AN INTRODUCTION TO STATISTICAL ANALYSIS

BY

GEORGE GAILEY CHAMBERS, Ph.D.

PROFESSOR OF MATHEMATICS IN THE
UNIVERSITY OF PENNSYLVANIA

F. S. CROFTS & CO.
NEW YORK
1925

AN INTRODUCTION TO
STATISTICAL ANALYSIS

THE PLIMPTON PRESS · NORWOOD · MASSACHUSETTS
PRINTED IN THE UNITED STATES OF AMERICA

PREFACE

This book is a presentation of the elements of statistical analysis with special applications to biological, anthropometrical and mental measurements. The symbols and processes of mathematics are used when, and to the extent that, they seem incidentally helpful in developing the statistical concepts. When the point is reached at which such symbols and processes are needed, they are then presented and developed.

It is the result of an effort to present a few of the most elementary statistical concepts in a manner comprehensible to first-year students in normal schools, schools of education and colleges. In this effort, however, to make these concepts comprehensible to first-year students, the author has tried to be especially careful not to build up concepts which would have to be discarded by those students who later carry forward either their statistical or mathematical study to a more advanced stage.

In manuscript form, the greater part of this presentation of the subject has been used in classes of first-year students in the University of Pennsylvania for four consecutive semesters, by four different teachers. The author has had the benefit of the experiences of his colleagues in revising the manuscript for publication. For their helpful and sympathetic criticisms, he now, in this public manner, most cordially thanks his colleagues, Professor Maurice J. Babb, Professor John R. Kline, and Dr. Joseph M. Thomas.

Also the author thanks Dean John H. Minnick, of the School of Education of the University of Pennsylvania, for his helpful criticisms of considerable portions of the manuscript, and Mr. Elmer O. Delancy, a recent graduate of the University of Pennsylvania, for his valuable suggestions.

However, we hasten to add that none of those just named should be held responsible for the content as now being published, because not all of their suggestions have been followed.

The subject matter is somewhat more than can be covered by freshmen in one semester of three hours per week. Teachers desiring to cover the later chapters satisfactorily should omit either Chapter II or all of Chapter III after the first presentation of the logarithmic concept.

If, to even a slight degree, this effort makes this live subject more comprehensible to prospective members of the teaching profession, the author will feel amply repaid.

UNIVERSITY OF PENNSYLVANIA,
PHILADELPHIA, PENNA.

EDITOR'S NOTE

The notable progress made in recent years in the use of statistical methods by investigators has effected a change in scientific procedure.

The biologist, the psychologist, the educational expert and their students now present the results of their work in a language which even twenty years ago was understood by comparatively few.

Teachers and those interested in the results of such investigations have come to recognize the need for a simple presentation of the fundamentals of statistical theory, one which shall assume on the part of the reader a minimum of mathematical equipment and which shall give a knowledge of the terms which are used and the processes which are employed.

The aim of the author of this book has been to prepare a text in which the general principles and theorems are few in number and in which the mathematical analysis depends in general upon the fundamental definitions.

As one who has had experience in presenting the work of his own statistical investigations and in lecturing to many teachers on statistical methods, the author is in a position to give a presentation of the subject which will be of real service to those who desire a knowledge of elementary statistics.

George H. Hallett

A PRELIMINARY FORWARD LOOK

For the Student

STATISTICS, as presented in this book, is a set of tools for investigation. In the first chapter the student is introduced to the kind of material to be investigated. Subsequent chapters develop these tools, one after another, and give practice in using them, and in interpreting the results.

The student should attempt to assume the attitude of an investigator. To help him to acquire that attitude a complete set of "raw" data is introduced in Chapter V in the form in which it would come to a statistician. In each of the succeeding chapters, as new tools of investigation are presented, the student is asked to apply them to this "raw" data.

Thus, at the end of the course, the student will have made an analysis of a set of statistical data, and will be able to give a statistical description of it.

CONTENTS

Chapter VI. Averages

Chapter VII. Measures of Dispersion

Chapter VIII. Correlation

AN INTRODUCTION TO STATISTICAL ANALYSIS

CHAPTER I

MEASUREMENTS AND OTHER APPROXIMATIONS

In order that we may approach the subject of statistics with a good background, it seems advisable that our attention should first be given to a brief study of measurements and other approximations.

1. Direct Measurements. Examples. (*a*) A schoolboy measured the length of a certain metallic rod, using a ruler with a scale graduated to tenths of an inch, and obtained 9.7 inches as the measurement.

(*b*) A physician measured the temperature of his patient with a clinical thermometer and obtained 101 degrees, the smallest divisions on the thermometer scale being degrees.

(*c*) A butcher measured the weight of a piece of meat and obtained 8 pounds, 6 ounces, as the measurement of the weight, the weight arm of his balance being graduated to ounces.

Whenever quantities are measured directly as in each of the examples above, the numbers obtained are called *direct measurements*.

2. Measurements with Varying Degrees of Accuracy. In the first example of the preceding section, a certain metallic rod was measured by a schoolboy with 9.7 inches as the result. The same rod was later measured by a mechanic in the shop of a machine factory. He found that 9.7 inches was not the correct length of the rod. He recorded 9.68 inches as his measurement.

Afterwards the same rod was measured by a physicist in a laboratory equipped with instruments of great precision, and he found that 9.68 inches was not the correct length of the rod, even when the temperature was the same as when the machinist measured it. The physicist recorded 9.683 inches as his measurement.

If a way should be found to measure that rod with still greater precision, might it appear that even the very accurate measurement of the physicist was not really the true length of the rod?

3. The Approximate Character of Measurements. The discussion above concerning the measurements of that metallic rod shows that none of the measurers could truthfully say that the measurement which he obtained was exactly the true value of the length of the rod. Each of the numbers obtained, namely, 9.7, 9.68, and 9.683, was merely *an approximation* to the *true value* of the length of the rod.

We will now *assume* that there is a number which was the true value of the length of the rod at the time the measurement was made, but that no measurer could have asserted that his measurement was that true value.

For convenience, we will represent the true length of the rod by a letter even though we can never know the exact value for which that letter stands. For that purpose we will use the Greek letter α (alpha). We can then say that α is approximately equal to 9.7, or to 9.68, or to 9.683.

The symbol $\doteq$ will be used to stand for " is approximately equal to."

Hence we can write

$$\alpha \doteq 9.7, \text{ or}$$
$$\alpha \doteq 9.68, \text{ or}$$
$$\alpha \doteq 9.683$$

Similarly the measurement of 101 degrees in the second example of Section 1 is merely an approximation to the true value of the temperature of the physician's patient. If we let the Greek letter β (beta) represent the true value of the temperature, we can then write

$$\beta \doteq 101$$

All measurements are merely approximations to the true values of the quantities measured, even though the measurements have been made with the greatest precision that human ingenuity can devise. Of course, it may happen that a measurement is exactly the true value of the quantity measured, but the measurer can never know when that is so.

In general, if a is a measurement of some quantity whose true value is α, we can write

$$\alpha \doteq a$$

This statement should be read "α is approximately equal to a."

Greek letters will be used to represent true values and letters near the beginning of the English alphabet to represent approximations.

4. Possible Errors of Measurements. When the schoolboy recorded 9.7 inches as the measurement of the length of that metallic rod, he meant merely that its true length was not less than 9.65 inches and not more than 9.75 inches. That is, the error was not greater than 0.05 inches. Similarly, the mechanic's measurement of the rod meant that the true length was not less than 9.675 inches and not more than 9.685 inches, or that the error was not greater than 0.005 inches.

Frequently a measurement is accompanied by a number, expressed or implied, giving a value which the unknown error almost certainly does not exceed. That number is called a *possible error* of the measurement. What are the implied possible errors of the other measurements mentioned in Sections 1 and 2?

When it is desired to express a possible error along with a measurement, the possible error is written after the measurement with the sign $\pm$ between the measurement and the possible error; for example, the boy's measurement of the rod would then be written, 9.7 ± 0.05, and the mechanic's measurement of the rod would be written 9.68 ± 0.005. The expression ± 0.05, for example, should be read " with the possible error of 0.05."

We could then write

$$\alpha \doteq 9.7 \pm 0.05, \text{ or}$$
$$\alpha \doteq 9.68 \pm 0.005, \text{ or}$$
$$\alpha \doteq 9.683 \pm 0.0005$$

In general, if α stands for the true value of some quantity and a stands for a measurement of that quantity, then Δa, (read delta a), will be used to stand for an accompanying possible error. We could then write

$$\alpha \doteq a \pm \Delta a,$$

which should be read "alpha is approximately equal to a with the possible error of delta a."

The *possible* error should not be confused with the *probable error*, a term which appears under quite different circumstances. See Chapter X, Section 118.

5. The Natural Number Scale. The concepts of the preceding sections may be made clearer by means of *the natural number scale* on a straight line.

To establish a natural number scale on a line we must first think of a straight line extending in each direction without end, and then draw a segment of that line, say AB in Fig. 1. Then

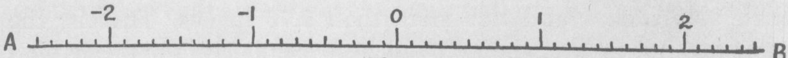

Fig. 1

select any point on that line, say O, as *the zero* of the scale. Next select any length as *the scale unit*, and lay off that unit from O a number of times in both directions. The point reached by laying off the unit once from O in either direction, say to the right, is called 1 on the scale. Then the points reached by laying off the unit twice, three times, etc. to the right are called 2, 3, 4, etc., and the points reached by laying off the unit once, twice, three times, etc. in the other direction, are called -1, -2, -3, etc.

The student should think of the line in Fig. 1 as extending to both the right and the left *without end*, (infinitely), and he should think of the integers 1, 2, 3, etc. and -1, -2, -3, etc. as being points located along that line *without end*, (infinitely), in both directions.

Next we will subdivide each unit segment into ten equal parts, each of these parts into ten equal parts, etc.

6. Representing Numbers on the Natural Number Scale. We will now explain how the numbers involved above in the measurements of that metallic rod may be represented on the natural number scale.

Those numbers are all between 9 and 10. We will therefore draw only the corresponding part of the scale, and will use a much larger unit, as in Fig. 2.

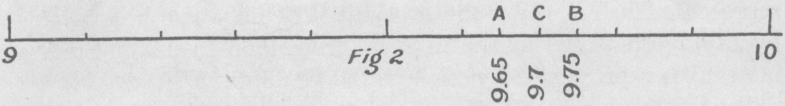

Fig 2

We can then locate the measurement of the rod by the schoolboy, namely 9.7, (C in the figure), and also the numbers 9.65, (A), and 9.75, (B), the limits beyond which the true value does not lie. We can then say that the true value, α, is somewhere in the segment AB, Fig. 2, and that the segment CB, or the segment AC, represents the possible error, namely 0.05.

To represent the measurement of the rod by the mechanic, we will make a drawing of the segment AB with an enlarged scale unit. (See Fig. 3.) In that figure D represents the measurement,

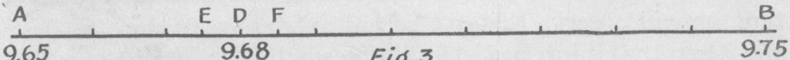

Fig. 3

and EF represents the interval within which α lies. Also the segment DF, or the segment ED, represents the possible error.

Exercises. **1.** Construct a drawing to represent the measurement of the metallic rod by the physicist.

2. On a wide piece of paper, or on a long blackboard, make one figure to represent all three of the measurements of this rod. In interpreting this figure, it is to be assumed that the true value of the length of the rod is unchanged as these successive measurements are being made. Notice that the successive approximations are closer and closer to the true value.

3. A small pebble was weighed three times. The first measurement of its weight was 2.6 grams; the second 2.62 grams; and the third 2.617 grams. Discuss these measurements as the measurements of the metallic rod have just been discussed, using drawings for that purpose.

It would be well for the student to practice making *freehand* drawings like those suggested above, trying to make them rapidly.

7. The True Value as a Limit. The three successive measurements of that rod and their relations to the true value α may be expressed in another manner. The first measurement,

$$\alpha \doteq 9.7 \pm 0.05,$$

represented in Fig. 2 by the segment AB, may also be expressed by saying that α is in the interval from 9.65 to 9.75. Such an interval is represented by the symbol [9.65, 9.75].

By the same method, the closer measurement

$$\alpha \doteq 9.68 \pm 0.005$$

would be expressed by saying that α is in the smaller interval [9.675, 9.685], represented in Fig. 3 by the segment EF.

Likewise the still closer measurement

$$\alpha \doteq 9.683 \pm 0.0005$$

would be expressed by saying that α is in the still smaller interval [9.6825, 9.6835].

When the student has worked Exercise 2 in the preceding section, he will observe that the second interval is within the first, and that the third is within the second. This means that the sequence of increasing numbers 9.65, 9.675, and 9.6825 is approaching α and also that the sequence of decreasing numbers 9.75, 9.685, and 9.6835 is approaching α from the other side.

Suppose now that, in some way, we were able to make a succession of closer and closer measurements of that rod without stopping.

If that could be done, then the sequence, 9.65, 9.675, 9.6825, could be extended without end, approaching closer and closer to α on one side and the sequence, 9.75, 9.685, 9.6835, would then be extended without end approaching α on the other side. Furthermore each of these sequences could be made so close to α that the difference would be less than any assigned number, however small. In such a case, α is called " the limit of either sequence."

8. Approximations Expressed as Inequalities. The approximation:

$$\alpha \doteq 9.7 \pm 0.05$$

may also be expressed by use of the sign $\leq$ (less than or equal to), as follows

$$9.65 \leq \alpha \leq 9.75$$

We now have three distinct methods of expressing an approximation; for example: