



PRENTICE-HALL
INTERNATIONAL
SERIES IN
INDUSTRIAL
AND SYSTEMS
ENGINEERING

INTRODUCTORY STATISTICS AND PROBABILITY FOR ENGINEERING, SCIENCE, AND TECHNOLOGY

ELWOOD G. KIRKPATRICK

*School of Industrial Engineering
Purdue University*

PRENTICE-HALL, INC., *Englewood Cliffs, New Jersey*

Library of Congress Cataloging in Publication Data

KIRKPATRICK, ELWOOD G

Introductory statistics and probability for engineering, science, and technology.

1. Mathematical statistics. 2. Probabilities.
I. Title.

QA276.K49 519.5 73-23040

ISBN 0-13-501627-4

© 1974 by Prentice-Hall, Inc.
Englewood Cliffs, New Jersey

All rights reserved. No part of this book
may be reproduced in any form or by any means
without permission in writing from the publisher.

10 9 8 7 6 5 4 3

Printed in the United States of America

PRENTICE-HALL INTERNATIONAL, INC., *London*
PRENTICE-HALL OF AUSTRALIA, PTY. LTD., *Sydney*
PRENTICE-HALL OF CANADA, LTD., *Toronto*
PRENTICE-HALL OF INDIA PRIVATE LIMITED, *New Delhi*
PRENTICE-HALL OF JAPAN, INC., *Tokyo*

INTRODUCTORY STATISTICS
AND PROBABILITY
FOR ENGINEERING,
SCIENCE,
AND TECHNOLOGY

PRENTICE-HALL INTERNATIONAL SERIES
IN INDUSTRIAL AND SYSTEMS ENGINEERING

W. J. Fabrycky and J. H. Mize, Editors

- BLANCHARD *Logistics: Engineering and Management*
CINLAR *Introduction to Stochastic Processes*
FABRYCKY, GHARE, AND TORGENSEN *Industrial Operations Research*
FRANCIS AND WHITE *Facility Layout and Location: An Analytical Approach*
GOTTFRIED AND WEISMAN *Introduction to Optimization Theory*
KIRKPATRICK *Introductory Statistics and Probability for Engineering, Science, and Technology*
MIZE, WHITE, AND BROOKS *Operations Planning and Control*
OSTWALD *Cost Estimating for Engineering and Management*
SIVAZLIAN AND STANFEL *Analysis of Systems in Operations Research*
SIVAZLIAN AND STANFEL *Optimization Techniques in Operations Research*
WHITEHOUSE *Systems Analysis and Design Using Network Techniques*

To Eileen and Tom

PREFACE

This textbook is an introduction to some of the basic concepts and techniques of statistical analysis. It is directed primarily at the undergraduate technology and engineering student who has completed one introductory course in calculus. No previous background in either probability or statistics is assumed.

The book can be used for a one or two semester course depending on the emphasis and time devoted to end-chapter applications problems. For a two semester sequence, it is presumed that the instructor will complement this book with a formal design of experiments text of his choice. This book has been designed to satisfy the need for an introductory statistics text oriented to engineering applications, a text that is concise, readable by the average student, of moderate length, and not too mathematical. Two basic objectives of the text are: (1) to select and construct subject matter requirements that are attainable by the average student, and (2) to identify and structure topics to introduce the subject to a student who is meeting statistics for the first time.

To satisfy these objectives at this level, it is necessary not only to acquaint the reader with statistical methods but also to indicate the mathematics required to establish the methods on a sound basis. Thus, the presentation is essentially a compromise between a book that gives only statistical

techniques and a book that attempts complete mathematical rigor. Some mathematical foundations are given to structure the subject and to stimulate the interested student to pursue specialized and more advanced courses in statistics. A nominal sacrifice of simplicity has been made at the beginning in order to introduce concepts that are fundamental to the development of the subject. For example, the distribution function and characteristic function are introduced in elementary fashion early in Chaps. 3 and 4. However, emphasis throughout the text is primarily on language, method, and applications problems.

An observation made by the author on many occasions is that junior and senior engineering students, who have had one or two statistics courses, frequently demonstrate a remarkable lack of statistical facility in other courses that use statistical tools for quantitative evaluations. One conjecture as to this puzzling end result is that the mathematics of applied statistics is not too difficult, but the amount of time that is required to pursue topics in depth leaves little time for applications and analysis practice. Operating from this assumption, this text is designed to save time wherever possible. Simple functions are used to illustrate concepts. For example, the uniform distribution function is used frequently in Chaps. 3 and 4 to illustrate obtaining distribution moments using either expected value or characteristic function methods. Other simple linear functions are used in examples, exercises, and problems. This is also the reason for an early presentation of the discrete random variable models.

Simple and direct exercises are given regularly throughout each chapter (more rigorous and applied problems are separated into the end-chapter problem assignments). Extensive use is made of graphical explanations and summaries, particularly for the problems of Chaps. 5, 7, 8, and 10. The purpose throughout is to take cognizance of the fact that technology and engineering students simply do not have the time to study statistics in a career fashion.

The author also feels that another important item related to the aforementioned student difficulty is the verbal factor. A first chore for the average reader is to sort the verbiage and condense it to a meaningful structure for himself—a task usually requiring extensive help by the instructor. To simplify this chore, use is made of discussion-paragraph asides blocked out separately from the main paragraphs. Generally, the main paragraphs read continuously and the discussion paragraphs contribute support material—intuitive explanations, of-interest-only derivations, additional argument, detail, exceptions, special cases, and so forth. Student difficulties seem to stem from too much material presented too rapidly and coincidentally. The pedagogical approach throughout this text is that of separation of material. The discussion paragraph asides separate out two types of supplementary material: (1) additional

intuitive explanations and aids for the below-average student, and (2) mathematical foundations to interest the above-average student.

In an introductory textbook, the selection of topics, their order of presentation, and extent of coverage are to some degree subjective decisions. There is no one correct sequence and there are a variety of effective teaching combinations. To meet objectives, frequent tradeoffs were made in developing the text, and sophisticated detail was necessarily omitted in some cases. Only elementary probability concepts are presented for the limited purpose of introducing the general statistical topics. A combinatorial rather than a set-theory approach is used. It is felt that the latter view is too abstract for an introductory text, and that the combinatorial approach is more adaptable to an intuitive counting of events and occurrences. Discrete distribution models are introduced early (Chap. 4) to afford simple examples of obtaining distribution moments and also as a means of introducing probabilities of events of the form $X \leq b$ and thereby gradually leading to the use of tables for making probability statements. The topic of sampling inspection (without analysis) is introduced early to obtain practical applications and examples for the binomial and Poisson models. The topics receiving the most coverage are descriptive statistics and hypothesis testing. The latter topic is perhaps the most important single topic in an introductory statistics textbook as it is fundamental to all advanced statistical topics.

The principal quality control topics (Chap. 8) follow immediately the presentation of hypothesis testing (Chap. 7) to further illustrate tests of hypotheses. Since engineering problems often require a presentation of data showing the observed relationship between variables, a separate chapter on curve fitting (Chap. 9) has been included. This treatment, without any of the assumptions required for regression, makes for a less abrupt entry into the regression topic. The chapter on regression that follows also lends itself to a gradual and more meaningful introduction to analysis of variance.

I am grateful to Mrs. Jong-Ping Hsu for her careful review of the manuscript, and to Dr. J. H. Mize, Oklahoma State University, for his review and many helpful suggestions. I am especially grateful to Mr. Kenneth Hopping for his thorough effort on the tedious job of proofreading.

I am indebted to the Literary Executor of the late Sir Ronald A. Fisher, F.R.S., to Dr. Frank Yates, F.R.S., and to Oliver and Boyd, Edinburgh, for permission to reprint Tables of the chi square, t -, and F -distributions (Appendixes E, F, and G) from their book *Statistical Tables for Biological, Agricultural and Medical Research*; to Dr. Frederick E. Croxton for permission to reprint the normal distribution Table (Appendix D) from his book *Elementary Statistics with Applications in Medicine and the Biological Sciences*, Dover Publications, Inc., 1959; to the *Annals of Mathematical Statistics* and Professors C. L. Ferris, F. E. Grubbs, and C. L. Weaver for permission to

reproduce their Operating Characteristic Curves for Significance Tests (Chap. 7); to Dr. George W. Snedecor for permission to reprint a random numbers Table (Appendix I) from his book *Everyday Statistics*; to Professors A. H. Bowker and G. J. Lieberman and the Stanford University Press for permission to reproduce the hypergeometric table (Appendix C) from their book *Tables of the Hypergeometric Probability Distribution*; and to the American Society for Testing Materials for permission to reproduce Table B.2 (Appendix H) from their *A. S. T. M. Manual on Quality Control of Materials*.

West Lafayette, Indiana

ELWOOD G. KIRKPATRICK

CONTENTS

Preface	xi
1 Introduction	1
1.1 Statistical Inference	1
1.2 Statistical Investigation	3
1.3 Population and Sample	4
1.4 Frequency Distribution	7
1.5 Frequency Distribution—An Analysis Tool	10
1.6 Other Frequency Distribution Representation	12
1.7 Summation Operations	15
1.8 Descriptive Measures	17
1.9 Measures of Average	18
1.10 Arithmetic Mean	18
1.11 Mean of Means	20
1.12 Other Measures of Average	21
1.13 Measures of Variation	28
1.14 Computation of Variance	32
	vii

1.15	Change of Scale (Data Coding)	34
1.16	Z-transformation	38
1.17	Moments	39
1.18	Significant Digits and Data Accuracy	41
2	Elementary Probability	50
2.1	Fundamental Principle for Counting	50
2.2	Factorials	52
2.3	Combinations	53
2.4	Probability	55
2.5	Operations with Events and Probabilities	57
2.6	Conditional Probability	60
2.7	Baye's Theorem of Inverse Probability	61
2.8	Relative Frequency Probability	62
2.9	Probability Based on Algebra of Sets	63
3	Probability Distributions	68
3.1	Random Variable	68
3.2	Probability Function	70
3.3	Cumulative Distribution Function	71
3.4	Continuous Random Variables	73
3.5	Probability Distribution Table	76
4	Discrete Distribution Models	84
4.1	Mathematical Expectation	84
4.2	Chebyshev's Inequality	88
4.3	Characteristic Function of a Distribution	89
4.4	Bernoulli Distribution Model	92
4.5	Binomial Distribution Model	94
4.6	Poisson Distribution Model	101
4.7	Hypergeometric Distribution Model	106
4.8	Summary—Discrete Models	108
4.9	Other Discrete Models	110
5	Normal Distribution Model	120
5.1	The Normal Distribution	120
5.2	Properties of the Normal Distribution	121
5.3	Normal Model in Standard Form	122
5.4	Normal Approximation to the Binomial	129
5.5	Normalizing Transformation	132
5.6	Central Limit Theorem	133
5.7	Sampling Distribution of the Mean	134
5.8	Specifications	136
5.9	Combinations of Independent Random Variables	139

6	Continuous Distribution Models	156
6.1	Degrees of Freedom	156
6.2	Chi-Square Distribution Model	158
6.3	Student's t -Distribution Model	163
6.4	F -Distribution Model	165
6.5	Distribution of \bar{x}	168
6.6	Distributions Related to s_x^2	170
6.7	Distribution of $\bar{x} - \bar{y}$	171
7	Hypothesis Testing and Estimation	178
7.1	Rationale of Hypothesis Testing	179
7.2	Nomenclature	181
7.3	Hypothesis Test Concerning the Mean, with Standard Deviation Known	182
7.4	Type II Error and β	185
7.5	One and Two-Sided Tests	191
7.6	α , β , n Relationships	195
7.7	Analytic Solution for β or n	198
7.8	Summary—Test of the Mean with Standard Deviation Known	201
7.9	Test of the Mean with Standard Deviation Unknown	202
7.10	Test That Two Means Are Equal	207
7.11	Test That Two Means Are Equal, Paired Observations	213
7.12	Test of the Standard Deviation of a Normal Population	218
7.13	Test That Two Standard Deviations Are Equal (Normal Populations)	226
7.14	Test of the Proportion	232
7.15	Test of Goodness of Fit	239
7.16	The Estimation Problem	247
7.17	Rationale of Estimation	249
8	Quality Control	266
8.1	Process Capability	266
8.2	Process Control	270
8.3	Sampling Inspection	279
8.4	Operating Characteristic Curve	280
8.5	Sampling Plan Evaluation	283
8.6	Sampling Plan Design	285
8.7	Variables Sampling Plans	287
9	Curve Fitting	298
9.1	Method of Moments	299
9.2	Method of Least Squares	302
9.3	Fitting a Polynomial	305

9.4 Fitting an Exponential	308
9.5 Choice of Model	310
9.6 Fitting a Distribution Function	311
10 Regression and Correlation	318
10.1 The Regression Problem	318
10.2 Significance of Regression	321
10.3 Hypothesis Tests—Slope and Intercept	325
10.4 Confidence Interval Estimates—Slope and Intercept	328
10.5 Confidence Interval Estimate of $\mu_{Y X}$	329
10.6 Confidence Interval Estimate of Y_0	330
10.7 Correlation	333
10.8 Inferences Regarding ρ	336
10.9 Multiple Linear Regression	339
10.10 Significance of Addition of a Variable to the Regression Equation	342
10.11 Additional Remarks—Regression and Correlation	344
11 Analysis of Variance	352
11.1 Single Variable, Completely Randomized Design	353
11.2 Randomized Complete Block Design	362
11.3 Latin Square Design	367
11.4 Introduction to Factorial Design	369
12 Processing of Statistical Data	382
12.1 Desk Calculator Computation	382
12.2 Digital Computer Calculation	389
Appendices	395
A Cumulative Binomial Distribution	397
B Cumulative Poisson Distribution	403
C The Hypergeometric Distribution	408
D Tail Areas—The Normal Distribution	411
E The Chi-Square Distribution	413
F The t -Distribution	416
G The F -Distribution	418
H Quality Control Chart Coefficients	429
I Random Numbers	431
Index	439

INTRODUCTORY STATISTICS
AND PROBABILITY
FOR ENGINEERING,
SCIENCE,
AND TECHNOLOGY

INTRODUCTION

Statistics is concerned with methods for the collection, analysis, and interpretation of quantitative data in such a way that conclusions based on the data can be objectively evaluated in terms of probability statements. The collection of data and computation of various indices such as averages and percentages is called descriptive statistics. Systematic drawing of conclusions from the data is called statistical inference.

1.1 Statistical Inference

The deductive method (or proof) makes inferences or conclusions from accepted principles. On the other hand, the inductive method draws conclusions from several known cases, reasoning from the particular to the general. The technology of deduction is mathematics, while that of induction is statistics. An old science, which has already developed a number of basic principles, can use mathematics to deduce more information from these principles. A young science depends on statistics to develop the basic principles. All sciences use statistical tools for reaching conclusions from experimental data.

1

Although inductive reasoning—going from the particular to the general—can lead to numerous difficulties, it is still possible to be precise about it in terms of the risk involved in making such an inference. The topic of statistical inference may be regarded in this light. The effort is not aimed at making statements each of which is guaranteed to be 100% correct, but rather at making inferences that have a certain probability of being correct.

1.2 Statistical Investigation

The important phases of a statistical investigation are:

1. Definition of the problem.
2. Design of the experiment.
3. Data collection.
4. Data reduction and computations.
5. Analysis.

A formulation or *definition* of the problem includes: (1) statement of the problem, (2) choice of dependent variable (or variables) to be studied, and (3)