# A FIRST COURSE IN EDUCATIONAL STATISTICS

BY

EDNA E. KRAMER, Ph.D.

*Instructor in Education*
*Graduate Division, Brooklyn College*

NEW YORK

JOHN WILEY & SONS, INC.

LONDON: CHAPMAN & HALL, LIMITED

Printed in the U. S. A. 11/38

THE HADDON CRAFTSMEN, INC.
CAMDEN, N. J.

# PREFACE

The material in the following text is the result of the author's experience with classes at the New Jersey State Teachers College at Montclair.

For the past few years, all sophomore students have been required to take a 52-hour course in educational statistics. Inasmuch as the students varied greatly with respect to mathematical background and ability, it was difficult to find a text to furnish so heterogeneous a group with the essentials of the subject in the limited time allotted to the course. The standard works in educational statistics proved too comprehensive, whereas "primers" and "drill-books" were inadequate.

Hence the author resorted, at first, to lectures and mimeographed notes, and finally embodied this material in a text lithoprinted by Edwards Brothers. This lithoprinted text has been used by several sets of students, and has been revised and enlarged to meet the difficulties encountered by them. The present text contains all these revisions and additions.

The material is presented with the hope that it may be of assistance to instructors and students in other colleges and normal schools who may be experiencing difficulties similar to ours. The author has attempted to attain the following objectives:

(1) To present in elementary fashion those statistical facts which form a necessary background for a proper understanding of educational literature and a minimum prerequisite for educational research.

Material has been selected with a view to avoiding details which might prove confusing in a first course in the subject. Emphasis has been placed on interpretation rather than computation, and numerical work has been stressed only where it is essential to a proper understanding of statistical formulas and processes.

(2) To reduce the mathematics involved to simplest form.

For the benefit of the student with little mathematical background, any arithmetic or algebraic facts needed, are developed at appropriate points. For example, there is work with square roots, interpolation, graphic methods.

On the other hand, the author has attempted to furnish logical, if elementary, justification of all facts and formulas. The Appendix gives a rigorous development of a number of formulas, for the benefit of students whose algebraic background is good.

(3) To stress the practical rather than the theoretic aspects of the subject.

The author has gone to some pains to select illustrative material from current educational periodicals rather than to resort to artificial or hypothetical data. She has attempted in this way to furnish students simultaneously with drill in numerical computation and practice in critical examination of real data.

The exercises in the text require students to find additional illustrative data in current periodicals, and, in this way, verify the application of the statistical processes developed in the text.

The author takes this occasion to thank Professor David R. Davis for reading the original manuscript and offering many valuable criticisms and suggestions. She wishes also to express her appreciation to President Harry A. Sprague and Professor John C. Stone, who sponsored and encouraged the use of experimental material with Montclair students. She is also indebted to the numerous authors who have graciously given their permission to use data from articles which they have written.

STATE TEACHERS COLLEGE — EDNA E. KRAMER
UPPER MONTCLAIR, N. J.
June 1, 1934

# CONTENTS

# A FIRST COURSE IN EDUCATIONAL STATISTICS

## INTRODUCTION

## THE TEACHER'S NEED FOR STATISTICAL KNOWLEDGE

**Comprehension of Educational Literature.** The modern teacher who is interested in new movements in the general field of teaching, or in new methods of instruction in some particular subject, will naturally have occasion to read the most recent educational texts, as well as articles in current issues of educational periodicals. For a proper comprehension of such literature, it is becoming increasingly necessary to possess at least an elementary knowledge of statistical terms and statistical methods. Such reading is very likely to include statements like the following:

> The two mean scores are shown to be 5.6 before seeing the film. This is a shift in the direction of friendliness toward the Germans. The difference between these two mean scores is 0.38. $\sigma_1 = 1.43$ and $\sigma_2 = 1.11$. The probable error of the difference is 0.0708 so that the ratio of the difference to the probable error is 5.37. We seem to be justified therefore in concluding that the film "Four Sons" had a significant effect in making children more friendly toward the Germans.[1]

> By considering the range for behavior-score medians from 47.5 to 88.3 one has a range that is 2.47 times the probable error of the distribution. The range of academic grade medians ranging from 71 to 89.3 and corresponding with a range of 40 to 115 in behavior scores is 2.54 times the probable error of the grade distribution.[2]

A teacher should be able to understand and interpret a table similar to Table 1. Words like *mean*, *median*, *mode*, *norm*, *range*, etc., should be part of the teacher's vocabulary. Terms like

[1] Ruth C. Peterson and L. L. Thurstone, "The Effect of a Motion Picture Film on Children's Attitudes toward Germans," *Journal of Educational Psychology*, April, 1932.

[2] Herbert Sorenson, "Some Factors for Pupil Control Measured and Related," *Journal of Educational Psychology*, January, 1932.

*standard deviation* and *correlation coefficient* should not be mere abstractions to him.

TABLE 1*

CORRELATION COEFFICIENTS

| Composite scores | Correlation with algebra marks | No. of cases | Probable error |
|---|---|---|---|
| 1. I.Q. and arithmetic................ | 0.66 | 112 | 0.038 |
| 2. I.Q. and Orleans test scores........ | 0.68 | 112 | 0.034 |
| 3. Or. Sc. and Ar.................... | 0.72 | 112 | 0.031 |
| 4. I.Q., Or. Sc., Ar................... | 0.68 | 112 | 0.034 |

* Data from "Ability Grouping in the High School" by Ferdinand Kertes, *Mathematics Teacher*, January, 1932.

**Educational Research.** In making an experimental study or a school survey, a teacher should be capable of using statistical methods in reporting and interpreting his findings.

**Graphic Methods.** Even a teacher who may not need an extensive knowledge of statistical theory in connection with his reading may have occasion to give graphic representation to school data or to material which he wishes to present to his pupils in vivid fashion.

**Testing and Rating Pupils.** Every teacher will have frequent need to administer tests to pupils, whether these tests be of his own composition or standardized, and, what is more important, to interpret the results of these tests. For a proper handling of such results, the statistical method is essential.

**Varied Needs.** Other uses for the statistical method in the field of teaching might be here discussed, but instead, we shall proceed, in the first chapter, to a specific example.

# CHAPTER I

## ELEMENTARY STATISTICAL PROCEDURES

**Treatment of a Set of Scores.** The grades listed in the second column of Table 2 were obtained in a geometry test given to a group of 23 high-school students. After a week of drill, a second test was given on the same subject-matter. The grades obtained in the second test are listed in the third column of the table.

TABLE 2

| Student | Test 1 | Test 2 |
|---|---|---|
| 1 | 83 | 90 |
| 2 | 77 | 89 |
| 3 | 40 | 45 |
| 4 | 46 | 45 |
| 5 | 49 | 63 |
| 6 | 60 | 76 |
| 7 | 60 | 65 |
| 8 | 81 | 91 |
| 9 | 98 | 91 |
| 10 | 63 | 85 |
| 11 | 44 | 65 |
| 12 | 88 | 99 |
| 13 | 87 | 88 |
| 14 | 63 | 65 |
| 15 | 60 | 50 |
| 16 | 87 | 85 |
| 17 | 60 | 80 |
| 18 | 98 | 98 |
| 19 | 39 | 50 |
| 20 | 96 | 94 |
| 21 | 42 | 48 |
| 22 | 71 | 70 |
| 23 | 51 | 60 |

If you were the teacher of this group, what questions would you ask yourself in attempting to interpret the results of these tests? You might ask some of the questions listed below. See if you can answer them.

**Ex. 1.** What is the "average" rating in Test 1? in Test 2?[1]

**Ex. 2.** What rating occurs most frequently in Test 1? in Test 2? Such a rating is called the *mode* (it represents the most fashionable rating).

**Ex. 3.** Is the mode of the first set of marks close to the average of the set? Answer the same question for the second set.

**Ex. 4.** Using the average ratings obtained in (1) as the basis of your judgment, compare the ratings of the group *as a whole* in the two tests. How would you account for the difference in results? Do the students who obtain high ratings in Test 1 obtain high ratings also in Test 2? Do the same students obtain low ratings in both these tests?

**Ex. 5.** Arrange the scores in each test in descending order of their magnitude. What is the rating of the student who stands at the head of the group in Test 1? at the foot of the group? at the middle point? (The rating of this student is called the *median.*) What is the *range* of ratings in Test 1? Compare the median rating with the average and the mode.

**Ex. 6.** Answer the same questions as in Ex. 5 for Test 2.

**Ex. 7.** The results of students 1, 2, 3, 4, 5 are represented in Fig. 1 by points $A$, $B$, $C$, $D$, $E$ respectively. Copy this diagram and put in the points representing the other students. Note that the points are not scattered all over the graph, but seem to form a diagonal band.

**Averages.** The definitions of the terms used in the above questions are:

The *average* or *arithmetic mean* of a series of scores is the sum of these scores divided by the number of scores.

The *mode* of a series of scores is the score which occurs most frequently.

The *median* of a series of scores arranged in order of their numerical value is the middle score.

The *range* of a series of scores is the difference between the lowest and the highest scores.

In what follows, the term *mean,* or *arithmetic mean,* will be used instead of "average" since the latter is used to denote any

[1] Find these averages to the nearest tenth, i.e., carry to two decimal places and round off to one. For example:

63.43 becomes 63.4
63.47 becomes 63.5
63.45 becomes 63.5

one of the three *measures of central tendency—mean, mode,* or *median*—since each one of these indicates *the nature of a series of scores as a whole,* or *on the average.*

The *median* is a quickly determined measure, if the scores are arranged. The *mean* is most often used, since it is statistically most reliable. The *mode,* although not generally as important a measure as the other two, is occasionally more indicative as a

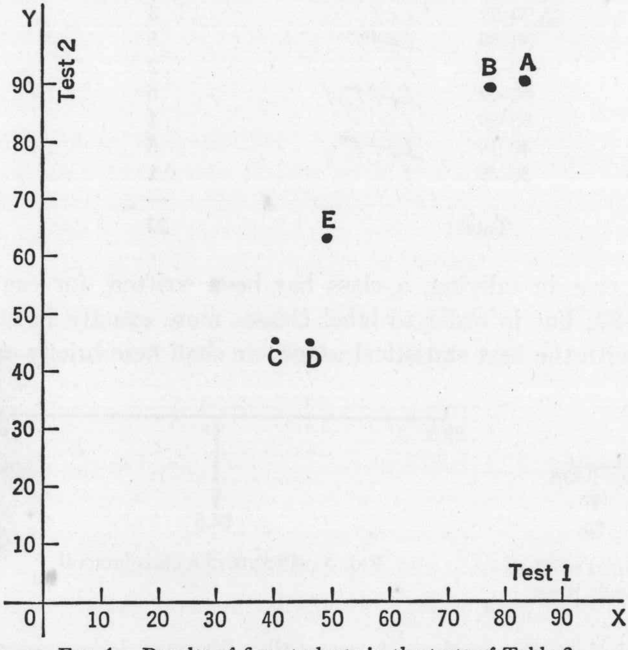


FIG. 1.—Results of five students in the tests of Table 2

measure of facts "on the whole." For example, in collecting for a school gift, in a class of 35 students, 30 students gave 25 cents each, 2 gave 10 cents, and 3 gave nothing. The mode, 25 cents, would be a better indication of the response of the students, as a group, than the mean, 22 cents.

**The Frequency Distribution.** In Table 3, the results of Test 1 (Table 2) have been grouped into a *frequency distribution,* by which is meant a *series of classes, and a set of corresponding frequencies.* The scores have been tallied in the fashion indicated.

The results of the test have been condensed by grouping the marks into classes. A *class-interval* of 10 was used in this classification, that is, the difference between the lowest score in each class and the lowest score in the next lower or next higher class is 10.

TABLE 3

| Class | | Frequency |
|---|---|---|
| 90–99 | /// | 3 |
| 80–89 | ~~////~~ | 5 |
| 70–79 | // | 2 |
| 60–69 | ~~////~~ / | 6 |
| 50–59 | / | 1 |
| 40–49 | ~~////~~ | 5 |
| 30–39 | / | 1 |
| Total | | 23 |

For ease in tallying, a class has been written, for example, as 80–89; but in order to label classes more exactly in accordance with the best statistical usage, we shall here briefly explain

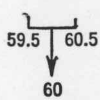


Fig. 2.—Picture of a single score

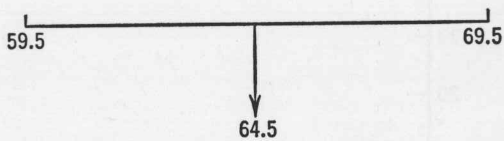


Fig. 3.—Picture of a class-interval

a statistical convention. A score like 60, say, is considered to have as its geometric picture a segment or interval of unit length with 60 at the mid-point. See Fig. 2. The boundaries of this interval are thus 59.5 and 60.5. This is in accordance with mathematical usage, since a measurement of 60 is a measurement made "to the nearest unit," which means that it is not exactly 60, but nearer to 60 than to 59 or 61, that is, something between 59.5 and 60.5. Hence, the class designated as 60–69 above, has as its picture Fig. 3, since the lowest score in the class is 60, whose lower boundary is 59.5, and the highest score in this interval is 69, whose upper boundary is 69.5. Hence the boundaries of this class are 59.5 and 69.5, and a better designation would be 59.5–69.5.

Throughout this book we shall designate classes in this way, for this is the customary method when scores are *integral*, and our data will involve *integral* scores. If scores are not integral, there are other ways of writing the class intervals. For example, if measurements are very fine, the limits of class intervals may be written 50–59.99, 60–69.99, 70–79.99, etc., or 50–60–, 60–70–, 70–80–, signifying that all scores equal to the lower limit, and up to but not including the upper limit, are included in a given class.

Since the value of many statistical measures depends on the boundaries and mid-points of class-intervals, numerical results will naturally vary with the method of selection of boundaries.

Table 4 shows Table 3 written with classes listed by their boundaries.

TABLE 4

| Class | Frequency |
|---|---|
| 89.5–99.5 | 3 |
| 79.5–89.5 | 5 |
| 69.5–79.5 | 2 |
| 59.5–69.5 | 6 |
| 49.5–59.5 | 1 |
| 39.5–49.5 | 5 |
| 29.5–39.5 | 1 |
| Total | 23 |

**Ex. 8.** Make a table similar to Table 4 for the results of Test 2, using a class-interval of 10.

**Histogram.** Fig. 4 is a bar graph or *histogram* of the frequency distribution of Table 4. The base of each rectangle in this figure is the length of the class-interval, and the altitude is the frequency of the particular class.

**Ex. 9.** Make a histogram to illustrate the frequency table which you prepared in answer to Ex. 8.

**Frequency Polygon.** In Table 5 the results of Table 4 have been rewritten so as to indicate the mid-point of each class.

Since the size of the class-interval is 10, the mid-point of each interval was found by adding half of 10, or 5, to the lower

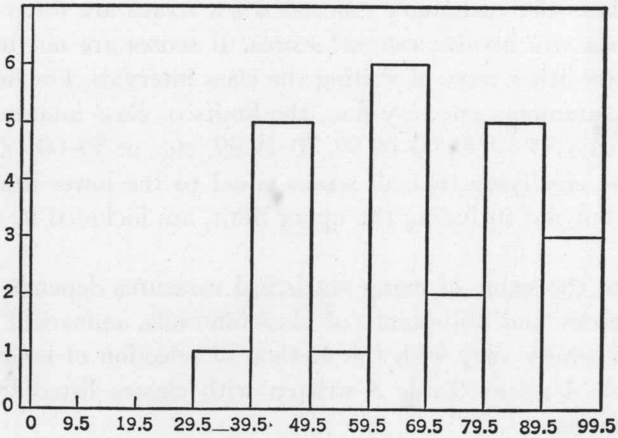


FIG. 4.—Histogram of Table 4

boundary of the interval. This is the way a frequency table is often written when it is to be used for approximate calculation or for graphic work. In such work, all scores in a class are regarded as concentrated at the mid-point of the class interval. A broken-line graph or *frequency polygon* of Table 5 is found in Fig. 5.

TABLE 5

| Mid-point of class | Frequency |
|---|---|
| 94.5 | 3 |
| 84.5 | 5 |
| 74.5 | 2 |
| 64.5 | 6 |
| 54.5 | 1 |
| 44.5 | 5 |
| 34.5 | 1 |
| Total | 23 |

**Ex. 10.** Rewrite, indicating the mid-point of each class-interval, the frequency table which you prepared in answer to Ex. 8, and then draw a frequency polygon to represent this table.

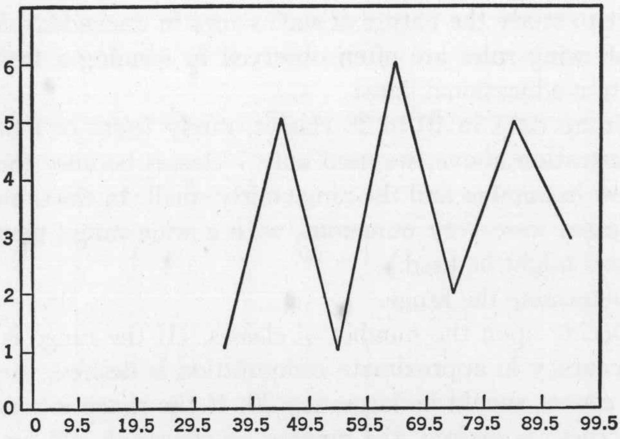


FIG. 5.—Frequency polygon of Table 5

Fig. 6 shows the histogram of Fig. 4 and the frequency polygon of Fig. 5 plotted together.

**Rules for Grouping Data.** Tables 3 and 4 indicate how a frequency distribution may be used to *organize a set of scores into compact form.* Of course, the use of such a table is hardly necessary for as small a set of scores as 23, but it is of great importance when the number of scores is large, since it would be

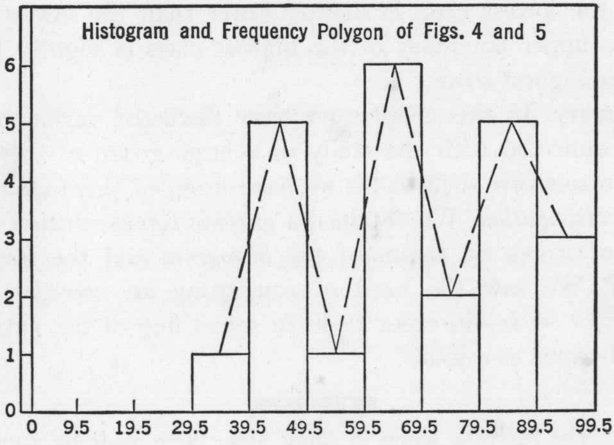


FIG. 6

difficult to study the nature of such scores in uncondensed form. The following rules are often observed in forming a frequency table from educational data:

1. Group data in 10 to 20 classes, rarely fewer or more. (In the illustration above, we used only 7 classes because the cases were few in number and the range fairly small. In the same way, if the cases were very numerous, with a wide range, more than 20 classes might be used.)

2. Determine the range.

3. Decide upon the number of classes. (If the range is large, or if accuracy in approximate computation is desired, the number of classes should be large, say 20. If the range is small and compactness is desired, the number of classes should be small, say 10.)

4. Divide the range by the number of classes to determine the class-interval. This will give the approximate size of the class-interval.

5. For computational purposes, class-intervals of 2, 5, 10, 15, 20 are the most convenient. Choose the number in this list nearest to the approximate value determined in item 4. If the approximate value is $2\frac{1}{2}$, choose 2. If the approximate value is $6\frac{1}{4}$, choose 5, etc.

6. Choose boundaries for your classes so that the lower boundary of the lowest class is slightly lower than the lowest score and the upper boundary of the highest class is slightly higher than the highest score.

**Summary.** In this chapter we have discussed various procedures connected with the study of a large group of scores. In order to *condense* such scores we have grouped them into a *frequency distribution*. We obtained a *graphic* representation of the group of scores by means of the *histogram* and the *frequency polygon*. We saw the need of computing an *average—mean, median*, or *mode*—in order to learn something of the nature of a set of scores *as a whole*.

## EXERCISES

**Ex. 11.** The following scores in Army Alpha were made by a group of sophomores specializing in science at the New Jersey State Teachers College