Applied Statistical Methods

Irving W. Burr

APPLIED STATISTICAL METHODS

Irving W. Burr

PURDUE UNIVERSITY

ACADEMIC PRESS New York San Francisco London

A Subsidiary of Harcourt Brace Jovanovich, Publishers

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1

Library of Congress Cataloging in Publication Data

Burr, Irving Wingate, Date
Applied statistical methods.

(Operations research and industrial engineering)
Includes bibliographical references.
1. Mathematical statistics. 2. Probabilities.
I. Title
QA276.B863 519.5 73-2077
ISBN 0-12-146150-5

AMS (MOS) 1970 Subject Classifications: 62N99, 62P99, 62N10

PRINTED IN THE UNITED STATES OF AMERICA

PREFACE

The purpose of this book is to provide the student with the fundamental understanding of statistical methods necessary to deal with a wide variety of practical problems. Two points have been emphasized in developing the text. First, the selection of topics for inclusion was based upon breadth of applicability to practical problems. Second, the topics covered are presented in a manner which stresses clarity of understanding, interpretation, and method of application.

The text is intended primarily for upper division undergraduate and graduate students in the mathematical, physical, and engineering sciences, and in economics, business, and related areas. In addition, researchers and line personnel in industry and government will find this book useful in self-study. The background required for complete understanding of the entire text is limited to an elementary knowledge of integral and differential calculus. However, much of the material covered does not require calculus and may therefore be pursued by students with a background limited to college algebra.

In general, results given are presented without proof or derivation. In some cases, however, proper comprehension requires a proof or derivation of a particular result. The development and discussion of the statistical methods treated is accompanied by an extensive set of examples to demonstrate the application of the material to practical problems. In many instances these examples have been taken directly from industrial problems encountered by the author. In addition, each chapter includes a liberal set of problems for student assignment. Answers for odd-numbered problems are provided; solutions for even-numbered problems are available upon request from the Publisher.

The text is intended for use in a one-semester course or a sequence of two-quarter courses in statistical methods. The variety of material presented should provide the instructor with a reasonable degree of flexibility in the choice of topics to be presented. Several sections within a number of chapters are starred. While these sections are of

significant importance in specific areas of application, they may be omitted without loss of continuity.

In Chapter 1 the importance of statistical analysis is illustrated, and this provides motivation for the study of the material in the remainder of the text. Chapters 2 and 3 introduce the student to methods of data summarization, including frequency distributions, cumulative frequency distributions, and measures of central tendency and variability. In Chapter 4 the fundamental principles of probability are discussed. The concepts of sample spaces, outcomes, events, probability, independence of events, and random variables are introduced. Chapters 5 and 6 treat discrete and continuous random variables and their characterization. The distributions of several important statistics such as the sample mean, the sample variance, the difference of sample means, and the ratio of sample variances are presented in Chapter 7.

In Chapter 8 the student is introduced to statistical tests of hypotheses, stressing the significance of sample size, type I and type II errors, and the design of statistical tests. Tests of hypotheses concerning a single mean and a single variance are then presented with an interpretation of the results of these tests. Two sample tests of hypotheses are presented in Chapter 9, including tests of hypotheses for the comparison of two means and two variances. Point and interval estimation as well as a discussion of the desirable properties of point estimates are treated in Chapter 10. Simple linear regression is the topic of Chapter 11. Here the method of least squares is discussed and illustrated. In addition, tests of hypotheses about and confidence intervals for sample estimates of the parameters of a simple linear regression model are presented. The elements of single- and two-factor analysis of variance are discussed in Chapter 12. The design of analysis of variance experiments is discussed along with tests for homogeneity of variance and multiple contrasts. The results of Chapter 11 are extended to multivariate regression analysis in Chapter 13. Relevant significance tests and matrix methods in multiple regression analysis are also presented. Finally, in Chapter 14, chi-square tests for goodness of fit and data categorized in contingency tables are treated and illustrated.

The author would much appreciate receipt of any errors or inaccuracies found in the text, or suggestions for future improvement.

ACKNOWLEDGMENTS

The author is much indebted to his colleague Professor Virgil L. Anderson who read the original manuscript and made helpful suggestions. He is also indebted to Professor J. W. Schmidt of Virginia Polytechnic Institute who reviewed the manuscript, making useful suggestions. Thanks are also due Mrs. Dorothy Penner, whose accurate typing was a great help.

The tables at the back of the book have been either calculated for this text (and checked against other sources) or else copied or abstracted, by kind permission, from Professor Donald B. Owen's "Handbook of Statistical Tables," published by Addison-Wesley Publishing Co., to whom thanks are tendered. In using rounding off in the few cases where the best last digit could not be ascertained, the larger digit has been used. Table VII has been reproduced by the kind permission of Dr. H. L. Harter of the Aerospace Research Laboratories, Wright-Patterson Air Force Base, from his "Order Statistics and Their Use in Testing and Estimation," Volume 1.

Special thanks go to my wife Elsie, ever a source of encouragement.

TABLE OF CONTENTS

Chapter 4 **Some Elementary Probability**

Chapter 5 **Some Discrete Probability Distributions**

Chapt **Estimation of Population Characteristics**

Chapter 11 Simple Regression

Chapter 13 **Multiple Regression**

Chapter 14 **Goodness of Fit Tests, Contingency Tables**

Chapter 1

INTRODUCTION

1.1. WHY STATISTICAL METHODS?

The word "statistics" has two meanings, both of which are of importance to us. The first of these is as a plural word, meaning facts or observations, especially numerical results or data. The second meaning is as a singular word, denoting a branch of knowledge.

Whenever outcomes or results vary unpredictably or in random fashion, we have statistical data, whether we like it or not, whether we know it or not. If we have statistical data, we ought to be using appropriate statistical methods to analyze them, so as to be objective and to avoid the many traps, pitfalls, and misinterpretations which lie in wait for the unwary. "Statistics" may in fact be defined as the science of proper collection, condensation, analysis, and interpretation of data which *vary* (from time to time, place to place, trial to trial, person to person, material to material, etc.). The objects of statistics are estimation, prediction, and decision-making from analysis of properly collected data which vary. In doing so some risks of wrong decisions must be accepted, but statistical methods enable us to measure such risks and to control them within economic limits. Basically there are two errors we can make from data: (1) fail to learn something we should; (2) "learn" something or take some action not justified by the data. Both errors are being made continually.

Now why is "statistics" an important discipline? The first reason is, as we have just seen, that of making sound use of data for decision-making, prediction, estimation, and so on, in our job or profession. The second reason is because of the universality of variation in all areas.

It is not at all difficult to show the need for statistical reasoning in each and every field of concentration in a large university, not only for those experimenting in the field but also for anyone reading about the field in question. Professional books and journals carry much statistical material. The same thing is true of our lives as citizens. We are continually bombarded by data and statements therefrom, in newspapers, magazines, radio and television, public speeches, conversation, advertisements, and in politics. We need statistical sense to be good citizens.

In essence statistics studies the relationships between populations and samples. Of course, if we have the entire population at hand and it is not too large a collection, we have a purely descriptive job at hand. (Statistical methods of condensation, averaging, etc., can help in this description.) But in general we can only obtain a sample, and from this we will wish to draw inferences about what the population is like. Or we may specify a hypothetical population and ask whether our observed sample could readily have come from such a population. In both cases we have the problem of the relation between population and sample.

Statistics provides us with a set of mathematical models of patterns of variation, which we "hold up to nature." The more accurately our model reflects nature, the safer are our conclusions.

In this book is presented a kit of statistical tools of remarkable versatility and utility for the analysis of data. These tools can either be applied directly or with a little ingenuity or imagination be adapted to the problem at hand. Regarding the latter it may even be said that there is always some approximation involved in attributing a mathematical or probabilistic model to nature. In fact one friend of the author's said he had never seen any *perfectly* pure application of statistical methods, while another used to give a course entitled "Messy Data." Nevertheless our applications are in general "good enough," if not absolutely perfect.

1.2. ADVICE TO THE STUDENT

Statistical reasoning may well seem rather foreign to the student if this book is his first contact with the field of statistics. But as the relations between population and sample become clear, the approaches will begin to make sense and to seem entirely natural.

The student is urged to work many numerical problems, for this helps him get the "feel" of the methods. Indeed the great early leader in statistics, Sir Ronald Fisher, is reputed to have said that all he ever learned was through the calculating machine. Nowadays, with such fantastic calculational potential in digital computers, the student may