Applied Statistics for Food and Agricultural Scientists

Subhash C. Puri
Canada Department of Agriculture
Ottawa, Ontario, Canada

Kenneth Mullen
University of Guelph
Guelph, Ontario, Canada

G. K. Hall Medical Publishers,
Boston, Massachusetts

G. K. Hall Medical Publishers
70 Lincoln Street
Boston, Massachusetts 02111

80 81 82 / 4 3 2 1

Bibliography
Includes index.
1. Mathematical statistics. I. Mullen, Kenneth, joint author. II. Title.
QA276.P85 5'9.5 79-15151
ISBN 0-8161-2161-3

Applied Statistics for Food and Agricultural Scientists

OTHER TITLES:

Statistical Quality Control for Food and Agricultural Scientists
Subhash C. Puri,
Daniel Ennis,
Kenneth Mullen

Dedicated to our teachers who taught us and encouraged us to think:

Dean W. F. Forbes, M. L. Tiku, K. R. Shah
Dean C. H. Cramer, Boyd Harshbarger, Arthur Klingel

Preface

There is no dearth of good textbooks on statistics, both applied and mathematical. In fact, few subjects have experienced such a bibliographic explosion. Notwithstanding, we have ventured into writing another textbook on applied statistics which, we hope, has some unique features. First, we have attempted to cover many different topics under one cover. Many of our colleagues working in the field of food and agricultural research have spoken of the inconvenience of searching through several textbooks to locate essential statistical methods. We hope this text will alleviate this problem and provide the researcher with a basic set of most commonly used statistical techniques. Second, we have directed our attention to the needs of food and agricultural scientists for whom few applied statistics books are available. Last, we have endeavored to keep the book as applied as possible, avoiding mathematical jargon, formulas, or their derivations. It is our belief that a knowledge of high school mathematics should be sufficient for understanding and using the statistical methods in this book. The book is intended as a working manual. Moreover, we hope that the simplicity and readability, coupled with a coverage of a variety of topics, will extend the usefulness of the book beyond the confines of food and agricultural sciences. In this sense, the book may be regarded as a general applied statistics text.

College and university students and teachers, research scientists, and those engaged in food industries and related organizations will

find this book useful. In fact, it should be of interest to anyone wishing to employ statistical techniques to analyze data.

Another novel feature is the sequence of topic presentation. It seems logical to start with scientists who have a set of data they wish to analyze. They are encouraged to use a given sequence for analyzing data. First, the data are organized and tabulated to make them presentable. Then the data are graphed in order to study their distribution pattern. Next, the location and dispersion parameters are calculated. It is desirable to understand the methods of analyzing the data. The validity of the relevant parameters and their applicability within defined limits are assessed. Interpretation of results and report writing follow as a final phase of data analysis.

We wish to thank all those who helped us, either directly or indirectly, in writing this text. First, we are indebted to those scholars in the past and present whose abilities and perseverance have made possible the development of the available techniques and ideas. The authors' special thanks are due to Dean W. F. Forbes of the Faculty of Mathematics, University of Waterloo, Ontario, Canada, and Dr. A. R. Sen of Canadian Wildlife Service, Canadian Department of Fisheries and Environment, whose encouragement, comments, and suggestions had a great influence on the writing of the book. We are particularly thankful to Dr. Steve Brown of the University of Waterloo, A. K. Sahney of the Ministry of Agriculture, Government of Jamaica, H. S. Bajaj of Canadian Department of Indian and Northern Affairs, and B. S. Nagpal of Canadian Department of Fisheries and Environment, for their helpful suggestions and comments. We are also pleased to acknowledge the contributions of the editorial staff of G. K. Hall & Company. The authors alone, of course, are responsible for any errors or omissions. Last, but not the least, we thank Leah Holmes and Linda Hohner for their careful and patient typing of the manuscript, and our families for their patience, encouragement, and endurance.

S. C. Puri
K. Mullen

Applied Statistics for Food and Agricultural Scientists

CONTENTS

1.

Introduction

Statistics deals with the development and application of methods and techniques for the collection, presentation, analysis, and interpretation of data. We collect data to obtain quantitative information about past results or to present observed phenomena and make predictions about their future behavior. The usefulness of statistics lies in its applicability to almost all fields of scientific study. Statistics can be broadly classified into two areas: theoretical statistics and applied statistics. Theoretical statistics provides a mathematical basis, whereas applied statistics deals with the applications of statistical methods for analyzing data.

In any scientific investigation, a food or agricultural scientist is generally faced with such problems as how to design an experiment efficiently to obtain valid data, how to analyze data properly through statistical methods, how to make relevant interpretations, and, finally, how to gauge the reliability of the results. The question of reliability arises because the investigator is generally dealing with a single sample of results and often wonders if repeated experiments would yield consistent results. Would all similar experiments provide reasonably consistent answers? This book offers guidelines for rational and orderly solutions to these problems. In this chapter we give some basic definitions and concepts, while the following chapters are devoted to applied statistical techniques essential for analyzing data.

1.1 Population and Sample

A *population* (sometimes called a *universe*) encompasses all possible measurements of the particular characteristic under consideration. A population can be finite or infinite, real or hypothetical. We may speak of the population of human beings in Canada, the height of men of a certain age group in a locality, the number of apples in an orchard, the number of grains of sand in the world, and so on.

A population is described by its *parameters*. If the population being studied is the height of all Canadian males over 21 years of age, then some parameters are the average height of the population, the range of heights of the population, or the most frequently occurring height of the population.

A *sample* is a subset of the population. It is made up of a selected number of individuals, each of which is a member of the population. Clearly, a sample will not tell us everything about the parent population from which it has been drawn, but an attempt is made to obtain information, as precisely as possible, about the parameters of the population. To do this requires good sampling procedure. An important type of sample in statistics is a *random sample*. A random sample is one that has been selected by a random process and is such that each measurement in the population has an equal and independent chance of being included in the sample. The theory of sampling is developed in a later section.

Whereas a population is described by its parameters, a sample is described by its *statistics*. If, from the population of heights of Canadian males over 21 years old, a sample of 1000 heights is obtained, some possible statistics are the average of the sample, the range of the sample, or the most frequently occurring value in the sample.

1.2 Continuous and Discontinuous Data

A measurable characteristic that changes from one object or individual to another is called a *variable*. Individual measurements of a variable are called *variates*. If the variable can assume only one value, it is called a *constant*.