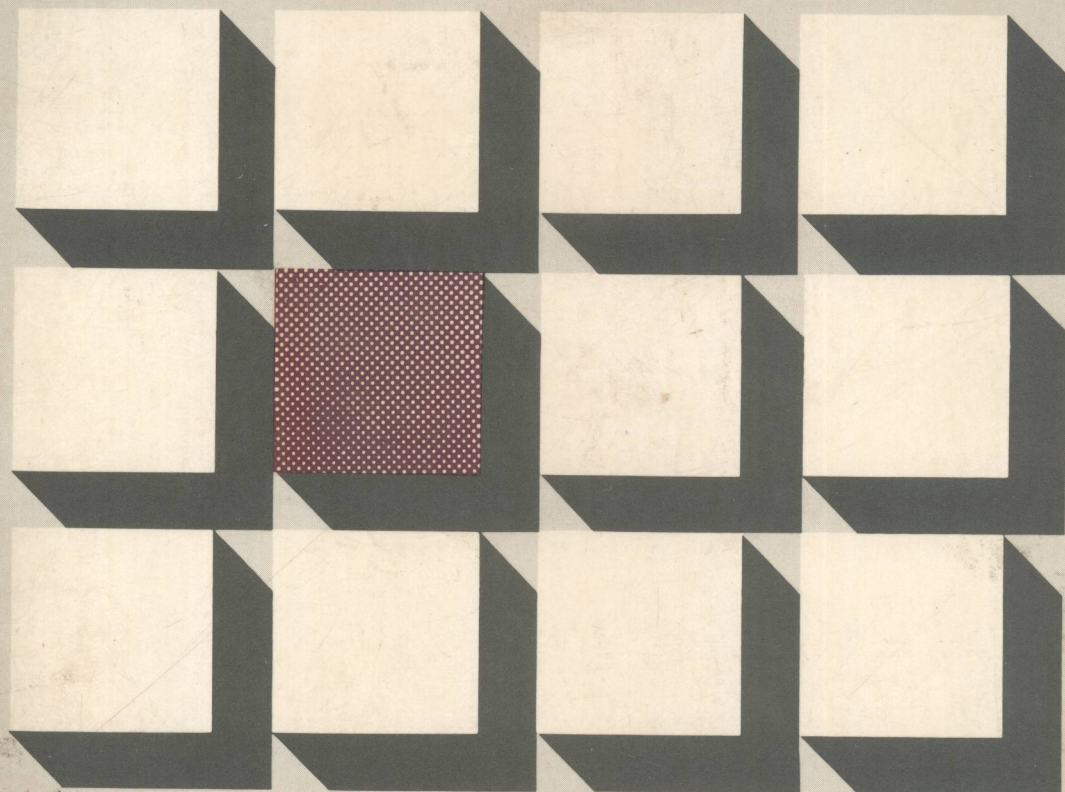

INTRODUCTION TO STATISTICS

Third Edition
RONALD E. WALPOLE



0212
W219-2
E.3

8364038

3rd
edition

Introduction to Statistics

Ronald E. Walpole

Professor of Mathematics and Statistics, Roanoke College



E8364038



Macmillan Publishing Co., Inc.
NEW YORK

Collier Macmillan Publishers
LONDON

Copyright © 1982, Ronald E. Walpole.

Printed in the United States of America.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the Publisher.

Earlier editions copyright © 1968 and 1974 by Ronald E. Walpole.

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 $\bar{x} = \frac{\sum x}{n}$

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

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 $\mu = \sum xf(x)$

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

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

$$h(x; N, p, k) = \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}}$$

Poisson distribution

$$p(x; \mu) = \frac{e^{-\mu} \mu^x}{x!}$$

Confidence interval for μ
(σ known)

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Confidence interval for μ
(σ unknown)

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

Sample size for estimating μ

$$n = \left(\frac{z_{\alpha/2} \sigma}{e} \right)^2$$

Confidence interval for $\mu_1 - \mu_2$
(σ_1 and σ_2 known)

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Pooled estimate of σ^2

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Confidence interval for $\mu_1 - \mu_2$
($\sigma_1 = \sigma_2$ but unknown)

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Confidence interval for μ_D
(paired observations)

$$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}}$$

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.

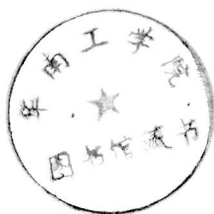
I wish to acknowledge my appreciation to all those who assisted in the preparation of this textbook. I am particularly grateful to Rhonda Haga and Kim Marzke for typing and proofreading this revised third edition and the accompanying *Solutions Manual*; to Dr. William D. Ergle for writing the major part 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.

8364038

Contents



1

Introduction

1

- 1.1 Descriptive and Inferential Statistics 2
- 1.2 Populations and Samples 6
- 1.3 Problems for the Statistician 9
- 1.4 Summation Notation 12

2

Statistical Measures of Data

21

- 2.1 Parameters and Statistics 22
- 2.2 Measures of Central Location 23
- 2.3 Measures of Variation 31
- 2.4 Chebyshev's Theorem 39
- 2.5 z Scores 40

3

Statistical Description of Data

47

- 3.1 Frequency Distributions 48
- 3.2 Graphical Representations 53
- 3.3 Symmetry and Skewness 57
- 3.4 Percentiles, Deciles, and Quartiles 61

xi

4

Probability

69

- 4.1 Sample Space 70
- 4.2 Events 72
- 4.3 Operations with Events 74
- 4.4 Counting Sample Points 82
- 4.5 Probability of an Event 89
- 4.6 Additive Rules 93
- 4.7 Conditional Probability 98
- 4.8 Multiplicative Rules 102
- 4.9 Bayes' Rule 108

5

Distributions of Random Variables

115

- 5.1 Concept of a Random Variable 116
- 5.2 Discrete Probability Distributions 118
- 5.3 Continuous Probability Distributions 120
- 5.4 Joint Probability Distributions 126
- 5.5 Mean of a Random Variable 133
- 5.6 Variance of a Random Variable 140
- 5.7 Properties of the Mean and Variance 145

6

Some Discrete Probability Distributions

155

- 6.1 Uniform Distribution 156
- 6.2 Binomial and Multinomial Distributions 157
- 6.3 Hypergeometric Distribution 167
- 6.4 Negative Binomial and Geometric Distributions 175
- 6.5 Poisson Distribution 177

7

Normal Distribution

183

- 7.1 Normal Curve 184
- 7.2 Areas Under the Normal Curve 186

- 7.3 Applications of the Normal Distribution 192
7.4 Normal Approximation to the Binomial Distribution 200

8

Sampling Theory

209

- 8.1 Sampling Distributions 210
8.2 Sampling Distributions of the Mean 211
8.3 t Distribution 219
8.4 Sampling Distribution of the Differences of Means 224
8.5 Simulated Experiments 233
8.6 Sampling Procedures 236

9

Estimation of Parameters

243

- 9.1 Statistical Inference 244
9.2 Classical Methods of Estimation 244
9.3 Estimating the Mean 247
9.4 Estimating the Difference Between Two Means 254
9.5 Estimating a Proportion 266
9.6 Estimating the Difference Between Two Proportions 271
9.7 Estimating the Variance 274
9.8 Estimating the Ratio of Two Variances 278
9.9 Bayesian Methods of Estimation 283
9.10 Decision Theory 287

10

Tests of Hypotheses

293

- 10.1 Statistical Hypotheses 294
10.2 Testing a Statistical Hypothesis 295
10.3 One-Tailed and Two-Tailed Tests 304
10.4 Tests Concerning Means 308
10.5 Tests Concerning Variances 320
10.6 Tests Concerning Proportions 325
10.7 Testing the Difference Between Two Proportions 328

- 10.8 Goodness-of-Fit Test 332
- 10.9 Test for Independence 336
- 10.10 Testing Several Proportions 339

11

Regression and Correlation

345

- 11.1 Linear Regression 346
- 11.2 Regression Analysis 353
- 11.3 Inferences Concerning the Regression Coefficients 357
- 11.4 Prediction 361
- 11.5 Test for Linearity of Regression 365
- 11.6 Exponential Regression 367
- 11.7 Multiple Regression 369
- 11.8 Linear Correlation 374
- 11.9 Multiple and Partial Correlation 379

12

Analysis of Variance

385

- 12.1 Analysis-of-Variance Technique 386
- 12.2 One-Way Classification 387
- 12.3 Test for the Equality of Several Variances 396
- 12.4 Multiple-Range Test 398
- 12.5 Two-Way Classification 403
- 12.6 Two-Way Classification with Interaction 411
- 12.7 Design of Experiments 422
- 12.8 Latin Squares 424

13

Nonparametric Statistics

433

- 13.1 Nonparametric Tests 434
- 13.2 Sign Test 434
- 13.3 Wilcoxon Signed-Rank Test 439
- 13.4 Wilcoxon Rank-Sum Test 446

13.5	Kruskal–Wallis Test	450
13.6	Runs Test	455
13.7	Rank Correlation Coefficient	460

References

467

**Appendix:
Statistical
Tables**

469

**Answers to
Selected
Exercises**

499

Index

513

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