STATISTICS IN PSYCHOLOGY AND EDUCATION

HENRY E. GARRETT

STATISTICS IN PSYCHOLOGY AND EDUCATION

BY

HENRY E. GARRETT

ASSISTANT PROFESSOR OF PSYCHOLOGY, COLUMBIA UNIVERSITY

WITH AN INTRODUCTION BY

R. S. WOODWORTH

PROFESSOR OF PSYCHOLOGY, COLUMBIA UNIVERSITY

LONGMANS, GREEN AND CO.

NEW YORK · LONDON · TORONTO

First Edition, January, 1926
Reprinted, November, 1926
Reprinted, October, 1927

MADE IN THE UNITED STATES

INTRODUCTION

MODERN problems and needs are forcing statistical methods and statistical ideas more and more to the fore. There are so many things we wish to know which cannot be discovered by a single observation, or by a single measurement. We wish to envisage the behavior of a man who, like all men, is rather a variable quantity, and must be observed repeatedly and not once for all. We wish to study the social group, composed of individuals differing one from another. We should like to be able to compare one group with another, one race with another, as well as one individual with another individual, or the individual with the norm for his age, race or class. We wish to trace the curve which pictures the growth of a child, or of a population. We wish to disentangle the interwoven factors of heredity and environment which influence the development of the individual, and to measure the similarly interwoven effects of laws, social customs and economic conditions upon public health, safety and welfare generally. Even if our statistical appetite is far from keen, we all of us should like to know enough to understand, or to withstand, the statistics that are constantly being thrown at us in print or conversation—much of it pretty bad statistics. The only cure for bad statistics is apparently more and better statistics. All in all, it certainly appears that the rudiments of sound statistical sense are coming to be an essential of a liberal education.

Now there are different orders of statisticians. There is, first in order, the mathematician who invents the method for performing a certain type of statistical job. His interest, as a mathematician, is not in the educational, social or psychological problems just alluded to, but in the problem of devising instruments for handling such matters. He is the tool-maker of the statistical industry, and one good tool-maker can supply many skilled workers. The latter are quite another order of statisticians. Supply them with the mathematician's formulas, map out the procedure for them to follow, provide working charts, tables and calculating machines, and they will compute from your data the necessary averages, probable errors and correlation coefficients. Their interest, as computers, lies in the quick and accurate handling of the tools of the trade. But there is a statistician of yet another order, in between the other two. His primary interest is psychological, perhaps, or it may be educational. It is he who has selected the scientific or practical problem, who has organized his attack upon the problem in such fashion that the data obtained can be handled in some sound statistical way. He selects the statistical tools to be employed, and, when the computers have done their work, he scrutinizes the results for their bearing upon the scientific or practical problem with which he started. Such an one, in short, must have a discriminating knowledge of the kit of tools which the mathematician has handed him, as well as some skill in their actual use.

The reader of the present book will quickly discern that it is intended primarily for statisticians of the last-mentioned type. It lays out before him the tools of the trade; it explains very fully and carefully the manner of handling each tool; it affords practice in the use of each. While it has little to say of the tool-maker's art, it takes great pains to make clear the use and limitations of each tool. As any one can readily see who has tried to teach statistics to the class of students who most need to know the subject, this book is the product of a genuine teacher's experience, and is exceptionally well adapted to the student's use. To an unusual degree, it succeeds in meeting the student upon his own ground.

R. S. Woodworth

Columbia University

PREFACE

THE present day emphasis on measurement and the quantitative treatment of results has made a knowledge of statistical method not only extremely useful but almost necessary to the student of psychology, education, and the social sciences. To those who have been well trained in mathematics, the acquisition of statistical technique offers no particular difficulty. To many otherwise capable students, however, either because of inadequate preparation in mathematics, or because their preparation is not very recent, the application of statistical method to data obtained from test and experiment is more than ordinarily difficult.

It is for this last group of students, especially, that this book has been written. Its primary purpose is to present the subject in a simple and concise form understandable to those who have no previous knowledge of statistical method. With this end in view, theory has everywhere been subordinated to practical application, and numerous illustrations of the various statistical devices have been provided. References have been given, however, for the benefit of those interested in the mathematical theory underlying the methods introduced.

The reader will note that in nearly all cases formulas have simply been stated without proof. This has been done, because the writer believes that most students of mental and social measurement are—and probably should be—more concerned with what a formula means and does than in how it is derived. There is considerable justification for such an attitude. In every science certain facts obtained from other fields must be taken on faith. We do not, to take a simple example, restrict the use of the radio or the microscope to those who understand the physical principles involved, and there seems to be no real reason why a student of psychology should not make good use of a correlation formula when he cannot derive it mathematically.

A chapter has been given to the subject of reliability—a topic too often passed over lightly—and considerable space has been devoted to correlation. An entire chapter, also, has been given to partial and multiple correlation. This method, while comparatively recent, is being widely used in educational research, and is probably destined in the near future to be more often used in the psychological laboratory. In the last chapter, the application of correlation and other statistical methods is shown to tests and testing.

Many have contributed to the making of this book of whom only a few can be mentioned. To Professors R. S. Woodworth and Mark A. May who read the manuscript, the writer is indebted for many useful and constructive criticisms. He is also grateful to Dr. M. R. Neifeld, to Mr. V. W. Lemmon, and to Miss Elizabeth Farber for computations and helpful suggestions.

HENRY E. GARRETT

COLUMBIA UNIVERSITY

CONTENTS

CHAPTER I

THE FREQUENCY DISTRIBUTION

CHAPTER II

GRAPHIC METHODS AND THE NORMAL CURVE

CHAPTER III

THE RELIABILITY OF MEASURES

CHAPTER IV

CORRELATION

CHAPTER V

PARTIAL AND MULTIPLE CORRELATION

STATISTICS IN PSYCHOLOGY AND EDUCATION

CHAPTER I

THE FREQUENCY DISTRIBUTION

I. The Tabulation of Measures into a Frequency Distribution

1. Measures in General: Continuous and Discrete Series

In the measurement of mental and social traits or capacities most of the facts with which we deal fall into what are known as continuous series. A continuous series may be defined simply as a series which is theoretically capable of any degree of subdivision. *IQ*'s, for example, are generally thought of as increasing by increments of 1 on a scale which extends from the idiot to the genius; however, there is actually no real reason—at least theoretically—why with more refined methods of measurement we should not be able to get *IQ*'s of 100.8 or even 100.83. Nearly all capacities measured by mental and educational tests and scales, as well as such attributes as height, weight, cephalic index, etc., have been found to be continuous, so that within the range of the scale used, any measure—integral or fractional—may exist and have meaning. Whenever gaps occur in a truly continuous series, therefore, these are usually to be attributed to our failure to measure enough cases, or to the relative crudity of our measuring instruments, or to some other fact of the same sort, rather than to the fact that no measures exist within the gaps.

There are, however, measures which do not fall into continuous series. Thus a salary scale in a department store may run from $10 per week to $20 per week in units of 50 cents or $1.00; no one receives, let us say, $17.53 per week. Or, to take another example, the average family in a certain locality may work out mathematically to be 4.57 children, although there is obviously a real gap between four and five children. Series like these, which contain real gaps, are called discrete or discontinuous.

It is probably fortunate—at least from the standpoint of the beginner in statistics—that nearly all of the measures which we make in psychology are continuous or can be treated as continuous. This considerably simplifies the problem, inasmuch as we may concern ourselves (for the present at least) almost entirely with methods of handling continuous data, postponing the discussion of discrete series to a later page.

2. The Classification of Measures in Continuous Series

Data collected from test or experiment are often merely a series of numbers or mass of figures without meaning or significance until they have been rearranged or classified in some systematic way. The first task that confronts us, then, is the organization of our material, and this leads naturally to a grouping of the measures into classes or categories. The procedure in grouping falls under three main heads, which are given in order below:

(1) The determination of the *range:* the interval between the largest and the smallest measures. The range is easily found by subtracting the smallest from the largest measure.

(2) Deciding upon the number and size of the groups to be used in classification. The number and the size of these steps or class-intervals depend largely upon the range and the kind of measures with which we are dealing.

(3) The tabulation of the separate measures within their proper step- or class-intervals.

TABLE I

ARMY ALPHA SCORES MADE BY 54 COLUMBIA COLLEGE MEN

1. THE ORIGINAL SCORES (UNGROUPED)

185	174	127	183	168	*126	177	154	157	189	172
*201	158	160	179	184	155	137	177	164	198	176
188	197	151	188	188	169	195	165	185	188	164
195	176	185	185	179	146	182	153	158	160	191
176	138	185	155	178	151	144	191	170	157	

* Maximum score = 201. * Minimum score = 126.

2. THE SAME SCORES GROUPED INTO A FREQUENCY DISTRIBUTION BY THREE METHODS

(A) (1) Scores	(2) Tabulation	(3) F	(B) Scores	F	(C) Scores	F
200 up to 205	/	1	200–204.99	1	200–204	1
195 “ “ 200	////	4	195–199.99	4	195–199	4
190 “ “ 195	//	2	190–194.99	2	190–194	2
185 “ “ 190	///// /////	10	185–189.99	10	185–189	10
180 “ “ 185	///	3	180–184.99	3	180–184	3
175 “ “ 180	///// ///	8	175–179.99	8	175–179	8
170 “ “ 175	///	3	170–174.99	3	170–174	3
165 “ “ 170	///	3	165–169.99	3	165–169	3
160 “ “ 165	////	4	160–164.99	4	160–164	4
155 “ “ 160	///// /	6	155–159.99	6	155–159	6
150 “ “ 155	////	4	150–154.99	4	150–154	4
145 “ “ 150	/	1	145–149.99	1	145–149	1
140 “ “ 145	/	1	140–144.99	1	140–144	1
135 “ “ 140	//	2	135–139.99	2	135–139	2
130 “ “ 135		0	130–134.99	0	130–134	0
125 “ “ 130	//	2	125–129.99	2	125–129	2
		$N = 54$		$N = 54$		$N = 54$

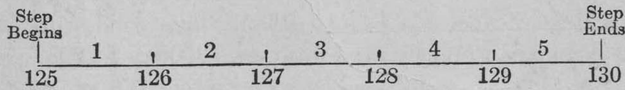
These three principles of classification are illustrated in Table I. The figures in this table represent the Army Alpha scores received by 54 college men. Since the highest score is 201, and the lowest 126, the range is found at once to be exactly 75 points. In deciding upon the number of "steps" or class-intervals to be used in grouping, the best general rule is to select by trial a step-interval which will yield not more than 20 nor less than 10 steps. The number of steps which a given interval will yield can be determined approximately (within one step) by dividing the range by the step tentatively chosen. In the present problem, for example, 75 (the range) divided by 5 (the step-interval) gives 15, which is one less than the actual number of steps, namely 16. A step-interval of 3 points will yield approximately 25 steps, while a step-interval of 10 points will yield approximately 7.5 steps. (Actually, for the given data, a step-interval of 3 points yields 26 steps, and one of 10 points 8 steps.)

The tabulation of the separate scores within their appropriate step- or class-intervals is shown in Table I(2A). In the first column of this table,—in the column marked "Scores,"—the step-intervals have been listed serially, with the smallest measures at the bottom of the column. The first interval, "125 up to 130," begins at 125 and ends at 130; the second interval "130 up to 135," begins at 130 and ends at 135 and so on. The last interval, "200 up to 205," begins at 200 and ends at 205. In column 2, marked "Tabulation," the separate scores have been listed opposite their proper intervals. The first score, 185 [see Table I(1)], is represented by a tally placed opposite step-interval "185 up to 190"; the second score, 201, by a tally placed opposite step-interval "200 up to 205"; the third score, 188, by a tally placed opposite "185 up to 190" and so on for the other scores. When all 54 scores have been listed, the total number of tallies on each step-interval (i.e., the frequency) is written in column 3, headed F (frequencies). The sum of the F column is called N. In the present case, of course, $N=54$. When the total frequency of each step-interval has been tabulated opposite its proper step-interval, as shown in column 3, our 54 Alpha scores are arranged into what is known as a Frequency Distribution.

The reader will note that the *lower limit* of the first step in the distribution (i.e., 125 up to 130) has been taken at 125 although the lowest actual score in the series is 126. This is due to the fact that when the step-interval equals 5 units, it facilitates tabulation as well as computations which come later on, if the lower limit of the first step-interval (and accordingly of each succeeding step-interval) is a multiple of 5. A step-interval of 126 up to 131 is just as good as a step-interval of 125 up to 130, theoretically; the second, however, is much easier to handle from the standpoint of the arithmetic involved.

3. Three Ways of Expressing the Limits of a Step-interval

Table I (2 A,B,C) illustrates three ways of writing the limits of a step-interval. In (A) the interval "125 up to 130" means that all scores from 125 up to but not including 130 fall on this step. In (B) the step-interval 125–129.99 means exactly the same thing. The upper limit is written 129.99 simply to emphasize the fact that this step-interval includes score 129 *plus* fractional parts up to 130, but does *not* include score 130. (C) expresses the same facts more clearly than (A) and not so exactly as (B). Thus 125–129 means that this step-interval begins *with* score 125 and ends *with* score 129. A diagram will indicate how (A), (B), and (C) are simply three ways of expressing the same facts.



Either method (B) or method (C) is advised as preferable to (A). It is fairly easy—even when one is on guard—to let a score of say 160 slip into the step-interval 155 up to 160 due simply to the presence of the 160 at the upper limit of the step. The accurate tabulation of a frequency distribution depends on getting each score into its proper step-interval, and for this reason one cannot be too careful in defining the limits of the steps.

In any frequency distribution we always assume that the scores within a given interval (i.e., the frequency) are spread evenly over the entire interval; and this assumption holds whether the length of the step is 3, 5 or 10 units. If we wish to represent *all* of the scores within a given interval by some single value, however, the midpoint of the interval is taken as