
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

Copyright © 1980 by G. K. Hall & Co.

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

All rights, including that of translation into other languages, reserved. Photomechanical reproduction (photocopy, microcopy) of this book or parts thereof without special permission of the publisher is prohibited.

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

Preface

Chapter 1 Introduction 1

- 1.1 Population and Sample 2
- 1.2 Continuous and Discontinuous Data 2
- 1.3 Scales of Measurement 3
 - 1.3.1 Nominal 3
 - 1.3.2 Ordinal or Ranking 3
 - 1.3.3 Interval 4
 - 1.3.4 Ratio 4
- 1.4 Summation Notation 4

Chapter 2 Organization and Presentation of Data 8

- 2.1 Frequency Table 8
- 2.2 Graphical Representation of Data 9
 - 2.2.1 Histogram 9
 - 2.2.2 Frequency Polygon 9
 - 2.2.3 Cumulative Frequency Polygon or Ogive 12
 - 2.2.4 Frequency Curve 12
 - 2.2.5 Bar Graph 14
 - 2.2.6 Pie Chart 15
- 2.3 Cheating by Charting 17

Chapter 3	Statistical Measures 21
3.1	Measures of Central Tendency 21
3.1.1	The Mean 21
3.1.2	The Median 24
3.1.3	The Mode 25
3.1.4	The Geometric Mean 26
3.1.5	The Harmonic Mean 27
3.2	Advantages and Disadvantages of Different Averages 27
3.3	Measures of Dispersion 28
3.3.1	The Range 28
3.3.2	The Mean Deviation or Average Deviation 28
3.3.3	The Variance and the Standard Deviation 29
3.3.4	The Coefficient of Variation 32
Chapter 4	Probability 38
4.1	Classical Probability 39
4.2	Probability as a Relative Frequency 39
4.3	Some Laws of Probability 40
4.4	Permutations and Combinations 43
4.5	Random Variables and Distributions 45
Chapter 5	Four Important Distributions 49
5.1	The Hypergeometric Distribution 49
5.2	The Binomial Distribution 50
5.3	The Poisson Distribution 52
5.4	The Normal Distribution 53
5.5	The Normal Approximation to the Binomial Distribution 59
Chapter 6	Statistical Estimation 64
6.1	The Distribution of the Sample Mean and the Sample Variance 64
6.2	t-Distribution 65
6.3	Chi-Square Distribution 68
6.4	Parameter Estimation and Confidence Intervals 70

- 6.4.1 Confidence Intervals for the Mean of a Normal Distribution 71
- 6.4.2 Confidence Interval for the Variance and Standard Deviation of a Normal Distribution 73
- 6.4.3 Confidence Interval for a Proportion 74
- 6.4.4 Confidence Interval for the Difference of Two Means 76

Chapter 7 Hypothesis Testing 83

- 7.1 Null and Alternate Hypotheses 83
- 7.2 Type I and Type II Errors 84
- 7.3 Types of Tests 87
- 7.3.1 One- and Two-Tailed Tests 87
- 7.3.2 Testing the Mean of a Normal Distribution 89
- 7.4 Comparison of Hypothesis Testing and Confidence Intervals 91
- 7.5 Testing the Variance and Standard Deviation of a Normal Distribution 92
- 7.6 Testing the Difference of Two Means 93
- 7.7 Testing the Difference of Two Means When the Observations Are Paired 95
- 7.8 Testing a Proportion 96
- 7.9 Testing for Normality 97
- 7.9.1 Skewness 98
- 7.9.2 Kurtosis 99

Chapter 8 Applications of the Chi-Square (χ^2) Distribution 104

- 8.1 The Chi-Square (χ^2) Test 105
- 8.2 χ^2 as a Test of Goodness of Fit 108
- 8.3 Test of Independence: Contingency Tables 113
- 8.3.1 Further Consideration of 2×2 Contingency Tables 117
- 8.4 Yates' Correction for Continuity 118
- 8.5 Fisher's Exact Test for 2×2 Tables 121

Chapter 9	Correlation and Regression	128
9.1	Correlation	128
9.2	Linear Regression	132
9.2.1	Partitioning of the Sum of Squares	136
9.2.2	Variance Estimates	138
9.2.3	Confidence Intervals and Tests of Hypothesis for A and B	139
9.2.4	X and Y Both Random Variables	141
9.2.5	Using \hat{Y} as a Predictor	141
Chapter 10	Analysis of Variance	146
10.1	Completely Randomized Design	147
10.1.1	Partitioning of the Sum of Squares	149
10.1.2	Construction of the Analysis of Variance Table	151
10.1.3	The Mathematical Model	152
10.1.4	Tests of Hypothesis	154
10.1.5	Estimation of Treatment Effects	156
10.1.6	Orthogonal Contrasts	157
10.1.7	Multiple Comparisons	159
10.1.8	Unequal Sample Sizes	162
10.1.9	The Random Effect Model	165
10.1.10	Testing the Homogeneity of Sample Variances	166
10.1.11	Data Transformation	169
10.2	Randomized Complete Block Designs	170
10.2.1	Missing Observations	177
10.3	Relative Efficiency of CRD and RCBD	179
Chapter 11	Nonparametric Methods	184
11.1	Wilcoxon Signed Rank Test	185
11.2	Wilcoxon Rank Sum Test	189
11.3	Kruskal-Wallis H-Test	191
11.3.1	Correction for Ties	192
11.4	Friedman's χ_r^2 -Test	194
11.5	Spearman's Rank Correlation Coefficient	196
11.5.1	Correction for Ties	197
11.6	McNemar Chi-Square Test for Related Samples	201

11.7	One-Sample Run Test for Randomness	202
------	------------------------------------	-----

Chapter 12 Survey Sampling 210

12.1	Principles of Sample Surveys	211
12.2	Steps of a Sample Survey	212
12.3	Sampling and Nonsampling Errors	214
12.4	Types of Sampling	216
12.4.1	Nonprobability Sampling	216
12.4.2	Probability Sampling	218

Chapter 13 Statistical Quality Control 233

13.1	Acceptance Sampling	234
13.1.1	Methods of Inspection	234
13.1.2	Characteristics of a Sampling Plan	235
13.1.3	Parameters Affecting Sampling Plans	235
13.1.4	Construction of an OC Curve	236
13.1.5	Types of Lot Sampling Plans	239
13.1.6	Risks and Quality Indices	240
13.1.7	Choosing a Sampling Plan	242
13.1.8	MIL-STD-105D	245
13.2	Control Charts	248
13.2.1	Nature of Control Charts	249
13.2.2	Types of Control Charts	249

Selected Bibliography 267

Appendix

1.	Binomial Coefficients $\binom{n}{x} = \frac{n!}{x!(n-x)!}$	273
2.	Values of Negative Exponential Function e^{-x}	273
3.	Normal Curve Areas	274
4.	Percentage Points of the t-Distribution	275
5.	Percentage Points of the χ^2 -Distribution	276
6.1	Binomial Confidence Intervals with $1 - \alpha = 0.95$	280
6.2	Binomial Confidence Intervals with $1 - \alpha = 0.99$	281

7. Percentiles of the Skewness Test Statistic a_3 282
8. Percentiles of Geary's Kurtosis Test Statistic 282
9. Transformation of r to Z $\left(\text{i.e., } Z = \frac{1}{2} \log e \frac{1+r}{1-r} \right)$
283
- 10.1 95% Significance Points of the F-Distribution
284
- 10.2 99% Significance Points of the F-Distribution
286
11. Upper Percentage Points of the Range for
Students, $q_d \bar{x}_{\max} - \bar{x}_{\min}$ 288
12. Significant Ranges for Students of 5% and 1%
Level: New Multiple-Range Test 290
13. Percentage Points of the Ratio, S_{\max}^2 / S_{\min}^2 292
14. Factorials and Their Common Logarithms 294
15. Critical Values for the Wilcoxon Test for Paired
Data 295
16. Critical Points of Rank Sums (Two-tailed
Alternatives) 296
- 17.1 Probabilities Associated with Obtaining Values
as Large as Observed Values with Friedman's
Test ($k = 3$) 297
- 17.2 Probabilities Associated with Obtaining Values
as Large as Observed Values with Friedman's
Test ($k = 4$) 298
18. Critical Values for a 5% Two-Sided Runs Test
299
19. Random Numbers 300
20. Summation of Terms of Poisson's Exponential
Binomial Limit 301
- 21.1 Sample Size Code Letters—MIL STD 105D
(ABC Standard) 304
- 21.2 Master Table for Normal Inspection (Single
Sampling)—MIL STD 105D (ABC Standard)
305
22. Factors for Computing Control Chartlines 306

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*.