MATHEMATICS OF STATISTICS

PART ONE

BY

JOHN F. KENNEY
University of Wisconsin

SECOND EDITION, SIXTH PRINTING

D. VAN NOSTRAND COMPANY, INC.

TORONTO NEW YORK LONDON

NEW YORK
D. Van Nostrand Company, Inc., 250 Fourth Avenue, New York 3
TORONTO
D. Van Nostrand Company (Canada), Ltd., 228 Bloor Street, Toronto
LONDON
Macmillan & Company, Ltd., St. Martin's Street, London, W.C. 2

First Published, May 1939
Reprinted January 1941, February 1942
July 1943, March 1944, June 1945

Second Edition, January 1947
Reprinted May 1947, September 1948, July 1950
August 1951, August 1952

PRINTED IN THE UNITED STATES OF AMERICA

Matri Meae

in Signum

Gratitudinis

PREFACE TO THE SECOND EDITION

Striking examples of the utility and scope of the science of statistics have occurred in recent years. As Professor Harold Hotelling remarks (*Annals of Mathematical Statistics*, vol. 11, 1940, pp. 457–470):

> Indeed it seems as if the exploitation of the business and manufacturing possibilities of statistical methods has only begun and that limitless further fields are coming into view.

The widespread use of statistical methods and the gratifying interest shown in the present book have made possible a second edition at this time. This opportunity has been used to polish and clarify certain portions of the text. For suggestions leading to the excision of obscurities I am indebted to many of my students and to a number of friends in other universities, particularly to Professors Irving W. Burr, John H. Curtiss, Henry Scheffé, Guy G. Speeker, and Howard E. Wahlert. Of course, full responsibility for any remaining errors or other defects is my own.

J. F. K.

PREFACE TO THE FIRST EDITION

The field of statistics is many sided and ranges over different levels. However, between the levels of clerical work at one extreme and mathematical research at the other extreme, there is a well-defined methodology, mathematical in nature, which underlies the specialized applications in the departments of economics, psychology, education, and biology.

This book is an elementary text dealing with the mathematics of statistics. Fortunately, a considerable part of the descriptive methodology of statistics can be understood by those having relatively little knowledge of college mathematics. Although no mathematics beyond the ordinary Freshman course in college algebra is required for a profitable reading of this text, a certain degree of mathematical maturity and intelligence is presupposed. To achieve the maximum success perhaps only the best of those students whose mathematical preparation is limited to the minimum prerequisite should be encouraged to study it. Occasionally, material is introduced to sharpen the interest and challenge the ability of the more advanced student without interrupting the main developments or discouraging those less mature.

In writing this book, considerable selection of material necessarily had to be made. The omission of certain topics will be noted in the table of contents. Judging from my own experience, and that of others, the theory of sampling cannot be taught satisfactorily at the level for which Part I is intended. At best only a superficial use of formulas could be hoped for. Consequently, I have elected to defer this subject to Part II where a systematic treatment can be given. With regard to time series analysis, Professor J. Neyman says in his *Lectures And Conferences On Mathematical Statistics* (p. 106),

> We start by trying to split each of the series into several parts, which we arbitrarily assume to be additive. One of these parts is the trend, which we estimate perhaps by fitting a low order parabola to the whole series available. The next part is the "business cycle." The third part is the "seasonal variation," which we frequently estimate by calculating moving averages. Finally, the remainder is considered to arise from random causes, and we concentrate on the question whether such a remainder in one of the variables is correlated

with that in some other. All this procedure seems to me very artificial and arbitrary. . . . In my opinion the whole problem of time series must be treated from a point of view that is quite different from the traditional one just described.

I concur in this opinion and I believe that no useful purpose would be served by drilling students in the traditional procedures.

Throughout the book the student is encouraged and stimulated to master fundamental principles and concepts. Essentially, the job of every statistician is to take hold of situations and disentangle them by the techniques of the science. Therefore, considerable emphasis is placed on technique. I have tried to develop in the student the ability to use symbolism creatively as a language. Numerous examples are given to clarify concepts and illustrate processes. Over two hundred exercises are included. It is intended that these exercises should be handled as in a mathematics course. No laboratory, so-called, is necessary.

Nowadays, no little importance is attached to motivation. I have constantly held in mind the necessity of making the subject interesting and stimulating to the beginning student. Nevertheless, I venture the opinion that the best motivation for intelligent students is the feeling that their teacher knows his subject.

In preparing the manuscript a large number of books and papers have been examined and perhaps leaned upon. No claim to originality is made except possibly in the matter of arrangement and pedagogical approach. Numerous references to the scholarly achievements of others are cited. It is hoped that the serious student will read some of these and thereby widen his perspective and enhance his interest.

In conclusion, I wish to express my deep appreciation to Professor Allen T. Craig and Dr. Mason E. Wescott who critically read the manuscript and made many suggestions for its improvement.

JOHN F. KENNEY

Evanston, Illinois.
April, 1939

CONTENTS

INTRODUCTION

CHAPTER I

FREQUENCY DISTRIBUTIONS

CHAPTER II

GRAPHICAL REPRESENTATION

CHAPTER III

AVERAGES

Contents

CHAPTER VII

CURVE FITTING

CHAPTER VIII

CORRELATION THEORY

MATHEMATICS OF STATISTICS

INTRODUCTION

1. Definition. The word *statistics* is used in at least two different senses. Construed as plural it refers to the systematic presentation of quantitative data. Used in a singular sense, the word *statistics* refers to the science which has for its object the classification and analysis of quantitative data so that intelligent judgments may be passed upon them.

It is usually clear from the context which meaning[1] is intended, although some persons prefer the expression *statistical methods* for this second meaning. Statistical methods are all those devices used in the collection and analysis of data. The *theory of statistics* is the exposition of statistical methods and is of a mathematical nature.

2. Scope. There used to be a widespread misapprehension that statistics is a branch of economics. As a matter of fact, statistical problems arise in many different fields — biology, economics, engineering, insurance, education, physics, and astronomy, as well as various branches of business. The exploration of certain aspects of nearly every field involves some phase of statistical theory. Indeed, certain types of statistical methodology may have almost unexpected applications — the discovery, for example, that the life of physical property[2] is governed by much the same statistical rules as govern the lives of human beings, and hence, that life tables may be applied to both. Physicists have discovered that many of the problems in the modern theory of the structure of the atom are essentially statistical in nature. In recent years industrial companies have placed an increasing reliance on statistical methods in controlling the quality of goods during manufacture.

Statistics as a science is making contributions to all the sciences. On the other hand, some sciences like biometry and physics have

[1] In addition to the two meanings given above, another has crept into the recent literature where reference is made to a *statistic*. This term will be explained later.

[2] *Life Expectancy of Physical Property* — E. B. Kurtz. Ronald Press.

contributed much in the development of statistics and its terminology. The following quotation from *Science* may appropriately be mentioned here:

> The extension of the scope of quantitative methods through the medium of statistical analysis is one of the most significant things going on in the scientific world at the present time.[1]

The importance of statistical method in present-day thinking has been well stated, as follows:

> More and more the modern temper relies upon statistical method in its attempts to understand and to chart the workings of the world in which we live. Particularly in those sciences which deal with human beings, whether in their physical and biological aspects or in their social, economic, and psychological relations, the spirit of our time asks that its conclusion be based not so much upon the distinctive reactions of one or two individuals as upon the observation of large numbers of individuals, the measurement of their common likenesses and the extent of their diversity. As the data thus gathered from mass phenomena become extensive, it becomes imperative to have methods of organization to bring the facts within the compass of our understanding, methods of analysis to make the essential relations appear out of the mass of detail in which they are hidden, and methods of classification and description to facilitate the presentation of the data for the study and consideration of other persons. Thus statistical method becomes a telescope through which we can study a larger terrain than would be accessible to our unaided vision.[2]

3. Statistical Methods in the Social Sciences. Because statistics is fundamentally the study of aggregates of individuals, rather than of individuals, whether these *individuals* be observations or measurements or persons, it is apparent that statistical methods are essential to social studies. Indeed it has been said that it is principally by the aid of such methods that these studies may be raised to the rank of sciences.

This particular dependence of social studies upon statistical methods is mentioned in a recent book [3] from which we quote the following:

> If, as seems probable, our present uncoördinated large-scale business is to be further developed into an efficiently managed instrument of production serving the needs of the people, then statistics, together with mathematical economics, will emerge among the most important tools of the social sciences. For it is by

[1] *Science.* January 18, 1929.

[2] *Mathematics and Statistics* — Walker. Sixth Yearbook, National Council of Teachers of Mathematics.

[3] Reprinted by permission from *Methods of Statistical Analysis* by Davies and Crowder, published by John Wiley and Sons, Inc.

means of averages, dispersions, coefficients of variability, trends, and regressions, as pictured in control charts, that management is able to visualize and direct the movements of large masses of population.

The work of the statistician is much like that of the map maker who presents the traveler with a sketch of important highways, showing the locations of towns and geographical features. The map is not a picture of reality. It shows cities as dots, and rivers as lines. It has purposely omitted the interesting details of scenery and the still more important features of human interest which lie along the route and which constitute the traveler's real objectives. Nevertheless, as a means of reaching these objectives, the map is extremely useful. And so it is with statistics in the hands of the business executive and statesman. Back of the charts are human beings with their varying characteristics and vital interests, few of which can be described in figures. Yet as a means of serving these interests, of keeping trade moving from one region to another, of allocating investment and labor, and of apportioning relief to maladjusted industries and dependent classes, statistics and mathematical methods are important, and are becoming increasingly important with the growing complexity of society.

It may be said that the study of statistics is not merely an attempt to describe what actually occurs, though it must begin at this point, but in its broader aspects it is the logical background of business and social management. Hence what appears now to be mere abstraction may later become the basic necessity of an applied science. Eventually, it may be assumed, the social arts of business and politics will rest upon as substantial a theoretical and mathematical background as physics, chemistry, and engineering.

4. Mathematics and Statistics. Statistical problems are of interest, therefore, not only to the worker in the particular field but also to the mathematician, inasmuch as methods adequate to the treatment of these problems can best be presented in the precise and accurate language of mathematics. Moreover, statistical methods are grounded in statistical theory which is a branch of applied mathematics.

Although it is true that some statistical problems are ultimately problems in advanced mathematics, many of which mathematicians have not yet been able to solve, nevertheless a large and interesting part of statistical analysis requires mathematics no more advanced than elementary algebra.

It has been said that sooner or later every true science tends to become mathematical. The notation of mathematics is simply a language and it is not limited to any particular field of knowledge. The following quotations are inserted to help the student approach the study of statistics in the proper spirit.

1. Mathematics, the science of the ideal, becomes the means of investigating, understanding, and making known the world of the real. — White.

2. Probably among all the pursuits of the university, mathematics preeminently demands self-denial, patience, and perseverance. . . . — Todhunter.

3. From time immemorial, there has been but one way to become a mathematician and there will never be another: it is a way interior to the subject and involves years of assiduous toil. Short-cuts to mathematical scholarship there are none, whether the seeker be a philosopher or a king. — Keyser.

4. Will is the creative force. Without the will to learn there is no learning. And when the will is feeble and confused, learning lags. — Mursell.

5. The theory of statistics is not easy, not so much because it is abstruse, as because the ideas are new to most people, and a good deal of hard thinking and patient work will be necessary. . . . Statistical work always involves a lot of computing [and] there is no better way of learning statistics than by working through examples. — Tippett.

5. Problem Assignments. The student should realize at the outset that statistical methods are not substitutes for thinking but are aids and supplements to it. A superficial knowledge of statistical technique cannot take the place of good judgment. Mere ability to substitute in formulas should not be confused with genuine statistical sophistication and insight. To the serious and capable student who intends to master this course, formulas will be a set of functioning concepts and tools rather than machines into which material may be fed to grind out a meaningless answer.

This opportunity is also taken to point out that even mathematical discourse consists of sentences. Punctuation should not be omitted in sequences of equations and other mathematical statements. (It is admitted, however, that many of us find this difficult to remember.)

Throughout the book exercises are inserted to give the student an opportunity to test his knowledge of the theory and methodology, and to develop his power of analysis. In grading the solutions, value will be attached to accuracy, thoroughness, neatness, and systematic arrangement of the work.

6. Calculating Machines.[1] A full description of the parts of a calculating machine and their operation may be obtained from an *Instruction Book* which is furnished by the manufacturer, so only a brief description will be given here.

A calculating machine is constructed to add and substract. By means of continued addition or subtraction, operations involving multiplication, division, and square root can also be performed with great speed.

[1] The early history of modern computing machines is outlined in the *American Mathematical Monthly*, vol. 31 (1924), pp. 422–429.

In addition to a keyboard on which numbers can be punched, most machines have a sliding carriage, carrying two dials one above the other. These dials are called *revolution register* (upper dial) and *product register* (lower dial). In finding a product nx, one of the factors n is punched on the keyboard and as the motive crank at the side is turned,[1] the other factor x appears on the upper dial. The product nx is then read from the lower dial.

An important property of the modern calculating machine is its adaptability to short cuts and combinations of operations. For example, one may multiply two numbers nx together and add the result to a third number k without tabulating the intermediate steps. This is accomplished by punching the number k on the keyboard, transferring it to the lower dial (product register), and then proceeding as in finding the product nx. The result $nx + k$ is then read from the lower dial. An extension of this procedure is especially useful in a series of computations where k and n are constant and various values are assigned to x. To describe the procedure, suppose it is required to calculate the successive values of $12 + 6x$ for $x = 5, 7, 15, 12$, etc. The number $k = 12$ is first registered on the lower dial, then the factor $n = 6$ is placed on the keyboard, and by turning the crank forward five times to make the first value of $x = 5$ appear on the upper dial, the result $12 + 6 \times 5$ appears on the lower dial. Instead of clearing the dial, the crank is now turned forward twice more to rebuild the value $x = 5$ into $x = 7$, and the result $12 + 6 \times 7$ can be read from the lower dial. In rebuilding $x = 15$ into $x = 12$ the crank is turned backwards. This procedure can be repeated until all the required values of $12 + 6x$ have been calculated. A process of this sort is called the *continuous method* of calculating.

In most of the exercises in this course, the computations are not laborious and calculating machines are not required. However, if machines are available they may be used to advantage in Chapters IV and VI. The student who desires to develop skill on a calculating machine should begin now to study an *Instruction Book* and practice the fundamental operations explained there.

7. Collateral Reading. Perhaps no single textbook can meet all the needs of all students of statistics. There are several good books on elementary statistics which, although not fundamentally different,

[1] The beginner will probably wish to practice on a manually operated machine before attempting to use the high-speed electric and automatic machines.

present different points of view on certain topics and treat them with varying degrees of emphasis depending upon the field of major interest. At least some of the books listed below should be readily available on the reserve shelf of the library. The list should be useful to those who wish to study more fully certain details in which they may be interested.

1. Bivins — *The Ratio Chart in Business.* Codex Book Co.
2. Burgess — *The Mathematics of Statistics.* Houghton Mifflin and Co.
3. Camp — *The Mathematical Part of Elementary Statistics.* D. C. Heath and Co.
4. Deming — *Statistical Adjustment of Data.* John Wiley & Sons, Inc.
5. Freeman — *Industrial Statistics.* Wiley.
6. Garrett — *Statistics in Psychology and Education.* Longmans, Green and Co.
7. Glover — *Tables of Applied Mathematics.* Wahr.
8. Haskell — *Graphic Charts in Business.* Codex Book Co.
9. Mills — *Statistical Methods, Revised.* Henry Holt and Co.
10. Pearl — *Medical Biometry and Statistics.* W. B. Saunders and Co.
11. Rider — *Statistical Methods.* Wiley.
12. Scarborough — *Numerical Mathematical Analysis.* The Johns Hopkins Press.
13. Snedecor — *Statistical Methods.* Collegiate Press, Inc., Ames, Iowa.
14. Treloar — *Statistical Reasoning.* Wiley.
15. Walker — *Elementary Statistical Methods.* Holt.
16. Yule and Kendall — *The Theory of Statistics.* Griffin and Co.