APPLIED STATISTICS

THIRD EDITION

JOHN NETER University of Georgia
WILLIAM WASSERMAN Syracuse University
G.A. WHITMORE McGill University

APPLIED STATISTICS

JOHN NETER University of Georgia
WILLIAM WASSERMAN Syracuse University
G.A. WHITMORE McGill University

ALLYN AND BACON, INC.
Boston London Sydney Toronto

To Dottie, Cathy, and Lonnie

Editorial-production service: Technical Texts, Inc.
Text designer: Sylvia Dovner
Cover designer: Hemenway Design Associates
Cover administrator: Linda Dickinson
Production administrator: Lorraine Perrotta
Manufacturing buyer: Ellen Glisker
Cover art: "View (Gray)" by Tetsuro Sawada. Published and distributed exclusively by Buschlen/Mowatt Fine Arts Ltd., 1467 West 47th Avenue, Vancouver, Canada V6M 2L9.

Library of Congress Cataloging-in-Publication Data

Neter, John.
Applied statistics.
Includes index.
1. Social sciences—Statistical methods.
2. Management—Statistical methods. 3. Economics—
Statistical methods. 4. Statistics. I. Wasserman,
William. II. Whitmore, G. A. III. Title.
HA29.N437 1988 519.5 86-14147
ISBN 0-205-10328-6

Printed in the United States of America.
10 9 8 7 6 5 4 3 91 90 89

TABLE C.1 Cumulative probabilities and percentiles of the standard normal distribution

(a) Cumulative probabilities

Entry is area a under the standard normal curve from $-\infty$ to $z(a)$.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

(b) Selected percentiles

Entry is $z(a)$ where $P[Z \leq z(a)] = a$.

a:	.10	.05	.025	.02	.01	.005	.001
$z(a)$:	−1.282	−1.645	−1.960	−2.054	−2.326	−2.576	−3.090
a:	.90	.95	.975	.98	.99	.995	.999
$z(a)$:	1.282	1.645	1.960	2.054	2.326	2.576	3.090

EXAMPLE: $P(Z \leq 1.96) = 0.9750$ so $z(0.9750) = 1.96$.

TEXT REFERENCE: Use of this table is discussed on pp. 214–220.

Preface

Applied Statistics, Third Edition, is written for students taking basic statistics courses in management, economics, and other social sciences, as well as for people already engaged in these fields who desire an introduction to statistical methods and their application. Our aim has been to offer a balanced presentation of fundamental statistical concepts and methods, along with practical advice on their effective application to real-world problems. Conceptual rigor is not sacrificed, however. The conceptual foundation of each subject is developed carefully up to that level needed for prudent and beneficial use of statistical methods in practice.

The third edition differs from the second in a number of important ways. We have extensively revised the chapters on statistical testing to unify the treatment of statistical tests throughout the book. As a result, all statistical testing is now carried out by means of standardized test statistics. In addition, we have expanded the coverage of *P*-values and integrated their use into the discussion of statistical testing. We have also made greater use of computer outputs and plots and have further emphasized the uses of statistical computer packages for data management and analysis.

The updating in the third edition to reflect current needs and practices has led to the addition or expansion of several important topics. The treatment of statistical quality control has been expanded by considering several additional types of control charts and their uses. In regression analysis, we have enlarged the treatment of indicator variables, curvilinear models, regression models with interaction effects, and statistical tests, and we have added a section on the use of the all-possible-regressions method for selecting the regression model. In time series analysis, we now include exponential smoothing for time series with trend and with trend and seasonal components. We also discuss the Durbin-Watson test and have added an introduction to ARIMA models.

In addition, we have greatly expanded the number of problems at the ends of the chapters and have added three more data sets in the appendix. We have also made many revisions in the text to improve the clarity of the presentation.

As in the second edition of this text, explanation of important principles and concepts always precedes discussion of detailed statistical procedures. Each new technique is illustrated by one or more examples drawn from real life. In addition, many case applications are presented, so that statistical concepts and methods can be understood in the context of their actual use. These case applications demonstrate the usefulness of statistical methods and statistically rigorous thinking in management, economics, and other social sciences.

The topical coverage of the text is broad, but it is in no sense a miscellany of statistical tools. Topics have been selected on the basis of two criteria: (1) their importance in real applications of statistics, and (2) their contribution to the development and understanding of material presented subsequently in the book. Units, chapters, and

sections have been organized and sequenced in a way that always keeps the main track of the subject clear to the reader. Technical notes and secondary observations are presented in *Comments* sections. Major topics of interest that are not essential to the main development of statistical ideas are presented in *Optional Topic* sections and may be included in the basic statistics course at the instructor's discretion to meet particular course objectives. These sections always appear at the ends of chapters and can be omitted without loss of continuity. Extensive use is made of figures and tables. Important definitions and formulas are set out in a distinctive manner to aid in learning and to facilitate ready reference.

Use of this book requires knowledge of college-entrance algebra but not calculus. We believe that fundamental statistical ideas can be introduced with little mathematics, and that this approach in an introductory text leads to a fuller appreciation of the uses of statistical methods than a more mathematical approach. Mathematical demonstrations are included where they make a significant contribution to the reader's understanding of the subject. All such demonstrations are self-contained and illustrated by numerical examples. Occasionally, a mathematical demonstration is presented that requires calculus. These sections are marked "calculus needed" and may be omitted without loss of continuity. Those readers who will continue their study of statistics and, therefore, must become more familiar with the mathematical basis of the subject will find that this book provides a strong foundation upon which they can build.

We assist the reader with the mathematical and computational aspects of the subject in a number of ways. Notation is used only where needed, and then a uniform and straightforward notational system is employed throughout the text. A mathematical review is included in Appendix A for readers desiring a brief summary of: (1) summation notation, (2) rules for exponents and logarithms, (3) set notation, operations, and rules, and (4) the basics of permutations and combinations. Because statistical computer packages play a major role in real applications of statistics, we have sought to familiarize readers with ways in which they assist in statistical analyses (such as regression) and in data handling (such as tabulations). Computer printout is presented throughout the text to illustrate its form and its usefulness in various practical contexts.

Large numbers of problems and questions are given at the ends of chapters. The *Problems* sections contain basic problems and drill questions. The *Exercises* sections present questions dealing with mathematical concepts and extensions of ideas developed in the chapters. Finally, the *Studies* sections contain major, comprehensive problems and case studies. The problems, exercises, and studies each follow the sequence of topics in the text. Numerical answers for selected problems (identified by asterisks) are given at the end of the book to facilitate immediate checking by the reader. The problems, exercises, and studies have been designed to assist the reader's understanding of concepts and to enable him or her to obtain experience in applying statistical techniques in practical situations and in interpreting results of statistical investigations. Many different types of applications are employed in order to expose the reader to the rich and varied settings in which statistics is applied in real life. A number of the problems and studies draw on the data sets in Appendix D.

Several teaching and learning aids, specifically keyed to this edition of *Applied Statistics*, are available. The aids include: (1) an Instructor's Manual containing fully

worked numerical and discussion answers for all problems, exercises, and studies; (2) a self-learning Study Guide for students, prepared by Professor Kenneth C. Schneider; (3) a Test Manual for instructors containing a large collection of multiple-choice test questions, also prepared by Professor Schneider; and (4) a computer diskette containing the four data sets in Appendix D and other data sets of the text, to support computer-aided learning and instruction.

The book is divided into seven units, as shown in Figure I.

Unit One, Data (Chapters 1–3). However sophisticated a statistical procedure, its successful application depends on data. This unit is concerned with the acquisition, classification, and summarization of data. Appropriately, the unit emphasizes large data sets, computerized data handling, and exploratory data analysis. Since a large portion of day-to-day dealings with statistics for many persons is concerned with raw data and their examination by descriptive (as opposed to inferential) means, the chapters in this unit are designed to prepare the reader for these common encounters. The important role of computer data-processing packages, data banks, and retrieval systems is also recognized in this unit, and the groundwork is laid for data handling in subsequent chapters.

Unit Two, Probability (Chapters 4–7 and Appendix B). Probability theory is the foundation of statistical inference, so this unit comes next in the book. The chapters of this unit contain an integrated presentation of basic probability concepts, random variables, and common probability distributions utilized in applications. The last topic is covered in two chapters, the first presenting common discrete distributions and the second, common continuous distributions. Appendix B covers the χ^2, t, and F distributions and explains their relationships to one another and to the standard normal distribution. These three distributions are presented in the appendix as a handy reference so that they can be accessed readily from any place in the main text where they might be encountered for the first time in a course. Extensive probability tables, prepared in a consistent fashion, are provided in Appendix C to support this and later units.

Unit Three, Estimation and Testing—I (Chapters 8–14). This unit opens with two chapters discussing sampling and the sampling distribution of $\overline{X}$. In contrast to conventional theoretical explanations, an empirical demonstration of the sampling distribution of $\overline{X}$ and the central limit theorem is presented first, using simulated, repeated sampling from an actual population. Then the key theorems are stated and explained. This approach allows the reader to anticipate the theoretical results on the basis of empirical experience, and it should lead to greater understanding of these conceptually important topics. Next, estimation and testing for a population mean are presented, followed by a discussion of the sampling distribution of $\overline{p}$ and inference procedures for a population proportion. Inference procedures in comparative studies involving two population means or proportions and inferences for population variances are discussed in the next-to-last chapter of this unit. The final chapter considers statistical quality control procedures, as well as sampling of finite populations and various sampling procedures other than simple random sampling.

Unit Four, Estimation and Testing—II (Chapters 15–17). In the first chapter of this second unit on estimation and testing, nonparametric procedures that have found extensive use are discussed. The approach is not one of presenting a grab bag of statis-

tical tools — rather, procedures are presented which match those discussed in the previous unit, but which do not depend on large samples or restrictive distributional assumptions. Goodness-of-fit procedures and techniques for multinomial populations are taken up in the next two chapters. Separation of these topics into two chapters emphasizes the separate classes of problems involved and is a departure from the usual textbook coverage.

Unit Five, Linear Statistical Models (Chapters 18–21). This unit meets a pressing need for a thorough treatment of regression and analysis of variance to prepare the reader for the extensive use of these techniques in management, economics, and social sciences in general. The topics of regression and analysis of variance are handled in a unified fashion in these chapters. Computer-assisted analysis is stressed, as is the facility to interpret standard printout from a computer regression package. Also stressed are practical aspects in using linear statistical models, such as study of the aptness of the model employed.

Unit Six, Bayesian Decision Making (Chapters 22 and 23). This unit presents a discussion of Bayesian decision analysis, showing how the decision-theoretic aspects of a statistical problem can be handled formally. The relation between statistical decision making and statistical testing is made clear by our initial emphasis on normal-form analysis and later presentation of extensive-form analysis.

Unit Seven, Time Series Analysis and Index Numbers (Chapters 24–26). A compact treatment of the classical time series model is presented in the first chapter of this unit. The second chapter is devoted to exponential smoothing, regression time series models, and an introduction to ARIMA models. The emphasis again is on concepts and procedures needed for effective application. In this same spirit a third chapter is devoted to price and quantity indexes, which play an important role today for many administrators, economists, and social scientists.

This book can be used for a wide variety of one- or two-quarter or one- or two-semester courses. Figure I shows the logical interdependencies of the units and their chapters, with the arrows indicating prerequisite chapters. As can be seen, various chapter sequences may be used, depending on the length of the course and the emphasis desired. Here are four examples:

1. A course covering descriptive statistics and the basics of statistical inference might include data (Unit One), probability (Unit Two), inferences for population mean and proportion (Chapters 8–12), simple linear regression (Chapters 18 and 19), and the descriptive portions of time series analysis and index numbers (Chapters 24 and 26).
2. A course on inferential statistics emphasizing linear statistical models might include data (Unit One), probability (Unit Two), inferences for population mean (Chapters 8–11), linear statistical models (Unit Five), and regression models for time series analysis (Chapter 25).
3. A course covering both parametric and nonparametric inferences might include data (Unit One), probability (Unit Two), parametric inferences (Unit Three), nonparametric inferences (Unit Four), and as much of linear statistical models (Unit Five) as time permits.

FIGURE 1 Structure of *Applied Statistics*. Chart shows chapter interdependencies—arrows point to prerequisite chapters.

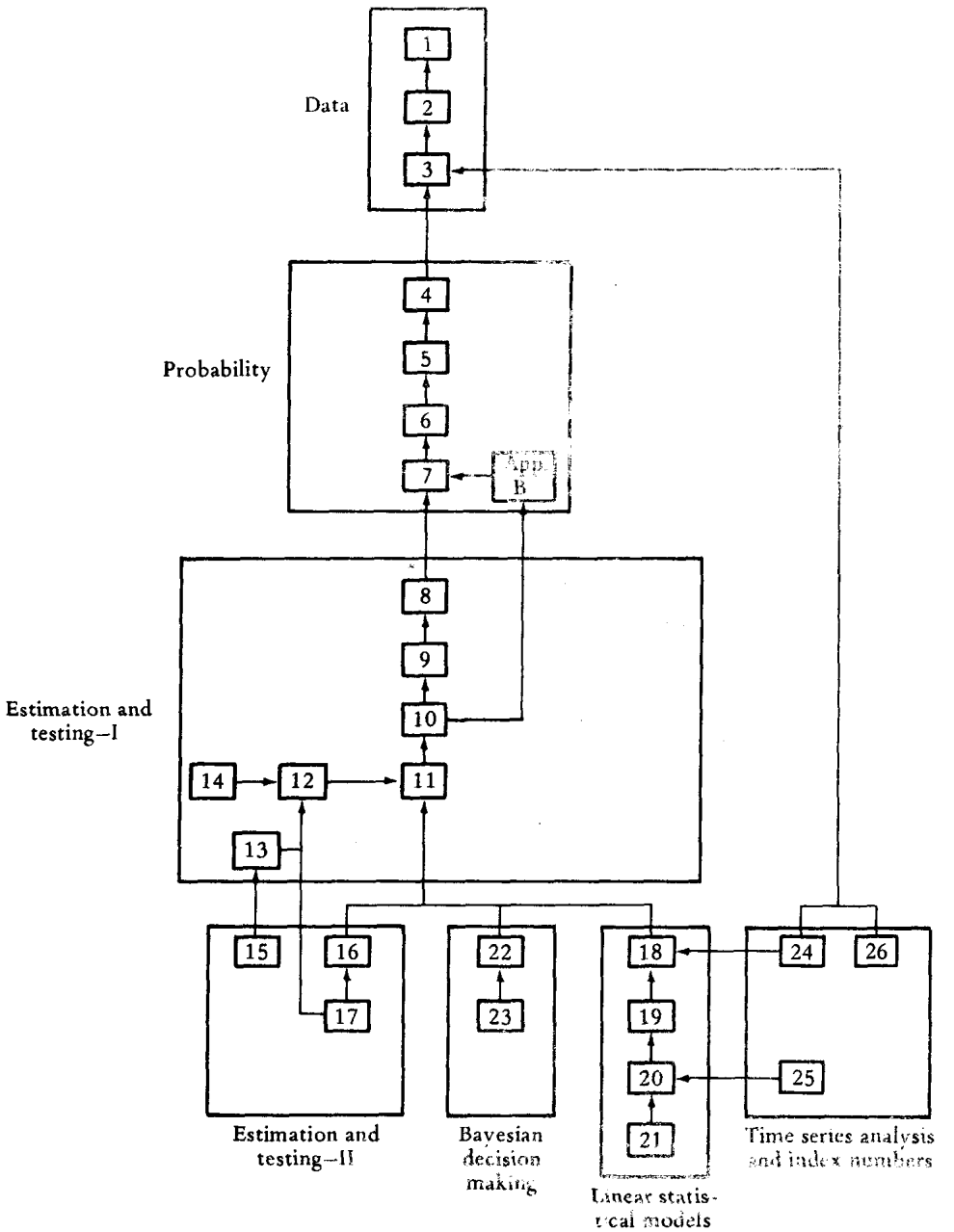

FIGURE I Structure of *Applied Statistics* (continued)

Unit One: Data
1. Data Acquisition
2. Data Analysis I: Classification and Distribution Patterns
3. Data Analysis II: Summary Measures

Unit Two: Probability
4. Basic Probability Concepts
5. Random Variables
6. Common Discrete Probability Distributions
7. Common Continuous Probability Distributions

Unit Three: Estimation and Testing—I
8. Statistical Sampling
9. Sampling Distribution of $\bar{X}$
10. Estimation of Population Mean
11. Tests for Population Mean
12. Inferences for Population Proportion
13. Comparisons of Two Populations and Other Inferences
14. Quality Control and Other Applications of Sampling

Unit Four: Estimation and Testing—II
15. Nonparametric Procedures
16. Goodness of Fit
17. Multinomial Populations

Unit Five: Linear Statistical Models
18. Simple Linear Regression
19. Inferences in Simple Linear Regression
20. Multiple Regression
21. Analysis of Variance

Unit Six: Bayesian Decision Making
22. Bayesian Decision Making I: No Sample Information
23. Bayesian Decision Making II: Sample Information

Unit Seven: Time Series Analysis and Index Numbers
24. Time Series and Forecasting I: Classical Methods
25. Time Series and Forecasting II: Exponential Smoothing and Regression Methods
26. Price and Quantity Indexes

Appendices
A. Mathematical Review
B. Chi-Square, t, and F Distributions
C. Tables
D. Data Sets

4. A course emphasizing decision theory might include data (Unit One), probability (Unit Two), the major elements of statistical inference (Chapters 8–13), Bayesian decision making (Unit Six), and other topics as time permits.

We are greatly indebted to many individuals and organizations who have helped us in the preparation of this book. Our sincere thanks go to all who have provided us with case materials and illustrations that demonstrate the usefulness of statistical methods in management, economics, and other social sciences. Many persons, including colleagues and reviewers, have made helpful suggestions on the manuscript and otherwise assisted us, for which we are most grateful. We particularly wish to thank for their help Robert F. Berner, R. V. Erickson, Edgar Hickman, Oswald Honkalehto, H. K. Hsieh, Allan Humphrey, Raj Jaganathan, William Meeker, Robert Norland, Jr., Thomas Pray, Thomas Rothrock, Barbara Rufle, J. Michael Ryan, Kenneth C. Schneider, Randolph Shen, Erland Sorensen, Stephen Vardeman, Dean Wichern, and Morty Yalovsky.

We are also indebted to our respective universities for providing an environment in which the undertaking could be brought to fruition. A number of people were most helpful in working with us in preparing and checking the manuscript. We wish to express our appreciation to Brenda Adams, Karen Robertson, and Johanne Thiffault for their valuable assistance. The staffs of Technical Texts, Inc. and Allyn and Bacon, Inc. helped us in many ways. Of course, we thank our families for their patience and encouragement while we completed this book.

Introduction

What is Statistics?

Almost everyone uses statistics and is affected by applications of statistics. The word *statistics* refers in common usage to numerical data. Vital statistics, for example, are numerical data on births, deaths, marriages, divorces, and communicable diseases; business and economic statistics are numerical data on employment, production, prices, and sales; social statistics are numerical data on housing, delinquency and crime, education, and social security and welfare.

Statistics has an additional meaning that is more specialized. In this second sense, *statistics* refers to the methodology for the collection, presentation, and analysis of data, and for the uses of such data. Unless data are accurate, properly presented, and correctly analyzed, they may be dangerously misleading. Since we all are "consumers" of statistics, it is important for all of us, not only professional statisticians, to acquire some knowledge of statistical methodology.

The word *statistician* also has several meanings. It can refer to: (1) a person who performs routine operations with statistical data; or (2) an analyst who is highly trained in statistical methodology and uses this methodology in the collection and interpretation of data; or finally, (3) an applied mathematician who utilizes advanced mathematics in the development of new statistical methods. Statisticians are needed in all these capacities in order to make statistical data most useful.

The Expanding Role of Statistics

Statistical data have been used for many centuries by governments as an aid in administration. In antiquity, statistics were compiled to ascertain the number of citizens liable for military service and taxation. After the Middle Ages, governments in Western Europe were interested in vital statistics because of the widespread fear of devastating epidemics and the belief that population size could affect political and military power. As a result, data were compiled from registrations of christenings, marriages, and burials. In the sixteenth through eighteenth centuries, when mercantilistic aspirations set nation-states in search of economic power for political purposes, data began to be collected on such economic subjects as foreign trade, manufacturing, and food supply.

Today, data are collected, classified, stored, and retrieved in diverse and comprehensive information systems that supply individuals and organizations with the statistical intelligence required to carry out their activities. The expansion in the collection, transmission, storage, and retrieval of statistical data, facilitated by computers, has been accompanied by the rapid development of statistical methodology and data analysis.

Statistical concepts have exercised a profound influence in almost every field of human activity and have been incorporated into the basic principles of such sciences as

physics, genetics, meteorology, and economics. Statistical methods have been used to improve agricultural products, to design space equipment, to plan traffic control, to forecast epidemics, and to attain better management in business and in government. Students of the natural and social sciences study statistics to become better scientists; students of economics study statistics to become better economists; students of administration study statistics to become more effective administrators.

Uses of Statistics

Some knowledge of statistics is essential today for people pursuing careers in almost every area of industry, government, public service, or the professions. Not only are more comprehensive networks of data available to serve as a basis for drawing valid conclusions and making decisions, but the purpose in assembling the data has shifted from record keeping to evaluation and action, based on timely information. Until recently, statistics were collected primarily as a record of past events. Although such statistics were analyzed to gain insights into current problems, the emphasis was essentially upon the past. At the present time, the collection of numerical information for the record still takes place, but because of the needs for improved planning and control, data-collection systems and data repositories have been designed to provide data that are as up-to-date as possible. Statistical analysis, in turn, has become chiefly concerned with the present and the future, rather than the past.

The increasing uses of statistics are part of the trend toward basing evaluations and making decisions on the most objective and scientific foundations possible. Modern organizations are becoming more dependent on statistical data to obtain factual information about their internal operations and their social, economic, and ecological surroundings. Statistical data are concise, specific, capable of being analyzed objectively with powerful formal procedures, and well suited for making comparisons. Hence, they are especially useful in such key organizational functions as choosing among alternatives, setting goals, evaluating performance, measuring progress, and locating weaknesses.

Uses of statistics have been greatly facilitated by the extensive array of statistical computer software packages now available. These packages not only expedite the application of statistical methods and the handling of data at lower cost, but they also permit the use of more powerful and complex statistical techniques. Statistical procedures and models are now integral parts of many management computer systems, such as decision support systems and expert systems. This expanded computerization of statistical methods increases the need for users of statistics to have an adequate understanding of statistical concepts and methods, making the subject of statistics an even more essential component of the professional education of businesspersons, economists, and administrators.

Contents