PRINCIPLES AND METHODS OF STATISTICS

BY

ROBERT EMMET CHADDOCK

Professor of Statistics, Columbia University

HOUGHTON MIFFLIN COMPANY

BOSTON · NEW YORK · CHICAGO · DALLAS · SAN FRANCISCO

The Riverside Press Cambridge

COPYRIGHT, 1925

BY ROBERT EMMET CHADDOCK

ALL RIGHTS RESERVED

The Riverside Press

CAMBRIDGE · MASSACHUSETTS

PRINTED IN THE U.S.A.

TO

R. A. C.

PREFACE

THIS book is an introductory text for general use in colleges and universities. The object is not to contribute new results of statistical analysis, nor to advance the knowledge of the author's associates in this field. The purpose is rather to introduce the subject to beginners by the use of simple illustrative materials, to present the elementary principles of statistics in such manner as to appeal to the logical faculties of the reader, and to foster a healthy skepticism toward the results of quantitative investigations.

Elementary economics and social science are being taught to sophomores in college. Professional training begins in the last two years of the undergraduate course. Special courses include a varied background of facts which are intended to describe and to measure the phenomena of organized society. Many sources of information are in quantitative form, but figures do not necessarily tell the truth. Training in the scientific method of approach to a problem; acquaintance with the tools needed in his work; and familiarity with the methods of testing how reliable so-called facts really are — these are matters of fundamental importance to the student of the social sciences, theoretical and applied.

Technique and method in the physical sciences have been well developed and are applied in the experimental laboratory by the beginning student. In the social sciences many students are still denied the opportunity to acquire a knowledge of the methods of assembling data and of sifting evidence. They have not formed the habit of independent and cautious generalization from known facts. Statistics may serve a purpose for the social sciences similar to that which the experimental laboratory serves in the physical sciences. The possibilities for experimentation in society are very limited. The investigator must observe, record and compare phenomena, for the most part, as they are. Statistical method deals with mass data in their numerical aspects. It involves precise measurement, a careful record, intelligent and logical analysis and grouping, and discriminating judgment as to the relative significance of groups of facts.

Training in statistics should be regarded as discipline in scientific method. The undergraduate should be equipped with statistical method as a tool. Especially in the social sciences, where factors are many and varied and where human sympathies are keen, it is difficult to free judgments from the bias of desires and feelings. The scientific

attitude of mind prompts the student to seek evidence which will appeal as true to minds other than his own. Quantitative data are objective and can be freed from bias more easily than subjective judgments.

The point of view from which the essential principles of quantitative analysis, description and comparison are approached in this treatise is logical and empirical rather than mathematical. A large number of technical and mathematical books on statistical methods have appeared in recent years. The author recognizes the value of these for the highly specialized student but doubts their utility for the beginner in the subject. The need at present is a clear understanding of fundamental principles in the treatment of numerical data.

Short methods have only a limited use in an elementary text since they may obscure for the beginner the meaning of the process which should be the all important consideration. Formulæ are regarded as short-hand statements of processes described and illustrated. The same problems are often carried forward from chapter to chapter, as the principles are developed, since by this plan the reader is freed from the necessity of becoming familiar with a new series of facts at a time when his attention should be concentrated upon the treatment of the materials.

Understanding in this subject results, for the most part, from doing. To find any value in the methods presented the student must make them a part of his experience and must be able to apply them with discrimination. Practice work, therefore, is the essential thing in acquiring statistical methods and in developing intelligence in their application. This book presents only a limited number of problems with detailed explanations of the methods applied to the data. In the author's opinion, it is the function of the teacher to select exercises for practice. In order to hold the interest of the student and to secure the best results the teacher will prefer to work out particular applications which interest himself and his students in specific fields. He will be able to select better materials for illustration than the author of a general text can possibly assemble for him. Besides, following a set of exercises prescribed by others emphasizes too strongly the mere routine of acquiring methods.

Most texts treat the methods described in Part III before those of Part II. This is the logical order in the collection and analysis of statistical data, and the teacher is at liberty to adopt this plan of presentation. The order of topics in this text has been adopted after trying out both plans in teaching, for the reason that very few have had experience in the actual collection of the original raw materials while most students have used to some extent existing sources of statistical data and possess a more or less extensive background of quantitative information. By

此为试读，需要完整PDF请访问：www.ertongbook.com

treating at once methods of classifying and analyzing data already assembled the teacher builds upon familiar materials and is able to introduce the methods of Part III as the opportunity arises.

The scope of the book has been carefully limited to include only what can be covered in a year's elementary instruction and practice. The continuation of the subjects treated and the introduction of specialized methods are left to more advanced treatises.

The author has received invaluable suggestions from numerous text books and writers, for which he wishes to make acknowledgment and to express appreciation. An effort has been made to give full credit in footnotes for materials and ideas utilized in this text, but no such statements can adequately describe the debt we owe to the workers who have made their contributions to the methods of quantitative analysis now in general use.

Special acknowledgment is gladly given to Warren M. Persons and to the Harvard Committee on Economic Research for materials used in Chapter XIII; to Seymour L. Andrew, Chief Statistician of the American Telegraph and Telephone Company, for permission to use for illustration a section of their curve representing business conditions; to Willard C. Brinton and to the Society of Mechanical Engineers for permission to reprint the publication on Standards for Graphic Presentation; and to the authors of Bulletin 165, Agricultural Experiment Station, University of Illinois, for illustrative materials in Chapter XI.

I am especially grateful to my colleagues for their helpful criticisms and suggestions and for their generous spirit of coöperation: to Franklin H. Giddings, leader in the scientific study of society, for his sympathetic interest in the entire work and for a critical reading of Chapter III; to Henry L. Moore for reading the entire manuscript of Parts I and II and for discussing with me many of the topics presented; to Frank A. Ross, who has been closely associated with me in the work of statistical instruction for several years; to Wesley C. Mitchell for the use of materials and ideas from Bulletin 284 of the United States Bureau of Labor Statistics on Index Numbers, and for reading the manuscript of Chapter X; and to Russell G. Smith for assistance in revising Part III.

The author is indebted also to Mrs. Dorothy Reddy Van Der Veer for her excellent work in the final revision and preparation of the manuscript, and to Mrs. Charles A. Gulick for her assistance; to Mr. and Mrs. Roy E. Stryker for checking arithmetical computations and especially for their painstaking care in the preparation of the original drawings for the diagrams. He is under special obligation to Miss Estella T. Weeks for a discriminating reading of the proofs of the book.

ROBERT E. CHADDOCK

EDITOR'S INTRODUCTION

In spite of a widespread human interest in numerical facts, statistics, not long ago, was a subject which had only a small, though loyal, following. In recent years, the rapid progress that has been made in the prosecuting of statistical inquiries and the perfecting of statistical technique has been matched by an equally rapid general growth of interest in statistical studies. To this the increasing number of excellent introductory textbooks available to the student of statistics bears witness. It is no disparagement of other books to say that many teachers and students will find that Professor Chaddock's work will best meet their own particular needs.

It is the product not only of sound scholarship but also of long and conspicuously successful experience in the teaching of statistics. Throughout the book Professor Chaddock is the teacher as well as the statistician. This shows itself in the general plan and arrangement of the book, in the apportioning of space and of emphasis, and in the painstaking care given to every detail of exposition. It will be clear, furthermore, that Professor Chaddock has carefully appraised the needs, the purposes, and the initial equipment of the average student. For one thing, no mathematical knowledge is assumed, beyond the most elementary processes of algebra.

The book has another conspicuous merit. It treats of statistical methods as general rather than special tools. This, in my opinion, is as it should be. The student of vital statistics, for example, will be better equipped for his work if he knows something about the general uses, possibilities, and limitations of statistical method. So also, to take another example, with the student whose interest is primarily in the light which statistical analysis may throw upon business problems and policies. In particular, it is a mistake to put a fence around a narrow field, and dub it "economic statistics," or "social statistics," as the case may be. Students of economics are likely to profit quite as much by studying population statistics as by studying index numbers. And who will venture to say just what portions of the field that is coming to be called "business statistics" are not of importance to the student of economics? Professor Chaddock nowhere permits his interest in any one field of inquiry to obscure the importance of other uses of statistical method. In this respect, as in others, his book displays judgment and balance.

Whether statistics is a field of knowledge or a method is an old problem respecting which there was once much fruitless dispute, especially in German treatises. The dispute was fruitless because the issue was factitious. Statistics is not a field of knowledge so much as it is a particular type of knowledge. To organize this type of knowledge, to handle it effectively, calls for a special technique. The statistical method, then, is merely the utilizing of this special sort of knowledge, with the aid of this special technique, in any field of inquiry within which the method may be fruitful. The method, therefore, has a unity of its own, and this is reflected in Professor Chaddock's treatment of it.

The successful use of the method, however, requires a command of more than formulas and processes. Like any other scientific pursuit it calls on the one hand for the power of constructive imagination, and on the other hand for constant awareness of the necessity of guarding against mistaken inferences. In statistics the most dangerous sources of error are of a very elementary sort. There is just enough difference between statistical processes and the common types of reasoning that serve us well enough in the daily routine of life to make the statistical method a dangerous tool in the hands of the unwary. Professor Chaddock has done all that any writer could be expected to do, and more than most writers have done, to put students on their guard against the commonest forms of statistical error and to imbue them with the habit of critically scrutinizing their sources, their processes, and their inferences.

Throughout the book, moreover, he puts a needed emphasis upon analysis, as contrasted with the routine of statistical technique. I mean that he insists that the student should *understand* every element in the formulas he uses and every step in the processes he applies, and that he does all that he can to help the student toward such an understanding.

It is here that Professor Chaddock has been most lavish in his expenditure of care and pains. The impatient student, very likely, will find some of this detail irksome. But the student with special faculty in mathematics or with a special aptitude for statistics will find nothing that will delay or hinder his progress. And for the great majority of students mastery of detail is the one sure road to a mastery of the subject.

ALLYN A. YOUNG

CONTENTS

PART I

INTRODUCTION — PRELIMINARY CONSIDERATIONS

CHAPTER I. THE APPEAL TO FACTS 3

The basis of social action — Social legislation based on facts — Vital statistics in the service of sanitary science and health administration — Statistics and court decisions — Statistics and modern business — The modification of old theories by new facts — Theories of government — SUMMARY.

CHAPTER II. MISUSES OF STATISTICAL DATA 10

The importance of method — COMPARISON OF NON-COMPARABLE DATA — Criminality among the foreign-born — Death-rate from disease among American soldiers in the Philippines — Comparison of coal-mine accidents in different countries — The measurement of unemployment — Incomparability of data in the statistical sources of different States — Tuberculosis among men and women in the garment trades — Comparative size of the family in two generations — WRONG USES OF PERCENTAGES — Need for both absolute numbers and percentages — Percentages as proportions of a total of one hundred — Averaging or combining percentages — NEGLECT OF IMPORTANT FACTORS — Smoking and failure in college — Physical defects and repeaters in school — THE BIASING INFLUENCE OF PREVIOUS CONVICTIONS — Labor's share as described by the agitator — Vaccination and its Enemies — AN ILLUSTRATION FROM AN OFFICIAL SOURCE — Criticism of the method and scope — SUMMARY.

CHAPTER III. STATISTICS IN THE SERVICE OF SCIENCE 23

Essentials of scientific method — The impersonal character of scientific conclusions — The testing of premises — The scientific explanation of the spread of yellow fever — Induction and deduction — Statistics, a method of study — Descriptive and scientific statistics — The case method — The statistical method — The non-mathematical aspects of statistics — STATISTICS IN THE SERVICE OF THE SOCIAL SCIENCES — The experimental method — Limitations on the experimental method in the social sciences — Dependence of the social sciences upon statistical method — THE SERVICE OF STATISTICS TO ECONOMICS AND SOCIOLOGY — STATISTICS IN THE SERVICE OF EDUCATION — SUMMARY.

PART II

CLASSIFICATION AND DESCRIPTION OF MASS DATA

CHAPTER IV. CLASSIFICATION OF STATISTICAL DATA 43

First steps in statistical analysis — Simplification of data — The basis of valid comparison — Attributes and measurable characteristics — Types of series in the classification of quantitative data — Adaptation of methods to the specific type of series — SUMMARY.

CHAPTER V. THE FREQUENCY DISTRIBUTION 53

CLASSIFICATION OF QUANTITATIVE DATA — The array — Magnitude classes — The frequency distribution — Advantages of grouping — Assumptions underlying the formation of frequency classes — The size of the class-interval — Importance of uniform class-intervals — THE GRAPHIC REPRESENTATION OF AN ARRAY AND A FREQUENCY DISTRIBUTION — THE EFFECT OF THE SIZE OF THE CLASS-INTERVAL UPON THE POLYGON AND HISTOGRAM — The characteristic distribution of heights — THE BELL-SHAPED SYMMETRICAL CURVE — CONTINUITY OF VALUES ALONG THE SCALE — Discrete data — Continuous data — Suggested experiment — POSITION OF THE LIMITS OF THE CLASS — The effect of inaccurate grouping.

CHAPTER VI. METHODS OF SUMMARIZATION AND DESCRIPTION — THE ARITHMETIC AVERAGE 81

THE CONCEPT OF THE AVERAGE AS EMPLOYED IN EVERYDAY EXPERIENCE — THE NATURE OF AN AVERAGE — The general functions of an average — I. A summation of values — A warning in comparing averages — The importance of homogeneity — II. A type value from which to measure variability — Grouping about a central value — THE KINDS OF AVERAGE — THE ARITHMETIC AVERAGE — THE MEAN — Computation from ungrouped data — Computation of the mean from grouped data — THE WEIGHTED ARITHMETIC AVERAGE — Combining relative quantities — Combining averages — SHORT METHOD OF COMPUTING THE MEAN — Short method — Short method when the intervals are not uniform — CRITERIA FOR JUDGING AN AVERAGE — Limitations of the mean — Effect of extreme variants upon the mean — SUMMARY.

CHAPTER VII. METHODS OF SUMMARIZATION AND DESCRIPTION — THE MEDIAN AND THE GEOMETRIC MEAN 107

The median – a position average — The effect of extreme variants upon mean and median — METHODS OF DETERMINING THE MEDIAN VALUE — Ungrouped data — The proportion method for grouped data — The location of the median exemplified in detail — The quartiles and their determination — The significance of the position of the quartiles — The deciles and percentiles — The graphic method of locating median and quartiles — USE OF THE CUMULATIVE FREQUENCY DISTRIBUTION — The use of cumulative frequency curves for interpolation — THE EFFECT OF DIFFERENT SIZED CLASS-INTERVALS UPON MEDIAN AND QUARTILES — Comparison of the mean and median weight — Computation from an odd number of cases — APPLICATIONS OF THE MEDIAN — SUMMARY OF THE MEDIAN AVERAGE — THE GEOMETRIC MEAN — Computation of the geometric mean — The use of the geometric mean to measure the rate of population growth.

CHAPTER VIII. METHODS OF SUMMARIZATION AND DESCRIPTION — THE MODE 128

Possibility of more than one mode — THE DETERMINATION OF THE MODE — Indefinite location of the mode in grouped data — Smoothing the distribution by a moving average of frequencies — The relative positions of mean, median, and mode — An empirical rule for locating the mode — Another method of locating the mode within the class — The "true mode" in continuous data — APPLICATIONS OF THE MODE — Wages and hours — Prices — Population statistics — The mode in the analysis of industrial accident data — CAUTION IN THE USE OF THE MODE — The mode in time series — The mode in geographic series — SUMMARY.

CHAPTER IX. VARIATION AND ITS MEASUREMENT 150

The importance of variability — Degree of homogeneity or similarity in a series — MEASUREMENT OF ABSOLUTE VARIABILITY — The range as a measure of variation — The semi-interquartile range (Q) — The average deviation ($A.D.$) — The proper average from which to measure average deviation — Computation of the average deviation by a short method — The standard deviation (σ or sigma) — Computation of the standard deviation — The subtraction of the correction factor — The class-interval as a fraction — The class-interval as a single unit — RELATIONS OF THE THREE ABSOLUTE MEASURES OF DISPERSION — Ratio of one measure of variability to another — MEASURES OF RELATIVE VARIABILITY — The importance of the zero origin in graphic representation — The coefficient of variation (V) — Applications of the coefficient of variation — CONCLUSIONS AS TO MEASURES OF DISPERSION — MEASURES OF SKEWNESS OR ASYMMETRY.

CHAPTER X. INDEX NUMBERS — AN APPLICATION OF STATISTICAL METHODS 175

Changes in real wages – an illustration — REASONS FOR INTRODUCING INDEX NUMBERS AT THIS POINT — A LEVEL OF PRICES — A VARIETY OF POSSIBLE USES FOR INDEX NUMBERS — STEPS IN THE CONSTRUCTION OF A PRICE INDEX — NUMBER AND KIND OF COMMODITIES — COLLECTION OF PRICE DATA — METHODS OF COMBINING PRICE DATA — Method of averaging relatives — Shifting the base period for comparison — An advantage of the geometric average — Method of comparison of aggregate values — CHOICE OF THE BASE PERIOD — Relating chain indexes to a fixed base — The choice of a fixed-base period — RELATIVE IMPORTANCE OF COMMODITIES – WEIGHTING — Bias resulting from weighting — Changing importance of commodities — THE USE OF AVERAGES IN CONSTRUCTING INDEX NUMBERS — THE IDEAL FORMULA RECOMMENDED BY PROFESSOR FISHER — GRAPHIC METHOD OF SHOWING PRICE CHANGES — SUMMARY.

CHAPTER XI. MEASURES OF UNRELIABILITY — PROBABLE ERROR 207

DESCRIPTION OF A FREQUENCY DISTRIBUTION — IDEAL FREQUENCY CURVES — CAUSES OF VARIATION — Errors and their characteristics — Kinds of errors classified — PROBABILITY AND THE SYMMETRICAL BELL-SHAPED CURVE — The work of Gauss and LaPlace — THE SIGNIFICANCE OF PROBABILITY — Mathematical characterization of probability — Compound events — The more complex problem — The probability curve — Mathematical description of a curve — The straight line — Description of a circle — The symmetrical bell-shaped curve — PRACTICAL USES OF THE BELL-SHAPED CURVE — The ideal curve superimposed upon actual data — An example of fitting the bell-shaped curve — DESCRIPTION OF THE DISTRIBUTION OF ERRORS — Division of the area under the probability curve — The overlapping of frequency curves — UNRELIABILITY AS MEASURED BY THE BELL-SHAPED SYMMETRICAL CURVE — An experiment with the heights of 1000 freshmen — The unreliability of an average — Use of the probability curve to describe unreliability — The necessity for a shorter method — FORMULÆ FOR COMPUTING THE UNRELIABILITY OF MEASURES OBTAINED FROM RANDOM SAMPLES — The unreliability of a mean — The unreliability of a standard deviation — The unreliability of a difference between two measures — The unreliability of a coefficient of correlation — Unreliability measured in terms of probable error — Degrees of probability — GAINS IN WEIGHT – AN APPLICATION OF PROBABLE ERROR — PROBABLE ERROR AND REPRESENTATIVENESS OF A SAMPLE — Probable error as a warning — THE IMPORTANCE OF THE SIZE OF THE SAMPLE.

CHAPTER XII. DISCOVERY AND MEASUREMENT OF RELATIONSHIPS — CORRELATION 248

NATURE OF THE INQUIRY INTO CAUSE AND EFFECT RELATIONS — LIMITATIONS OF THE EXPERIMENTAL METHOD IN THE SOCIAL SCIENCES — MEASURING DEGREES OF RELATIONSHIP — THE SCATTER DIAGRAM — Representation of vertical and horizontal distributions by averages — A DOUBLE-ENTRY TABLE OF RELATED VALUES – THE CORRELATION TABLE — Representation of the distributions in columns and rows by mean ages — THE GRAPHIC REPRESENTATION OF RELATIONSHIP — The equation of a straight line passing through zero origin — Fitting the straight line in the correlation diagram — Measurement of the degree of association — The standard deviation – a common denominator — A more complete description of the line of relationship (RR_1) — A second line of the means of the correlation table — COMPUTATION OF r FROM THE CORRELATION TABLE – PEARSON'S FORMULA — The product-deviations ($d_x d_y$) from the table — An algebraic sum — THE UNRELIABILITY OF THE COEFFICIENT OF CORRELATION — Significance of a coefficient — THE PREDICTION EQUATIONS FOR Y AND X — Limits of error in the use of the prediction equations — NEGATIVE OR INVERSE CORRELATION — The regression equations of Y on X — Locating the regression line of Y on X — Measure of scatter about the regression line of Y on X — CORRELATION OF UNGROUPED DATA — NON-LINEAR RELATIONSHIPS BETWEEN TWO VARIABLES — A more general measure of correlation — The relation of η to r — METHODS OF MEASURING THE DEGREE OF ASSOCIATION FROM THE RELATIVE POSITIONS OF THE VALUES — The method of correlation from ranks — An illustration of the computation of ρ — INTERPRETATION OF THE COEFFICIENT OF CORRELATION.

CHAPTER XIII. THE TIME SERIES — HISTORICAL STATISTICS 306

CHARACTERISTICS PECULIAR TO THE TIME SERIES — THE DETERMINATION OF SECULAR TREND — THE MOVING AVERAGE — The moving average as a zero line from which to measure short-time fluctuations — Correlation of the short-time fluctuations about the trends of two series — Limitations of the moving average — FITTING A STRAIGHT LINE TO HISTORICAL DATA DESCRIBING THE SECULAR TREND — Regression of Y on X – the secular trend — The secular trend, 1903–1914 — Annual values on the trend line — An abbreviation in the method of fitting the trend line — DEVIATIONS FROM THE TREND EXPRESSED IN COMPARABLE UNITS – PERCENTAGES – STANDARD DEVIATIONS — Deviations in terms of standard deviation units — A COMPOSITE CURVE SHOWING SHORT-TIME FLUCTUATIONS — USE OF MOVING AVERAGES TO DESCRIBE TRENDS OTHER THAN STRAIGHT LINES — TESTING FOR LAG AND DETERMINING ITS AMOUNT — SEASONAL VARIATIONS – MONTHLY DATA — AN INDEX OF SEASONAL VARIATION — I. Method of simple monthly means — II. Method of a twelve-month moving average — A method based upon the moving averages — III. Method of link relatives — Median link relatives in row (A), Table 76 — Chain relatives in row (B) — Adjustment of chain relatives in row (C) — The seasonal indexes in row (D) — IV. A simpler method — APPLICATION OF THE INDEX OF SEASONAL VARIATION — Elimination of seasonal fluctuations — CYCLES EXPRESSED IN UNITS OF σ — IMPORTANCE OF GRAPHIC COMPARISON — CORRELATION OF SERIES EXPRESSED IN UNITS OF σ — MEASURING THE LAG FROM MONTHLY DATA — Utility of the measurements of lag in forecasting — A problem proposed — SUMMARY.

PART III

THE GATHERING AND PRESENTATION OF STATISTICAL DATA

CHAPTER XIV. COLLECTION OF STATISTICAL DATA 371

A PRELIMINARY STUDY OF THE PROBLEM — A WORKING HYPOTHESIS — A CLEARLY DEFINED STATISTICAL UNIT — METHODS OF GATHERING DATA IN A FIRST-HAND INQUIRY — 1. Personal inquiry — 2. Enumerators or special agents — 3. Questionnaires - Schedules filled out by informants — Registration of births and deaths — 4. Method of estimates — A REPRESENTATIVE INVESTIGATION — Sampling in a representative investigation — Review of certain aspects of sampling — The meaning of random selection — Practical test for the size of the sample — THE SCHEDULE OF QUESTIONS OR ITEMS — General considerations in preparing a schedule — Important rules in determining the items or questions — THE USE OF DATA GATHERED BY OTHERS — Primary and secondary sources — Guiding principles in the use of sources — SUMMARY.

CHAPTER XV. PRESENTING STATISTICAL DATA IN TABLES 397

EDITING RETURNS ON SCHEDULES — The function of tabulation — ACCESSIBILITY OF THE ORIGINAL SCHEDULES — THE TAKING OF THE DATA FROM THE SCHEDULES — KINDS OF TABLES — CONSIDERATIONS IN THE CONSTRUCTION OF TABLES — IMPORTANCE OF CHECKING ACCURACY IN TABULATION — EXPLANATION AND INTERPRETATION FROM TABLES — The government statistician — COMPARABILITY OF TABLES — SUMMARY.

CHAPTER XVI. GRAPHIC REPRESENTATION 418

THREE WAYS OF PRESENTING DATA ON FAMILY EXPENDITURES — USES AND ADVANTAGES OF GRAPHIC REPRESENTATION — ILLUSTRATIONS OF THE FAULTY CONSTRUCTION OF GRAPHS — 1. Omission of a zero base line — 2. The comparison of rectangular or bar diagrams — 3. Comparison of circles of varying size — 4. Wrong impressions arising from the choice of scales — 5. The use of pictorial representations — 6. Quantities entered within the field of the diagram — THE FIELD UPON WHICH A GRAPH IS DRAWN — RECTANGULAR COÖRDINATES — TYPES OF DIAGRAMS — 1. Line diagrams — 2. Area diagrams — 3. Maps — 4. The ratio diagram — STANDARDS FOR GRAPHIC PRESENTATION.

APPENDIX A 447

Alphabetical list of references.

APPENDIX B 453

Summary of symbols, equations, and formulæ.

APPENDIX C 462

Ordinates of the normal probability curve expressed as fractional parts of the mean ordinate y_o^N.

APPENDIX D 463

Fractional parts of the total area (10,000) under the normal probability curve.

APPENDIX E 464

Values of r for corresponding values of ρ.

INDEX 465

PRINCIPLES AND METHODS OF STATISTICS

PART I

INTRODUCTION

PRELIMINARY CONSIDERATIONS

The scientific study of any subject is a substitution of business-like ways of "making sure" about it for the lazy habit of "taking it for granted" and the worse habit of making irresponsible assertions about it. To make sure, it is necessary to have done with a careless "looking into it" and to undertake precise observations, many times repeated. It is necessary to make measurements and accountings, to substitute realistic thinking (an honest dealing with facts as they are) for wishful or fanciful thinking (a self-deceiving day-dreaming) and to carry on a systematic "checking up" . . . science is nothing more nor less than getting at facts, and trying to understand them.

FRANKLIN H. GIDDINGS

"Societal Variables," *The Journal of Social Forces*, March, 1928

PRINCIPLES AND METHODS OF STATISTICS

CHAPTER I

THE APPEAL TO FACTS

The basis of social action. An increasing number of communities are seeking exact information concerning themselves. Many classes of social and economic phenomena may be briefly and accurately described by quantitative statements which furnish precise measures of existing social conditions and significant changes from period to period. But more than description is sought. Careful comparisons of the facts of experience usually reveal relations of cause and effect, which make additions to our knowledge by explaining phenomena, not merely describing them.

The manner in which social and economic conditions affect individual and community welfare is measured in terms of family expenditures in relation to income, by industrial accident and mortality rates, by the decline of death-rates through the control of preventable disease, by the records of crime and dependency, by the length of the efficient working life, and by the changing standards of living of the population. These facts are as necessary for our enlightenment as those concerning the amount of exports and imports, the production of steel, the changes in commodity prices, and the volume of bank clearings.

The gathering and analysis of facts suggests the answers to many interesting and important questions. Is the Nation's production of food and raw materials keeping pace with the growth of population? Is the laborer better or worse off to-day than a decade ago, as measured by what he can afford to buy? Which occupations are dangerous to life and health, and why? What effect has social legislation on individual and social welfare? Are the wastes of community life growing less and the gains of coöperative activities growing greater? Is the immense increase in wealth really contributing to welfare as it should? Whether in estimating resources, or in measuring wastes, or in laying the foundation for intelligent public opinion concerning what is necessary for social and economic progress, quantitative measurements play a most essential part,

Social legislation based on facts. More and more frequently exact knowledge concerning conditions precedes the enactment of laws designed to modify them. The Government has been extending its activities constantly into fields where investigation is a necessary preliminary to wise action. Collection of facts concerning hazards in modern employments and a growing realization of the economic and social consequences of neglect, have been followed by agitation for better protection of workmen, by preventive methods or by social insurance. Before either policy could be adopted careful investigation was needed to determine the causes of industrial accidents and to place the responsibility upon specific processes or trades; and to determine what burden insurance must meet. During the last decade almost all our commonwealths have enacted compensation laws.

Observation has shown the comparatively low pay of women and minors in certain industries and the resulting low standard of living with its social consequences. It has been proposed to protect this class of wage-earners in a special manner by minimum-wage laws. But before the legislature and the administrative authority can act intelligently, it is necessary not only to know what wages are being paid in specific industries but also what wage is required in order to meet the reasonable needs of the worker. These are questions of fact and must be determined by inquiry.

Vital statistics in the service of sanitary science and health administration. Modern sanitary science owes its existence to the registration of deaths and their causes. Records of death and sickness constitute the bookkeeping of the public health movement. They direct the activities of the sanitary expert as chart and compass guide the navigator. Vital statistics furnish a definite measure of the value of sanitary improvement and of the progress of preventive medicine, as the following facts indicate:

TYPHOID FEVER IN PITTSBURGH, 1907–1911 [a]

YEAR	NUMBER OF CASES	NUMBER OF DEATHS	DEATH-RATE PER 100,000 POPULATION
1907	4,514	502	130.8
1908	1,833	255	46.6
1909	955	130	24.6
1910	998	149	27.7
1911	768	140	25.9

a From the Annual Report of the Department of Health, Pittsburgh, Pennsylvania, February 1, 1911, to January 31, 1912, p. 56.