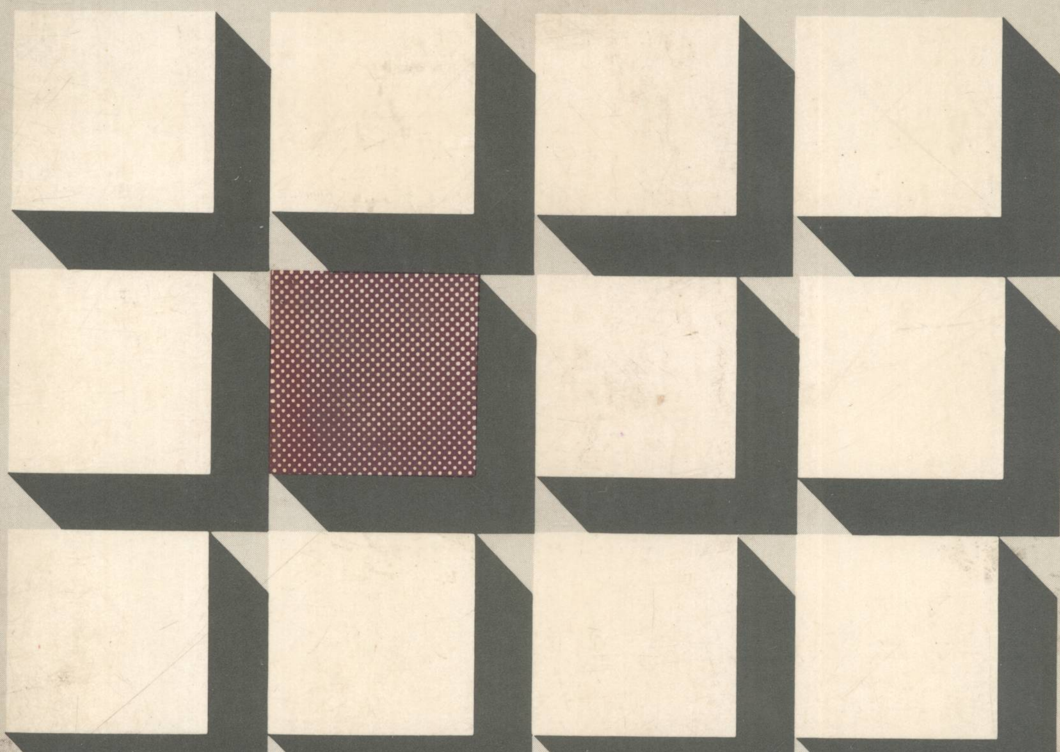
INTRODUCTION TO STATISTICS

Third Edition

RONALD E. WALPOLE



3rd edition

Introduction to Statistics

Ronald E. Walpole

Professor of Mathematics and Statistics, Roanoke College

Macmillan Publishing Co., Inc.
NEW YORK

Collier Macmillan Publishers
LONDON

Macmillan Publishing Co., Inc.
866 Third Avenue, New York, New York 10022

Collier Macmillan Canada, Inc.

Library of Congress Cataloging in Publication Data

Walpole, Ronald E.
Introduction to statistics.

Bibliography: p.
Includes index.
1. Statistics. I. Title.
QA276.12.W36 1982 519.5 81-8189
ISBN 0-02-424150-4 (Hardbound) AACR2
ISBN 0-02-977650-3 (International Edition)

Printing: 1 2 3 4 5 6 7 8 Year: 2 3 4 5 6 7 8

A study guide containing numerous worked examples and practice problems has been prepared to assist the student in mastering the various statistical concepts that are presented in this text. This *Student Study Guide—Introduction to Statistics, 3rd edition,* by William D. Ergle and Ronald E. Walpole may be purchased from your local bookstore.

Frequently Used Formulas

Sample mean $$\bar{x} = \frac{\Sigma x}{n}$$

Sample variance $$s^2 = \frac{n\Sigma x^2 - (\Sigma x)^2}{n(n - 1)}$$

Additive rule of probability $$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Multiplicative rule of probability $$P(A \cap B) = P(A)P(B|A)$$

Conditional probability $$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Mean of a discrete random variable $$\mu = \Sigma x f(x)$$

Variance of a discrete random variable $$\sigma^2 = \Sigma(x - \mu)^2 f(x)$$

Binomial distribution $$b(x; n, p) = \binom{n}{x} p^x q^{n-x}$$

Binomial mean	$\mu = np$
Binomial variance	$\sigma^2 = npq$
Hypergeometric distribution	$h(x; N, p, k) = \dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$
Poisson distribution	$p(x; \mu) = \dfrac{e^{-\mu}\mu^x}{x!}$
Confidence interval for μ (σ known)	$\bar{x} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$
Confidence interval for μ (σ unknown)	$\bar{x} \pm t_{\alpha/2}\dfrac{s}{\sqrt{n}}$
Sample size for estimating μ	$n = \left(\dfrac{z_{\alpha/2}\sigma}{e}\right)^2$
Confidence interval for $\mu_1 - \mu_2$ (σ_1 and σ_2 known)	$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$
Pooled estimate of σ^2	$s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$
Confidence interval for $\mu_1 - \mu_2$ ($\sigma_1 = \sigma_2$ but unknown)	$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$
Confidence interval for μ_D (paired observations)	$\bar{d} \pm t_{\alpha/2}\dfrac{s_d}{\sqrt{n}}$

Continued on rear endpapers

Introduction to Statistics

Dedicated to the memory of my parents

Preface

Like the previous editions, this third edition of *Introduction to Statistics* has been written to serve as an introductory statistics text for students majoring in any of the academic disciplines, whether it be sociology, psychology, economics, business administration, or one of the sciences. A knowledge of high school algebra is sufficient to comprehend the basic concepts of descriptive and inferential statistics that are presented, although experience seems to indicate that a student benefits more from such a course when it is preceded by either a semester course in elementary functions or finite mathematics.

In preparing this third edition, all examples and exercises dealing with numerical measurements have been converted to metric units. Unlike the first two editions in which the exercises were placed only at the end of each chapter, now the numerous illustrative exercises, many of which are new, have been placed immediately following appropriate sections within each chapter. Optional statistical concepts and procedures defined throughout many of the exercise sets may be included at the discretion of the professor. A *Student Study Guide,* which provides additional insight and practice in solving the various types of problems presented throughout the text, accompanies this edition.

The concepts of descriptive statistics, which previously were scattered throughout the text, have now been placed in the first three chapters. To complement this area, the treatment of percentiles, quartiles, and deciles has been expanded to cover both grouped and ungrouped data, the Pearsonian coefficient of skewness is defined, and new material has been added to illustrate the use of a random number table in selecting a random sample. The field of modern statistics, with an increased emphasis on statistical inference, is based primarily on the theory of probability. An introduction to the basic concepts of probability theory using set notation is, therefore, presented in Chapter 4. The material in Chapter 5, introducing random variables and their mathematical expectations, and discrete and continuous probability distributions and the various properties describing these distributions, has been retained from the second edition, although the discussion on mathematical expectations has been completely revised.

An introduction to several discrete probability distributions is presented in Chapter 6, followed by a discussion of the normal distribution in Chapter 7. This then naturally leads to the treatment of sampling theory, estimation theory, and hypothesis testing in Chapters 8, 9, and 10, respectively. Chapter 8 has been revised somewhat by including a discussion of simulated experiments as well as a brief treatment of systematic, stratified, and cluster sampling procedures, while delaying our study of the chi-square and F distributions until Chapter 9. A new section on Bayesian methods of estimating the binomial parameter and the normal mean has also been included in Chapter 9. Chapter 11 presents the general techniques of curve fitting followed by a discussion of regression theory and an expanded treatment of correlation analysis that includes tests of significance and a new section on partial and multiple correlation. An introduction to analysis of variance with an expanded treatment of experimental designs covering randomized block and Latin square designs is presented in Chapter 12. Although some authors prefer to use a regression approach to analysis of variance, no attempt has been made in this text to relate the two chapters. Either chapter may be considered without a knowledge of the other. For greater emphasis, the material on nonparametric tests has been greatly expanded and now constitutes the final chapter of the text.

The text contains sufficient material to allow for flexibility in the length of the course and in the selection of topics. A semester course, meeting three hours a week, should include most of the material in Chapters 1 through 5; Sections 6.1, 6.2, and 6.3 of Chapter 6; Chapters 7 through 10, perhaps excluding Sections 9.9 and 9.10 of Chapter 9 and Sections 10.8, 10.9, and 10.10 of Chapter 10; and Sections 11.1, 11.2, 11.3, 11.4, and 11.8 of Chapter 11.

I wish to acknowledge my appreciation to all those who assisted in the preparation of this textbook. I am particularly grateful to Rhonda Haga and Kim Marzke for typing and proofreading this revised third edition and the accompanying *Solutions Manual;* to Dr. William D. Ergle for writing the major part of the *Student Study Guide;* to the Macmillan Publishing Co. Inc., for their editorial assistance; and to the many teachers and reviewers of the first two editions for their helpful suggestions in preparing this third edition.

I am indebted to the Literary Executor of the late Sir Ronald A. Fisher, F.R.S., Cambridge, and to Oliver & Boyd Ltd., Edinburgh, for their permission to reprint Table A.5 from their book *Statistical Methods for Research Workers;* to Professor E. S. Pearson and the Biometrika Trustees for permission to reprint in abridged form, as Tables A.6 and A.7, Tables 8 and 18 from *Biometrika Tables for Statisticians,* Vol. I; to D. Van Nostrand Company, Inc., for permission to reproduce in Table A.3 material from E. C. Molina's *Poisson's Exponential Binomial Limit*. I wish also to express my appreciation for permission to reproduce Table A.8 from a publication by the American Cyanamid Company, Table A.9 from the *Bulletin of the Educational Research at Indiana University,* Tables A.10 and A.14 from the *Annals of Mathematical Statistics,* Table A.11 from *Biometrics,* and Table A.13 from the *Journal of the American Statistical Association*.

R.E.W.

Contents

1 Introduction

The processing of statistical information has a history that extends back to the beginning of mankind. In early biblical times nations compiled statistical data to provide descriptive information relative to all sorts of things, such as taxes, wars, agricultural crops, and even athletic events. Today, with the development of probability theory, we are able to use statistical methods that not only describe important features of the data but methods that allow us to proceed beyond the collected data into the area of decision making through generalizations and predictions.

1.1 Descriptive and Inferential Statistics

In the study of statistics we are basically concerned with the presentation and interpretation of **chance outcomes** that occur in a planned or scientific investigation. For example, we may record the number of accidents that occur monthly at the intersection of Driftwood Lane and Royal Oak Drive, hoping to justify the installation of a traffic light, we might classify responses in an opinion poll as "yes" or "no," or we may be interested in the amount of residue deposited in a chemical reaction when the concentration of the catalyst is varied. Hence the statistician is usually dealing with either **numerical data** representing **counts** or **measurements,** or perhaps with **categorical data** that can be classified according to some criterion.

We shall refer to any recording of information, whether it be numerical or categorical, as an **observation.** Thus the numbers 3, 1, 0, and 2, representing the number of accidents that occurred for each month from January through April during the past year at the intersection of Driftwood Lane and Royal Oak Drive, constitute a set of observations. Similarly, the measurements 2.5, 3.1, and 1.8 grams of residue deposited in the chemistry experiment are recorded as observations.

Example 1. The students at Hidden Valley High are given blood tests to determine their type of blood. A person's blood can be classified in 8 ways. It must be AB, A, B, or O, with a plus or minus sign, depending on the presence or absence of the Rh antigen. Of course, this results in a descriptive or categorical representation of the data rather than a numerical measurement or count.

Statistical methods are those procedures used in the *collection, presentation, analysis, and interpretation of data*. We shall categorize these methods as belonging to one of two major areas called **descriptive statistics** and **statistical inference.** First, let us consider the field of descriptive statistics.

DEFINITION

Descriptive Statistics. Descriptive statistics comprises those methods concerned with collecting and describing a set of data so as to yield meaningful information.

Let it be clearly understood that descriptive statistics provides information only about the collected data and in no way draws inferences or conclusions concerning a larger set of data. The construction of tables, charts, graphs, and