discriminant Analysis
 GROSS and CLARK • Survival Distributions: Reliability Applications in the Biomedical Sciences
 GROSS and HARRIS • Fundamentals of Queueing Theory
 GUPTA and PANCHAPAKESAN • Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations
 GUTTMAN, WILKS, and HUNTER • Introductory Engineering Statistics, *Third Edition*
 HAHN and SHAPIRO • Statistical Models in Engineering
 HALD • Statistical Tables and Formulas
 HALD • Statistical Theory with Engineering Applications
 HAND • Discrimination and Classification
 HARTIGAN • Clustering Algorithms
 HILDEBRAND, LAING, and ROSENTHAL • Prediction Analysis of Cross Classifications
 HOAGLIN, MOSTELLER, and TUKEY • Understanding Robust and Exploratory Data Analysis
 HOEL • Elementary Statistics, *Fourth Edition*
 HOEL and JESSEN • Basic Statistics for Business and Economics, *Third Edition*
 HOLLANDER and WOLFE • Nonparametric Statistical Methods
 IMAN and CONOVER • Modern Business Statistics
 JAGERS • Branching Processes with Biological Applications
 JESSEN • Statistical Survey Techniques
 JOHNSON and KOTZ • Distributions in Statistics
 Discrete Distributions
 Continuous Univariate Distributions—1
 Continuous Univariate Distributions—2
 Continuous Multivariate Distributions
 JOHNSON and KOTZ • Urn Models and Their Application: An Approach to Modern Discrete Probability Theory
 JOHNSON and LEONE • Statistics and Experimental Design in Engineering and the Physical Sciences, Volumes I and II, *Second Edition*
 JUDGE, HILL, GRIFFITHS, LÜTKEPOHL and LEE • Introduction to the Theory and Practice of Econometrics
 JUDGE, GRIFFITHS, HILL and LEE • The Theory and Practice of Econometrics
 KALBFLEISCH and PRENTICE • The Statistical Analysis of Failure Time Data
 KEENEY and RAIFFA • Decisions with Multiple Objectives
 LAWLESS • Statistical Models and Methods for Lifetime Data
 LEAMER • Specification Searches: Ad Hoc Inference with Nonexperimental Data
 McNEIL • Interactive Data Analysis
 MANN, SCHAFER and SINGPURWALLA • Methods for Statistical Analysis of Reliability and Life Data
 MARTZ and WALLER • Bayesian Reliability Analysis

MEYER • Data Analysis for Scientists and Engineers
 MIKÉ and STANLEY • Statistics in Medical Research: Methods and
 Issues with Applications in Cancer Research
 MILLER • Survival Analysis
 MILLER, EFRON, BROWN, and MOSES • Biostatistics Casebook
 MONTGOMERY and PECK • Introduction to Linear Regression Analysis
 NELSON • Applied Life Data Analysis
 OTNES and ENOCHSON • Applied Time Series Analysis: Volume I, Basic
 Techniques
 OTNES and ENOCHSON • Digital Time Series Analysis
 POLLOCK • The Algebra of Econometrics
 PRENTER • Splines and Variational Methods
 RAO and MITRA • Generalized Inverse of Matrices and Its Applications
 RIPLEY • Spatial Statistics
 SCHUSS • Theory and Applications of Stochastic Differential Equations
 SEAL • Survival Probabilities: The Goal of Risk Theory
 SEARLE • Linear Models
 SEARLE • Matrix Algebra Useful for Statistics
 SPRINGER • The Algebra of Random Variables
 UPTON • The Analysis of Cross-Tabulated Data
 WEISBERG • Applied Linear Regression
 WHITTLE • Optimization Over Time: Dynamic Programming and
 Stochastic Control, Volume I
 WHITTLE • Optimization Under Constraints
 WILLIAMS • A Sampler on Sampling
 WONNACOTT and WONNACOTT • Econometrics, *Second Edition*
 WONNACOTT and WONNACOTT • Introductory Statistics, *Third Edition*
 WONNACOTT and WONNACOTT • Introductory Statistics for Business
 and Economics, *Second Edition*
 WONNACOTT and WONNACOTT • Regression: A Second Course in
 Statistics
 WONNACOTT and WONNACOTT • Statistics: Discovering Its Power
 ZELLNER • An Introduction to Bayesian Inference in Econometrics

Tracts on Probability and Statistics

AMBARTZUMIAN • Combinatorial Integral Geometry
 BARNDORFF-NIELSEN • Information and Exponential Families in
 Statistical Theory
 BHATTACHARYA and RAO • Normal Approximation and Asymptotic
 Expansions
 BIBBY and TOUTENBURG • Prediction and Improved Estimation in
 Linear Models
 BILLINGSLEY • Convergence of Probability Measures
 KELLY • Reversibility and Stochastic Networks
 RAKTOE, HILTON, and FEDERER • Factorial Designs
 TOUTENBURG • Prior Information in Linear Models

ACKNOWLEDGMENTS

We express our appreciation to the following individuals whose reviews of this book contributed greatly to smoothing and clarifying its presentation.

Eileen Boardman
Colorado State University
Fort Collins, Colorado

Francisco Samaniego
University of California, Davis
Davis, California

John Boyer
Kansas State University
Manhattan, Kansas

Chanchal Singh
St. Lawrence University
Canton, New York

Gerry Hobbs
West Virginia University
Morgantown, West Virginia

Vidya S. Taneja
Western Illinois University
Macomb, Illinois

Dallas E. Johnson
Kansas State University
Manhattan, Kansas

Thanks are also due to Mike Connors for drawing the cartoons. He demonstrated a great deal of innovation in bringing the cartoons to life after being presented with only a punch line and very rough sketches. Lastly, we are grateful to the professional dedication of the staff of John Wiley & Sons. In particular, the efforts of Robert Pirtle, editor for this book, Judith Watkins, and Alej Longarini are greatly appreciated. Their dedication to review and quality has led to the final form of this book.

R. L. I.
W. J. C.

ABOUT THE AUTHORS

Ronald L. Iman is a statistical consultant at Sandia National Laboratories. He holds a Ph.D. in statistics from Kansas State University. He previously worked for The Upjohn Company and has several years of university teaching experience. His research papers are in applied nonparametric statistics and sensitivity analysis techniques for use with computer models. Dr. Iman is a Fellow of the American Statistical Association and has served on the ASA Council. He is chairman of the Steering Committee for the DOE Statistical Symposium, an associate editor of *The American Statistician*, and a member of the Editorial Board of *Selected Tables in Mathematical Statistics*.

W. J. Conover is Horn Professor of Statistics and Associate Dean of Research and Graduate Programs, College of Business Administration, Texas Tech University. He holds a Ph.D. in mathematical statistics from the Catholic University of America. Professor Conover is also a consultant at Sandia National Laboratories and a visiting staff member at Los Alamos National Laboratory. An earlier book written by him is *Practical Nonparametric Statistics*, Second Edition, Wiley, 1980. His research papers include both theoretical and applied nonparametric statistics. He is a Fellow of the American Statistical Association.

PREFACE

To the Student

As you begin this course you may well wonder what lies in store for you. At this time in your life with limited job experience it is difficult to be familiar with situations where the application of statistical methods would be appropriate. Thus, the authors have tried to provide a variety of problem settings, examples, and exercises that will allow you to realize some of the potential uses of statistics and how statistics can be important to you in your future or present employment. Additionally, even though you master the material and gain an appreciation for the potential use of statistics while taking this course, it may be difficult at some time in the future to recall all of the details of some particular statistical procedure. Thus, a great deal of effort has gone into making this text readable and organizing the various methods in a form that would permit this book to serve as a ready reference at some time in the future when the need arises.

RONALD L. IMAN W. J. CONOVER

PREFACE

To the Instructor

In general the approach used in this book is conventional in many ways, but also unconventional in many other ways. Both authors have had extensive experience teaching and consulting, and this book reflects the authors' experiences, especially regarding the inadequacies and incompleteness of the presentation made in most basic statistics textbooks. In this vein, throughout the text the emphasis is on training the reader to go from a data-oriented situation to the proper statistical method. Most topics are introduced by means of a realistic problem setting, and then the method of solution unfolds in the subsequent discussion where the authors attempt to communicate how to analyze a real situation.

The problems, examples, and exercises further emphasize realistic settings in which these methods are useful. The problem settings and exercises are based on actual applications in the fields of biology, business, economics, psychology, sociology, education, home economics, agriculture, pharmacology, engineering, physical education, physiology, genetics, actuarial science, and others.

This book is intended as an introduction to modern statistical techniques in a one-semester course with the topics divided into lesson-size sections. Each section concludes with a set of exercises; review exercises occur at the end of each chapter. In all, the book contains approximately 540 exercises and numerous worked-out examples. Answers to selected exercises are given at the end of the text.

Each methodology in the book is presented in a self-contained format, complete with assumptions, explanation of notation, statements of hypotheses, and decision rules. This format is used for the convenience of the reader in referring back to the various statistical procedures. It is the authors' experience that most texts present methodology and assumptions in a helter-skelter fashion that makes it extremely difficult if not impossible for the reader to decide when to use each methodology. These same texts almost always fail to provide any guidance as to what to do if the assumptions are not satisfied. Thus, a question

of ethics arises. Should the assumptions be considered? Should texts make some attempt to deal with these issues? The authors do not feel that the reader should be left on his or her own, assuming that all is well if the mechanics are known, but that these issues are quite important and should be addressed.

In order to avoid these shortcomings, the sections presenting testing procedures clearly state the assumptions required for validity of the testing procedure. Many of these testing procedures rely on the underlying assumption of normality. This assumption is easily checked using graphs specifically developed for this book for use with the Lilliefors test for normality. Thus, the cumbersome chi-square goodness-of-fit test for normality is avoided, enabling the assumption of normality to be checked easily throughout the text. For situations where the normality assumption is not satisfied, guidance toward the appropriate nonparametric test is provided. This means that the parametric and nonparametric tests are presented side-by-side throughout the text. However, the transition to the nonparametric tests is made easy by presenting them as analogues to parametric tests. This is done by applying the parametric procedure to rank transformed data. The result is a test statistic that is functionally equivalent to the usual presentation of the nonparametric test statistic. This way of presenting nonparametric tests makes them easier to learn than when the standard method is used, and it eases the burden on instructors who may not have a strong nonparametric background. This approach was described and verified in recent research papers by the authors. A tutorial paper on the subject, with appropriate references, appeared in *The American Statistician* in August 1981.

The first chapter concentrates on sampling procedures and includes a section on sample surveys reprinted with the permission of the American Statistical Association. Due to the ease with which graphs are used, it is a temptation for many instructors to skim through techniques for displaying sample data such as occur in Chapter 2. The authors hope the instructor will not yield to this temptation since most students really do not know how to display sample data, and these techniques will be used throughout the text to aid in displaying and understanding the data as well as checking the reasonableness of many decisions. This is particularly true of scatterplots and the empirical distribution function. The graph of the empirical distribution function simplifies many of the ideas that usually give students difficulty, such as working with normal probabilities, grasping the concept of power, finding sample quantiles, and performing goodness-of-fit tests.

Years of study can be devoted to the fascinating study of probability. This book is intended to teach statistics rather than probability, so in Chapter 3 only the topics in probability that are necessary to teach statistics are included.

Descriptive sample statistics such as the sample mean, standard deviation, median, mode, and proportion and the associated population parameters are given in Chapter 4. Also included in this presentation are the empirical distribution function, quantiles, the sample correlation coefficient, and the rank correlation coefficient, since these statistics are integrated into the presentation throughout the text.

The commonly used binomial and normal distributions are introduced in

Chapter 5 along with instructions on how to use their corresponding Tables A1 and A2. These tables as well as others presented in this book are all based on cumulative distributions and therefore tie in closely with the authors' repeated emphasis on empirical and cumulative distribution functions. The table of the cumulative normal distribution has the advantage over the density function approach (i.e., area under a curve) emphasized in most statistics books because students grasp its use almost immediately—something that cannot be said of the tables based on the density function, as anyone who has spent countless hours explaining how to use it knows only too well. The Lilliefors test for normality is introduced in Chapter 5. Its use requires plotting an empirical distribution function of standardized sample values in the graphs developed for this text. The Lilliefors test provides an easy and accurate test for normality and is used throughout the text to challenge the normality assumption. Sampling distributions and the Central Limit Theorem are covered in Chapter 5 as are the closely related exponential and Poisson distributions. A companion Lilliefors test for exponential distributions is given.

Point and interval estimates are considered in Chapter 6 as well as properties of estimators. In addition to the usual confidence interval for the mean of a normal population, a procedure for finding a confidence interval for the median of a population is given. This procedure may be more appropriate in nonnormal populations. Confidence intervals are also provided for the population proportion. The framework for hypothesis testing is given in Chapter 7. Hypothesis testing is demonstrated in sections covering each of the population mean, proportion, and median. Power is explained in terms of the cumulative distribution function.

The matched pairs problem is presented in Chapter 8 from the standpoint of being a good experimental design to control unwanted variation. The paired t -test is presented as the parametric method of analysis. The normality assumption is checked with the Lilliefors graphs and the Wilcoxon signed ranks test is given as the alternative nonparametric procedure when the normality assumption is not satisfied. The Wilcoxon signed ranks test is presented as a rank transform test. That is, the parametric paired t -test is applied to rank transformed differences. The result is a test statistic that is functionally equivalent to the standard presentation of the Wilcoxon signed ranks test statistic; however, the transition is quite easy for the student as all that is needed is to apply the just learned parametric paired t -test to the ranks. The critical values are approximated by the Student's t -distribution, which is quite accurate even for small sample sizes. The basis for this approach is contained in a number of research papers by the authors.

The case of two independent samples is considered in Chapter 9. Large sample techniques are given first followed by the two-sample t -test. Assumptions of normality and equal variance are checked. Graphs of the empirical distribution functions are presented in the examples to aid the student in understanding the two-sample problem and resultant decisions. The Wilcoxon rank sum test is presented as a rank transform test obtained from the two-sample t -test.

Chi-square test procedures are considered in Chapter 10, including 2×2

and $r \times c$ contingency tables, and goodness-of-fit tests. Chapter 11 reviews correlation on raw and rank transformed data prior to the study of regression techniques in Chapter 12.

Methods for linear regression are given in Chapter 12, including the usual least-squares computation procedures as well as a study of the model for linear regression and hypothesis testing. A nonparametric test for slope based on rank correlation is also given. Methods of monotone regression are presented as linear regression computations applied to rank transformed data. These methods are appropriate for either linear regression or monotonic nonlinear regression.

The completely randomized design and associated one-way analysis of variance are presented in Chapter 13. These techniques represent a generalization of the two-sample procedure of Chapter 9. Fisher's least significant difference is given as a multiple comparisons procedure; research (including that by the authors) has shown this procedure to be robust and powerful compared to other multiple comparisons techniques. The Kruskal-Wallis test is given as the rank transform analogue to the one-way analysis of variance for the completely randomized design. Computer output is utilized for the analyses in this chapter.

Depending on the emphasis desired by the instructor, a number of sections would be omitted in a 3- or 4-semester-hour course without loss of continuity. In particular, any one of the following groups of sections could be eliminated totally independent of the elimination or inclusion of any of the other groups: (1.2, 1.3, and 1.4); (4.6, 11.2, 12.5, and 12.6); 5.5; (6.4 and 7.4); and (10.1, 10.2, and 10.3).

The authors hope you will find it a refreshing experience to teach from this text and would welcome any correspondence regarding strengths or weaknesses of the book that would affect any possible future editions.

RONALD L. IMAN

W. J. CONOVER

CONTENTS

CHAPTER 1

The Relationship Between Sampling and Statistics

PRELIMINARY REMARKS 1

Section 1.1 Populations and Samples 2

The Importance of Sampling 3

Sampling in a Small Business 3

Designing a Sample Survey 4

Why Use a Sample 5

Examples of Sampling 6

Exercises 6

Section 1.2 Type of Sampling Techniques 7

Target Populations and Sampled Populations 7

Six Types of Samples 8

Sampling Error and Nonsampling Error 12

Exercises 13

Section 1.3 What is a Survey 13

Characteristics of Surveys 13

How is a Survey Carried Out 16

Using the Results of a Survey 19

Budgeting a Survey 21

Where to Get More Information 22

Exercises 23

Section 1.4 Meaning and Role of Statistics 24

Quantitative and Qualitative Data 24

Preparing a Questionnaire 25

Summarizing Data from a Questionnaire 26

Exercises 27

Section 1.5 Review Exercises 28

Bibliography 29

CHAPTER 2

Displaying Sample Data

PRELIMINARY REMARKS 30

Section 2.1 Frequency Distributions 31

The Stem and Leaf Plot 31

Constructing a Histogram from a Stem and Leaf Plot 32

Constructing a Histogram with Equal Class Widths 33

Number of Classes 34

Exercises 36

Section 2.2 More on Frequency Distributions 38

Histograms with Unequal Class Widths 38

Frequency Polygon 40

Graphing Time Series 41

Individualized Graphical Methods 42

Exercises 44

Section 2.3 Cumulative Frequency Distributions 46

Constructing an Ogive 46

Real Number Line Plot 48

The Empirical Distribution Function 48

Exercises 49

Section 2.4 Scatterplots 50

- Univariate and Bivariate Data 50
- Scatterplot 51
- The Log Transformation as an Aid to Graphing Data 52
- Log Scatterplot 54
- The Rank Transformation 54
- Use of the Stem and Leaf Plot to Assign Ranks 55
- Rank Transform Scatterplot 55
- Exercises 57

Section 2.5 Review Exercises 60

- Bibliography 61

CHAPTER 3

Probability

PRELIMINARY REMARKS 62

Section 3.1 Sample Spaces, Events, and Probabilities 63

- Sample Variability 63
- Sample Space 63
- Probability 64
- Other Definitions of Probability 66
- Probability Models 67
- Joint Probabilities 67
- Marginal Probabilities 68
- Conditional Probability 68
- Independence 69
- Exercises 70

Section 3.2 Random Variables and Their Distributions 73

- Random Variables 73
- Discrete Random Variables 73
- Probability Function of a Discrete Random Variable 74
- The Cumulative Distribution Function 75
- Continuous Random Variables 76
- Empirical Distribution Functions Versus Cumulative Distribution Functions 78
- Histograms Versus Density Functions 80
- Independent Random Variables 81
- Exercises 82

Section 3.3 Review Exercises 83

- Bibliography 84

CHAPTER 4

Descriptive Statistics and Population Characteristics

PRELIMINARY REMARKS 85

✓ **Section 4.1 The Sample Mean 86**

- The Stem and Leaf Plot 87
- The Empirical Distribution Function 87
- The Sample Mean 88
- Calculating the Mean from Grouped Data 88
- Exercises 91

✓ **Section 4.2 The Population Mean 93**

- The Population Mean 94
- The Relationship Between the Sample Mean and the Population Mean 96
- The Population Distribution Function 97
- Exercises 98

Section 4.3 The Standard Deviation for the Sample and the Population 99

- What is the Sample Standard Deviation 99
- Calculation of the Sample Standard Deviation 100
- Back to the Test Scores 101
- A Useful Rule of Thumb 102
- A Quick Check of the Computations 103
- The Coefficient of Variation 103
- Standardized Observations, or z-Scores 104
- An Application of z-Scores 104
- More About z-Scores 104
- Calculating the Standard Deviation from Group Data 105
- The Population Standard Deviation 106
- Exercises 108

✓ **Section 4.4 Modes, Quantiles, and Proportions 109**

- The Sample Mode 109
- The Sample Median 110
- Insensitivity of the Sample Median to Outliers 110
- Skewness 111
- Sample Percentiles 112
- Other Sample Quantiles 113
- The Population Median 114
- Population Quantiles 115
- The Interquartile Range 115
- The Sample Proportion 116

Population Proportion 117
Exercises 117

✓ **Section 4.5 The Correlation Coefficient 118**

A Scatterplot 119
The Sample Correlation Coefficient 120
Possible Scatterplots for the Sample Correlation Coefficient 120
How to calculate r 121
Changing the Example 122
Sample Covariance 124
The Population Covariance 125
The Population Correlation Coefficient 125
Exercises 125

Section 4.6 Rank Correlation 126

Spearman's Rho 126
An Example With Outliers 127
How to Calculate Spearman's Rho 128
An Interpretation of Spearman's Rho 129
The Population Counterpart to Spearman's Rho 129
Exercises 129

Section 4.7 Review Exercises 130

CHAPTER 5

Some Useful Distributions

PRELIMINARY REMARKS 131

Section 5.1 The Binomial Distribution 132

A Binomial Random Variable 132
The Binomial Probability Function 134
Graph of the Binomial Probability Function 135
The Binomial Distribution Function 135
Tables for the Binomial Distribution Function 136
Population Mean 137
Population Standard Deviation 137
Estimating p in the Binomial Distribution 138
Examples of Binomial Random Variables 138
Exercises 139

✓ **Section 5.2 The Normal Distribution 141**

Histograms of the Data 141
Normal Random Variables 142
The Normal Distribution 142
Population Parameters and Their Estimates 143
The Standard Normal Distribution 144
Converting to a Standard Normal

Distribution 144
How to Find Quantiles from Table A2 145
How to Find Probabilities from Table A2 146
Exercises 149

Section 5.3 The Importance of the Normal Distribution 150

Normal Approximation to the Binomial 150
The Continuity Correction 152
Lilliefors Test for Normality 153
An Application of the Lilliefors Test 154
Exercises 155

✓ **Section 5.4 Sampling Distributions and the Central Limit Theorem 156**

The Sampling Distribution of a Statistic 156
The Sampling Distribution of the Sample Mean 157
The Distribution of \bar{X} When the Population is Normal 159
The Distribution of \bar{X} When the Population is Not Normal 159
An Illustration of the Central Limit Theorem in Action 160
The Approximate Normality of the Sample Proportion 164
Exercises 165

Section 5.5 The Exponential Distribution 166

Decisions Based on Experimental Results 166
The Exponential Random Variable 166
Lilliefors Test for Exponential Distributions 168
Estimating the Parameter of an Exponential Population 169
The Poisson Distribution 171
The Relationship Between Poisson and Exponential Distributions 172
Exercises 173

Section 5.6 Review Exercises 174

Appendix 175
The Binomial Coefficient 175
Example of the Combination Counting Formula 175
Interpolation for Normal Probabilities in Table A2 177
Bibliography 177