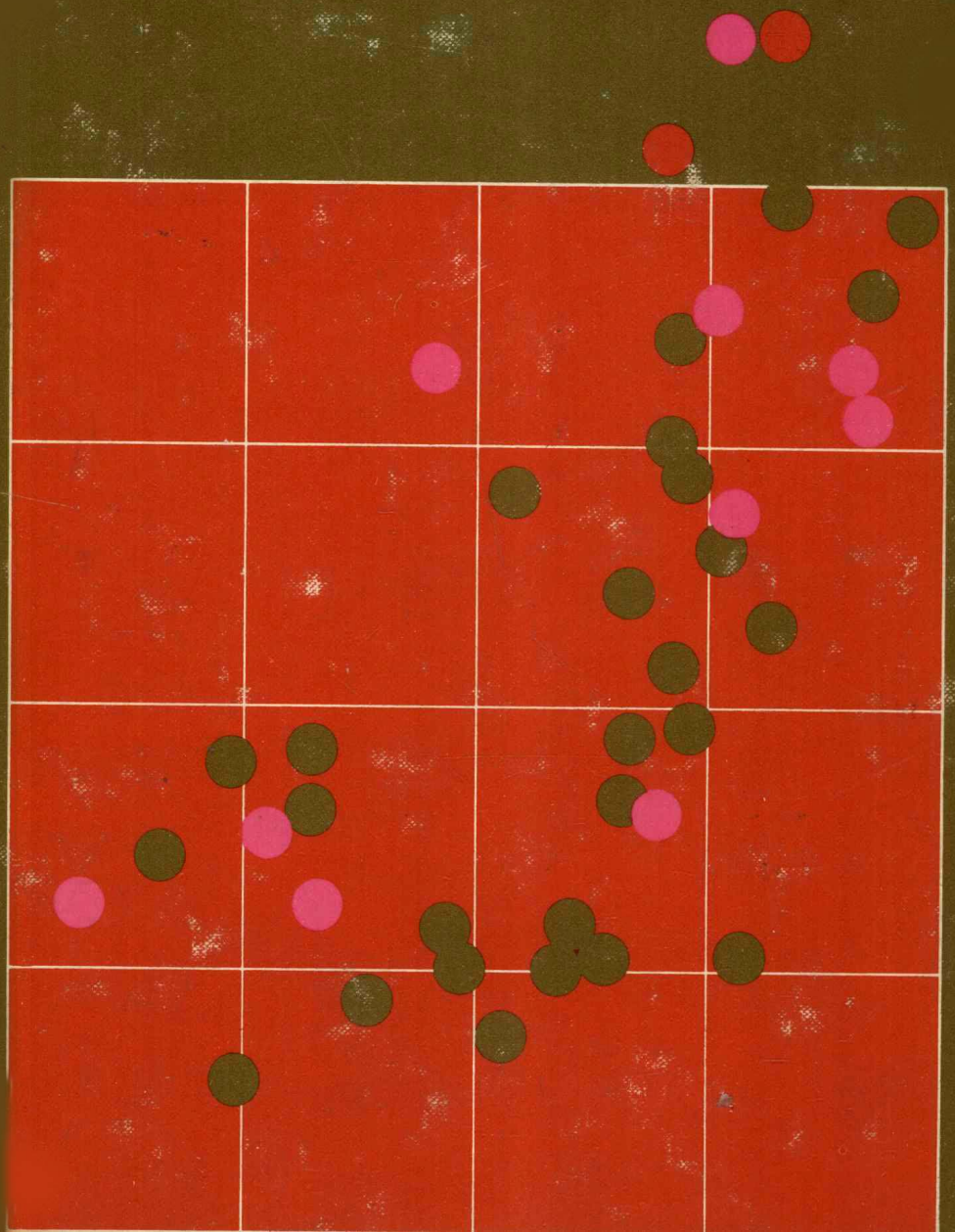
Paul G. Hoel

# elementary statistics

ed.



# Elementary Statistics

## Fourth Edition

Paul G. Hoel

Professor of Mathematics

University of California, Los Angeles

John Wiley & Sons, Inc.

New York London Sydney Toronto

*Library of Congress Cataloging in Publication Data*

Hoel, Paul Gerhard, 1905–
Elementary statistics.

(Wiley series in probability and mathematical statistics)
Includes index.
1. Statistics. I. Title.
HA29.H662 1976 519.5 75-33400
ISBN 0-471-40302-4

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

## WILEY SERIES IN PROBABILITY AND MATHEMATICAL STATISTICS

ESTABLISHED BY WALTER A. SHEWHART AND SAMUEL S. WILKS

Editors

*Ralph A. Bradley* — *David G. Kendall*

*J. Stuart Hunter* — *Geoffrey S. Watson*

*Probability and Mathematical Statistics*

ANDERSON · The Statistical Analysis of Time Series

ANDERSON · An Introduction to Multivariate Statistical Analysis

BARLOW, BARTHOLOMEW, BREMNER, and BRUNK · Statistical Inference Under Order Restrictions

DE FINETTI · Probability, Induction and Statistics

DOOB · Stochastic Processes

FELLER · An Introduction to Probability Theory and Its Applications, Volume I, *Third Edition,* Revised

FELLER · An Introduction to Probability Theory and Its Applications, Volume II, *Second Edition*

FISZ · Probability Theory and Mathematical Statistics, *Third Edition*

FRASER · Statistics—An Introduction

FRASER · The Structure of Inference

HANNAN · Multiple Time Series

HANSEN, HURWITZ, and MADOW · Sample Survey Methods and Theory, Volumes I and II

HOEL · Introduction to Mathematical Statistics, *Fourth Edition*

KEMPTHORNE · The Design and Analysis of Experiments

LARSON · Introduction to Probability Theory and Statistical Inference

LARSON · Introduction to the Theory of Statistics

LEHMANN · Testing Statistical Hypotheses

PARZEN · Modern Probability Theory and Its Applications

PURI and SEN · Nonparametric Methods in Multivariate Analysis

RAGHAVARAO · Constructions and Combinatorial Problems in Design of Experiments

RAO · Linear Statistical Inference and Its Applications, *Second Edition*

RIORDAN · An Introduction to Combinatorial Analysis

SCHEFFÉ · The Analysis of Variance

WILKS · Collected Papers: Contributions to Mathematical Statistics

WILKS · Mathematical Statistics

ZACKS · The Theory of Statistical Inference

*Applied Probability and Statistics*

BAILEY · The Elements of Stochastic Processes with Applications to the Natural Sciences

BARTHOLOMEW · Stochastic Models for Social Processes

BENNETT and FRANKLIN · Statistical Analysis in Chemistry and the Chemical Industry

BHAT · Elements of Applied Stochastic Processes

BOX and DRAPER · Evolutionary Operation: A Statistical Method for Process Improvement

BROWNLEE · Statistical Theory and Methodology in Science and Engineering, *Second Edition*

*Applied Probability and Statistics (Continued)*

CHAKRAVARTI, LAHA and ROY · Handbook of Methods of Applied Statistics, Vol. I

CHAKRAVARTI, LAHA and ROY · Handbook of Methods of Applied Statistics, Vol. II

CHERNOFF and MOSES · Elementary Decision Theory

CHIANG · Introduction to Stochastic Processes in Biostatistics

CLELLAND, deCANI, BROWN, BURSK, and MURRAY · Basic Statistics with Business Applications, *Second Edition*

COCHRAN · Sampling Techniques, *Second Edition*

COCHRAN and COX · Experimental Designs, *Second Edition*

COX · Planning of Experiments

COX and MILLER · The Theory of Stochastic Processes

DANIEL and WOOD · Fitting Equations to Data

DAVID · Order Statistics

DEMING · Sample Design in Business Research

DODGE and ROMIG · Sampling Inspection Tables, *Second Edition*

DRAPER and SMITH · Applied Regression Analysis

ELANDT-JOHNSON · Probability Models and Statistical Methods in Genetics

FLEISS · Statistical Methods for Rates and Proportions

GOLDBERGER · Econometric Theory

GUTTMAN, WILKS and HUNTER · Introductory Engineering Statistics, *Second Edition*

HAHN and SHAPIRO · Statistical Models in Engineering

HALD · Statistical Tables and Formulas

HALD · Statistical Theory with Engineering Applications

HOEL · Elementary Statistics, *Third Edition*

HOLLANDER and WOLFE · Nonparametric Statistical Methods

HUANG · Regression and Econometric Methods

JOHNSON and KOTZ · Distributions in Statistics

Discrete Distributions

Continuous Univariate Distributions-1

Continuous Univariate Distributions-2

Continuous Multivariate Distributions

JOHNSON and LEONE · Statistics and Experimental Design: In Engineering and the Physical Sciences, Volumes I and II

LANCASTER · The Chi Squared Distribution

LEWIS · Stochastic Point Processes

MILTON · Rank Order Probabilities: Two-Sample Normal Shift Alternatives

OTNES and ENOCHSON · Digital Time Series Analysis

RAO and MITRA · Generalized Inverse of Matrices and Its Applications

SARD and WEINTRAUB · A Book of Splines

SARHAN and GREENBERG · Contributions to Order Statistics

SEAL · Stochastic Theory of a Risk Business

SEARLE · Linear Models

THOMAS · An Introduction to Applied Probability and Random Processes

*continued on back*

# Preface to the Fourth Edition

This book is designed for a one- or two-semester course for the student whose background in mathematics is limited to high school algebra. A number of mathematics departments offer a service course of this kind; however, in many schools a department with strong interests in the applications of statistics gives the course.

The present edition differs from the preceding one principally in that the illustrations and problems have been supplemented with numerous interesting applications to current social problems. The data for most of these problems were obtained from journals in medicine and the social and life sciences. The only noticeable change in organization is the introduction of a new chapter on multiple regression, together with the deletion of such material in the chapter on correlation and regression.

The last section of each chapter contains additional illustrations of the methods presented in that chapter. It may be used by the student as a basis for reviewing the entire chapter or as a source of additional illustrations for the individual sections of the chapter. To facilitate this latter type of usage, the illustrations have been labeled with the proper section number. The last section of each of the problem sets has been labeled "General" because the solutions of those problems may involve several of the techniques of that chapter. Such problems are very useful as review exercises.

The material of the first nine chapters meets the needs of the ordinary one-semester course. The remaining chapters, which have been called "Additional Topics," contain material for amplifying the basic course, if time permits, or for organizing a two-semester course. These five chapters are independent of each other, and therefore the instructor may choose whichever ones appeal to him. Since a number of the chapters have been expanded, particularly the last five chapters, and since a large number of new problems have been added, there is ample material for a two-quarter or a two-semester course. Many of these new

problems are of current interest and thus justify spending considerable time in discussing them, rather than pushing on to new material.

Many instructors who used previous editions were kind enough to send me suggestions for improvement. I have incorporated most of the suggestions. Even though I may not have included all of their ideas, I very much appreciate their willingness to help me write a better book.

Instructors can obtain answers to all of the problems in pamphlet form by writing to the publisher.

*Los Angeles, California*
*June 1975*

PAUL G. HOEL

# Contents

## CHAPTER 1
## The Nature of Statistical Methods



## CHAPTER 2
## The Description of Sample Data



## CHAPTER 3
## Probability

## CHAPTER 4
## Probability Distributions



## CHAPTER 5
## Some Special Probability Distributions



## CHAPTER 6
## Sampling

## CHAPTER 7
## Estimation



## CHAPTER 8
## Testing Hypotheses



## CHAPTER 9
## Correlation and Regression

# Additional Topics

## CHAPTER 10
## The Chi-Square Distribution



## CHAPTER 11
## Analysis of Variance



## CHAPTER 12
## Nonparametric Tests



## CHAPTER 13
## Multiple Regression

## CHAPTER 14
## Statistical Decision

# CHAPTER 1

# The Nature of Statistical Methods

## 1. INTRODUCTION

From a very general point of view, statistical methods are methods for treating numerical data. Thus, methods for collecting and analyzing the data of any business or government enterprise may be considered as belonging to the field of statistics. Such a definition, however, is much too broad in scope to be useful. It is necessary to restrict both the nature of the data and the reasons for studying them before such methods can rightfully be called statistical.

Statistical methods are concerned with data that have been obtained from taking observations, in the form of measurements or counts, from a source of such observations. For example, in studying the cost of medical services in a city, a small percentage of the inhabitants of the city would be selected and asked about their medical costs. Or, in studying public opinion on a controversial issue being debated in congress, a set of voters would be selected from across the country and asked their views concerning the issue.

Statisticians take observations of the type just described for the purpose of

drawing conclusions about the source of the observations. Thus, the medical costs of the selected individuals would be used to determine the general level of medical costs throughout the city. Similarly, the purpose of questioning only a small percentage of the voters on a controversial issue is to determine to a satisfactory approximation the opinions of all the voters on that issue.

The set of observations that is taken from some source of observations for the purpose of obtaining information about the source is called a *sample,* whereas the source of those observations is called a *population.* In view of the preceding discussion, *statistical methods may be described as methods for drawing conclusions about populations by means of samples.* The single word "statistics" is often used in place of statistical methods. Thus a student who is taking a course in statistics is taking a course in statistical methods.

At first glance, the foregoing definition may seem to be rather technical and contrary to the popular notion about statistics. For example, many business people look upon statistical methods as methods for collecting and summarizing business facts. The Federal Government employs a number of statisticians whose principal duty is to design efficient ways of collecting and summarizing various kinds of information. According to the preceding definition of statistical methods, these statisticians do not appear to be using statistical methods because they do not apply the information they have collected for drawing conclusions about the sources of the information. This viewpoint, however, does not take into account the fact that such information is gathered for the consumption of others who will use it to reach conclusions. Business concerns do not collect and summarize business facts just to admire the information obtained. They expect to use the information to make decisions, and whether or not they openly arrive at conclusions concerning the sources of the information the fact remains that they do make decisions on the basis of samples.

That part of statistical methods concerned with the collecting and summarizing of data is usually called *descriptive statistics.* The part concerned with drawing conclusions about the source of the data is called *statistical inference.* Since the ultimate objective is to make inferences, that is, draw conclusions, the descriptive part of statistics should be looked upon as a sort of preliminary to the main bout.

The use of statistical methods has increased remarkably in the last few decades, particularly in the biological and social sciences. Such methods have also proved very useful in various branches of the physical sciences and engineering. Because of this varied and strong interest, these methods have developed rapidly and have increased in complexity and diversity; nevertheless, many of the most important techniques are quite simple and are the same for all branches of application. Some of these universal methods are studied in this book. They should prove useful

both to students who wish to understand how simple experimental data are handled and to students who need this type of background for more advanced work.

## 2. ILLUSTRATIONS

This section describes a few problems of the type that statistical methods were designed to solve. It does not begin to cover the broad class of problems capable of being solved by statistical methods but rather illustrates a few of the simpler ones that can be solved by using only the methods developed in this book. One problem is of academic interest, whereas the others are typical real-life problems.

(a) The mayor of a large city wishes to know what his chances are for winning the primary nomination of his party to a vacancy in Congress. To obtain such information an organization engaged in promoting political campaigns agrees to take a poll of the voters of the state to determine the mayor's popularity. By using statistical methods, such an organization can decide how large a poll will be necessary in order to estimate, within any desired degree of accuracy, the percentage of voters who would favor the mayor over other leading candidates.

(b) A medical research team has developed a new serum it hopes will help to prevent a common children's disease. It wishes to test the serum. In order to assist the researchers in carrying out such a test, a school system in a large city has agreed to inoculate half of the children in certain grades with the serum. Records of all children in those grades are kept during the following year. On the basis of the percentages of those children who contract the disease during that year, both for the inoculated group and for the remaining half, it is possible to determine by statistical methods whether the serum is really beneficial.

(c) An industrial firm is concerned about the large number of accidents occurring in its plant. In the process of trying to discover the various causes of such accidents, an investigator considers factors related to the time of day. He collects information on the number of accidents occurring during the various working hours of the day, and by using statistical methods he is able to show that the accident rate actually increases during the morning and also during the afternoon. Further statistical studies then reveal some of the major contributing factors involved in those accidents.

One might be tempted to say that statistical methods are not needed in a problem such as this, and that all one needs to do is to calculate percentages and look at them to decide what is happening. If one has a large amount of

properly selected data, such decisions will often be correct; however the high cost of collecting data usually forces one to work with only small amounts and it is precisely in such situations that statistical methods are needed to yield valid conclusions.

(d) A merchandizing firm wishes to determine the size of an advertising budget that will maximize the profit for one of its products. It decides to pick out three widely separated market areas and carry out promotional programs of varying intensity and costs that it believes are appropriate for that product. In doing so, it will need statistical methods on optimization to plan the study and estimate the best size budget to use.

(e) An instructor of an elementary statistics course is having difficulty convincing some of his students that the chances of winning from a slot machine are just as good immediately after someone has won some money as after a run of losses. For the purpose of convincing them, he and several of his students perform the following experiment on a slot machine in Las Vegas. The machine is played for one hour, or until the combined resources of instructor and students are exhausted, whichever occurs first. A record is kept of the number of wins and losses that occur immediately after a win, together with the amounts won, and also of the number of wins and losses, and amounts won, immediately after a run of, say, five losses. With data of this type available, the instructor should be able to apply statistical methods to convince the skeptics of his wisdom in this matter. Since a run of bad luck might make it difficult to demonstrate this wisdom, unless the machine were played a long while, the instructor would be well advised to come amply supplied with cash. An experiment of this type should also convince the students that slot machines are designed to extract money from naïve individuals.

## 3. ESTIMATION AND HYPOTHESIS TESTING

An analysis of the preceding illustrations will show that they properly belong to the field of statistics because all are concerned with drawing conclusions about some population of objects or individuals and propose to do so on the basis of a sample from that population.

It may also be observed that these problems fall into two general categories. They are concerned either with estimating some property of the population or with testing some hypothesis about the population. The first illustration, for example, is concerned with estimating the percentage of the voters of the state

who favor the mayor. The second illustration is one of testing the hypothesis that the percentage of children contracting a disease is the same for inoculated children as for children receiving no inoculation. The third illustration considers the problem of testing the hypothesis that the accident rate for a population of workers is constant over the day. The fourth illustration is concerned with estimating profit as a function of the amount of money spent on advertising the product. The fifth illustration is one of testing the hypothesis that the average amount of money won from a slot machine after a run of losses is the same as after a win.

Most of the statistical methods to be explained in this book are those for treating problems of these two types, namely, estimating properties of or testing hypotheses about populations. Although there are other types of conclusions or decisions that can be related to populations on the basis of samples, the bulk of those made by statisticians falls into one of the two aforementioned categories, and therefore they make up most of the material in this book.

## 4. PROBABILITY

In the problem of estimating the percentage of voters who favor a candidate or issue, the solution will consist of a percentage based on the sample and a statement of the accuracy of the estimate, usually in a form such as "the probability is .95 that the estimate will be in error by less than 3 percent." Similarly, in problems involving the testing of some hypothesis the decision to accept or reject the hypothesis will be based on certain probabilities.

It is necessary to use probability in such conclusions because a conclusion based on a sample involves incomplete information about the population, and therefore it cannot be made with certainty. The magnitude of the probability attached to a conclusion represents the degree of confidence one should have in the truth of the conclusion. The basic ideas and rules of probability are studied in a later chapter; meanwhile it should be treated from an intuitive point of view. Thus the statement that the probability is .95 that an estimate will be in error by less than 3 percent should be interpreted as meaning that about 95 percent of such statements made by a statistician are valid and about 5 percent are not. In the process of studying statistical methods one will soon discover that probability is the basic tool of those methods.

Probability is an exceedingly interesting subject, even for those who have little liking for mathematics or quantitative methods. Many people enjoy some of the events associated with probability, if not the study itself; otherwise, how can one