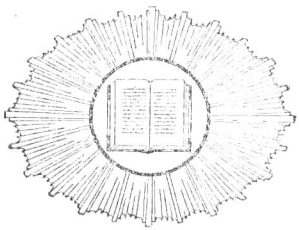
MEASUREMENT AND ADJUSTMENT SERIES

EDITED BY LEWIS M. TERMAN

STATISTICAL METHOD IN EDUCATIONAL MEASUREMENT

BY ARTHUR S. OTIS, PH.D.

Formerly Development Specialist with
Advisory Board of General Staff, U. S. War Department
Author of Otis Group Intelligence Scale
and other tests



Yonkers-on-Hudson, New York
WORLD BOOK COMPANY
2126 Prairie Avenue, Chicago
1926

WORLD BOOK COMPANY

THE HOUSE OF APPLIED KNOWLEDGE

Established 1905 by Caspar W. Hodgson

YONKERS-ON-HUDSON, NEW YORK
2126 PRAIRIE AVENUE, CHICAGO

Scientific method in education involves the careful measurement of each child's ability to learn and of the amount that he has learned. It also involves adjustment of organization, subject matter, and methods of instruction to the varying needs and abilities of pupils. This book is one of a series that sets forth the value, technique, and applications of educational measurement and adjustment. It describes in an elementary way those simple and useful methods and devices that are commonly employed in the interpretation of the data of mental and educational measurements

MAS : OSMEM-3

PRINTED IN U.S.A.

PREFACE

THIS book deals with those simple statistical methods that are needed in the interpretation of test results. It is designed for classroom teachers, administrators, students, and researchers.

The purpose of the book is to present the subject in such a way that it can be understood by those who know nothing whatever about statistical method. The reader is given an insight into the subject so that he will see the reasons for the various kinds of statistical procedure and will understand the meaning and significance of the results.

The attempt has been made to cover all the topics that a school teacher, administrator, or researcher is likely to have need for, but only those, the reader being referred to more technical books for explanation of special procedures that are seldom used. Particular attention is given to the explanation of the meaning and significance of correlation, and the simple and useful applications of partial and multiple correlation are explained.

The reader somewhat familiar with statistical method and terminology will find frequent footnotes which qualify, in the interest of exactness, the more general statements of the text. The lay reader need not trouble to study such notes, for they are not essential to a working knowledge of the method.

Several new charts for practical use are introduced, including a correlation chart and a percentile graph. These make easy certain important processes and calculations once known only to specialists in statistical method. The past few years have shown a very definite trend toward the development of such practical devices, and their more extensive use in future is certain; for this reason considerable space is devoted to the explanation of them.

Indeed, throughout the book the practical application of

methods has been kept in mind, and only such discussion of theory is given as is thought necessary to make the use of the methods and devices intelligent rather than rule-of-thumb. For example, the reader is shown the meaning of correlation by very simple illustrations with all refinements omitted, whereas the application of the complete formula is reduced by "job analysis" to a mere succession of simple arithmetic steps which may be performed with no thought of the formula. Attention may be centered on accuracy of the computation, which is reduced to a minimum by various tables.

It is believed that any person interested in scientific method in education may obtain from this book a good working knowledge and understanding of statistical method, no matter how new the subject may be to him.

ACKNOWLEDGMENTS

The writer is indebted to Lewis M. Terman for a very careful and helpful editing of the book, and to the following persons for reading the manuscript and offering valuable suggestions for its improvement: John R. Clark, Truman L. Kelley, William A. McCall, J. Ralph McGaughy, Rudolf Pintner, Godfrey H. Thomson, and Ben D. Wood.

John R. Clark also read the proof and has given especially helpful criticism.

Mrs. Otis has spent many patient hours with proof reading, computations, and copying of drawings.

Thanks are due the following persons who have contributed material for illustrations, exercises, etc.: E. J. Ashbaugh, George T. Avery, H. R. Bonner, Leo J. Brueckner, B. R. Buckingham, B. P. Heubner, A. M. Jordan, J. Cayce Morrison, H. H. Murphy, W. M. Proctor, Percival M. Symonds, Jacob Theobald, Elmer H. Webber, and Sara E. Weisman.

ARTHUR S. OTIS

EDITOR'S INTRODUCTION

In recent years extraordinary changes have taken place in the professional training of teachers, as a result of the development of scientific methods for the measurement of intelligence and educational achievement. Textbooks devoted to arm-chair discussions of educational principles, child study, and general methods are rapidly giving way to texts which deal with factual data resulting from the measurement of native abilities and of the effects produced upon such abilities by particular educational influences. This does not mean that the larger issues of educational theory and practice are being neglected, but only that we now recognize the futility of any attempt to dispose of such issues except in the light of investigational findings. At every point the consideration of educational theory must wait upon the discovery of facts regarding the nature of the raw material with which teachers work and regarding the changes which various kinds of teaching methods are capable of making in this raw material.

Present-day methods of educational investigation necessitate the constant use of statistical methods in the treatment of data. Familiarity with statistical procedures has for some time been considered necessary for the psychologist and the educational research worker, but we are rapidly coming to recognize that it is also an indispensable part of every teacher's equipment. Teachers cannot be expected to make intelligent use of test methods as long as the significance of test results is hidden from them in a maze of meaningless figures. But while every one will admit the *desirability* of acquainting the average teacher with the mysteries of statistical procedures, some may be inclined to doubt its possibility. Certainly it is impossible by the use of any of the textbooks which have hitherto been available. There are several statistical treatises which are reasonably suitable

for use with advanced graduate students of psychology and education, but in the editor's opinion this is the first text that is at all satisfactory for use in teachers' colleges or in teachers' reading circles. It is at the same time admirably adapted for use as an introductory text in colleges and universities.

It is generally recognized that the course in statistical methods presents, from the pedagogical point of view, exceptional difficulties. These are traceable in part to the inadequate mathematical preparation of the average student of education, but in part, also, to faults of exposition on the part of the textbook and the teacher. Principles which could be made concrete are presented abstractly; sequence is disregarded; insufficient drill is provided in fundamental everyday procedures. The average student who has taken such a course may have acquired a little "knowledge about" statistical methods, but he can make little or no use of them. The present textbook by Dr. Otis will do much to remedy this situation. Its characteristic features lie in the fact that it takes nothing for granted, that it relies upon the simplest and most straightforward explanations, that it gives concrete meaning to abstract terms and principles, that it teaches statistical procedure in connection with data of universal interest to teachers, and that it provides at every stage the practice necessary for making the procedures habitual. Students and teachers who have studied this text will be able to apply accepted statistical methods with all the ordinary kinds of measurement data, and will be vastly more able to read and understand the current literature of education and psychology.

Dr. Otis is widely known among American teachers as the author of several of the most useful tests of intelligence and of educational achievement. Among a smaller group he is known for his important contributions to statistical theory.

The book here presented rests, therefore, upon a solid foundation of mathematical knowledge and of practical experience in the use of measurement methods. From the pedagogical point of view it is without a rival in its field. It will add to the popularity of tests because it makes their results meaningful to the non-mathematical.

LEWIS M. TERMAN

CONTENTS

PAGE

EDITOR'S INTRODUCTION xi

CHAPTER

I. INTRODUCTION 1

Is there anything interesting about statistics? — The purpose of this book — Is there a science of mental measurement? — What kind of facts does statistical method reveal?

II. THE CENTRAL TENDENCY OF A DISTRIBUTION . . . 6

Comparison of groups — The mean — The median — Making a distribution — Range of score — Overlapping — Finding the median from a distribution — Special cases — Finding the mean score from a distribution — The mode — Summary

III. REPRESENTING THE WHOLE OF A DISTRIBUTION . . . 20

Comparison of individuals within a group — Rank order — Comparison of pupils of different groups — The percentile rank — An approximate method of finding a percentile rank — Interpreting the rule — The bar graph — The histogram — The step graph — Finding the median from a step graph — Quartile scores — The line of diagonals

IV. GROUPED DISTRIBUTIONS 36

Need for a new method of dealing with scores — Finding the mean of a grouped distribution — Histograms — The graph of a grouped distribution — Finding the median of a grouped distribution — Finding the median graphically — Finding quartile scores graphically — Calculating the median of a large group — A point of theoretical interest — Some theoretical considerations — Questions

V. THE PERCENTILE GRAPH 53

The percentile graph — The general utility of the percentile graph — How to draw a line of diagonals on a percentile graph — Cumulative errors — A table of per cents — Plotting points by means of the scale chart — Drawing the line of diagonals — The Universal Percentile Graph — Drawing two or more lines on one graph — The smoothing effect of grouping — The general shape of a graph — Finding a percentage graphically — Questions

VI. THE LAW OF NORMAL DISTRIBUTION 68

The law of probability in the flipping of coins — The chance element in scores — The normal surface of distribution — The normal curve of distribution — Skewed distributions

CHAPTER PAGE

VII. Percentile Curves 77

The nature of a sample — Effect of smoothing a line of diagonals — Percentile curves — The difficulty of smoothing a histogram — Advantages of the percentile curve

VIII. Variability of Scores 85

Need for measuring of variability — The total range — The interquartile range — Overlapping of classes — Other measures of variability — Median deviation — Probable error — Average deviation — Standard deviation — Variability of normal distributions — Questions

IX. Percentile Rank in a Normal Distribution . . . 95

The correspondence between scores and percentile ranks — Percentile rank in terms of variability — Testing the normality of a distribution — An accurate table of percentile ranks

X. The Correspondence between Tests 101

The problem — A method of finding the correspondence between scores in two tests — Making a table of correspondence — Use of percentile graph in finding correspondence — The line of relation — Alternative methods

XI. Averaging Measurements of a Pupil 118

How to average scores in two tests — Disadvantages of averaging percentile ranks — The method of transmuting scores — The need for a universal standard scale — *T* scores — Disadvantages of the *T*-score method — How to average scores and teacher's marks

XII. The Growth of Mental Ability 132

A growth chart — Comparison of growth curves — Growth in mental ability — Validity of mental-ability test scores — Units of measurement of mental ability — A curve of growth in mental ability — Interpreting the curve of growth — Individual differences — Bright, normal, and dull children — Mental age — Norms — The limit of true mental ages — Binet mental ages — Fictitious mental ages — Questions

XIII. The Measurement of Brightness 148

The intelligence quotient — The invalidity of the old IQ — The remedy — A popular misconception — The Coefficient of Brightness — Intelligence quotients from group tests — The Index of Brightness — A new method of finding an intelligence quotient — Percentile rank — The Interpretation Chart — Finding an intelligence quotient by means of the chart

CHAPTER PAGE

XIV. Norms 160

Need for representative sampling — Age norms — Grade norms — How to obtain grade norms from age norms — Grade status — Finding grade status from age norms — Finding age norms from grade norms — Interpolation — Subject ages — Fictitious subject ages — Educational quotient — The accomplishment ratio — Summary

XV. The Meaning of Correlation 175

The comparison of relative standing — The meaning of correlation — The natural method — Perfect correlation — Coefficient of correlation — Significance of the coefficient of correlation — A coefficient of correlation in relation to causation — Questions

XVI. The Calculation of a Coefficient of Correlation . 186

The product-moment method — The difference method — The difficulty with fractions — The use of an assumed mean — The Otis Correlation Chart — How to use the Correlation Chart — Negative correlation — Rank methods of computing correlation — The second rank method — The Spearman Footrule — Interpolation — Correlation by unlike signs — Other methods of measuring correlation — Questions

XVII. Correlation and Prognosis 217

The criterion — The coefficient of alienation — Further interpretation of a coefficient of correlation — Interpretation of a coefficient of correlation in terms of displacement — Errors of measurement — The effect of errors of measurement upon correlation — Reliability — Correlation for attenuation — Questions

XVIII. Partial Correlation 230

A prevalent misconception regarding correlation — The need for a more direct measure of correlation — How to find a coefficient of partial correlation — Nature of partial correlation — Effect of heterogeneity on correlation — How to correct a coefficient for a change in heterogeneity — Question

XIX. Multiple Correlation 238

Measuring prognostic value — The meaning of multiple correlation — The formula for multiple correlation — Finding the best weighting — Formula for weighting — Symbolization — Relation of a multiple correlation to the total correlation — Regression equations — Calculations with four or more variables

CHAPTER PAGE

XX. Reliability 247

The fallibility of test scores — Causes of variability of the scores of a test — Reliability — Measures of reliability — Probable error of a test — How to find the probable error of a test — The formula for finding the probable error of a score from the differences between pairs of scores of single individuals — Need for other measures of reliability — Probable error in terms of variability of scores — Reliability coefficient of correlation — Relation between reliability coefficient and probable error — Formula for finding the probable error of a score from the reliability coefficient — Interpretation of a reliability coefficient — Reliability and heterogeneity — Errors of measurement — Validity — The probable error of a coefficient of correlation — The formula for the probable error of a coefficient — How to find the probable error of a coefficient of correlation obtained by the Otis Correlation Chart — Significance of the probable error of a coefficient — The probable error of a difference — Probable errors of a mean and of a standard deviation — Probable error of coefficients of partial and multiple correlation — Reliability of averages — Practice effect — Question

XXI. Grading and Classifying 267

The problem — Regrading on the basis of mental ability — Need for an achievement test — Regrading by counting papers — Regrading by a percentile curve — A less drastic method of regrading — Classification within grades — Bright, normal, and dull sections — The need for varied courses — The three-track plan — The Interpretation Chart — Convenience of the Interpretation Chart — Classification for the three-track plan — Grading within bright, normal, and dull groups — No rigid classification — Taking account of teachers' marks in grading and classifying — How to take account of teachers' marks — Rating scale — A "five-point" scale

Appendix:

I. Abbreviations and Symbols 289
II. Statistical Tables 291
III. Supplementary Practice Material 306
IV. Answers to Exercises 317
V. Bibliography 321

Index 329
Age Calculator 333
IQ Slide Rule 337

STATISTICAL METHOD IN EDUCATIONAL MEASUREMENT

CHAPTER ONE

Introduction

Is there anything interesting about statistics? Most people would probably say no. They mean that most statistics are for them uninteresting. This is quite natural. But when a mother keeps a record of the growth of her baby from week to week, or when a teacher who suspects that her pupils are not as bright as those of another teacher undertakes to find out by testing them with a mental-ability test whether this is true, these data become for that mother or teacher very interesting.

The purpose of this book. In reality, this book does not deal with statistics as such. It deals with the method of interpreting statistics — one's own interesting data, collected painstakingly for an important purpose. When the teacher has taken anywhere from one hour to ten hours in giving and scoring a standard test and tabulating the scores, she wishes to get from these data all the significance and enlightenment that can be found in them. This book is written for the special purpose of enabling a teacher who knows nothing whatever about statistical methods to work over her scores and any other data in the most convenient way, so that those facts that are inherent in the data and which are earnestly sought will stand out vividly and in their true perspective.

After all, those things are likely to be most interesting about which we know most, and the writer firmly believes that the more the teacher or principal learns about the methods that may be invoked to make a seemingly incoherent mass of scores or other measurements yield fruit in those truths that are so vital to the progress of education, the more

interest that teacher or principal will find in the study of the method itself.

Is there a science of mental measurement? When one speaks of science we generally think of some organized system of relatively exact knowledge, such as mathematics, physics, or chemistry. Mental measurement, on the other hand, is so inexact that we may be inclined to feel that it should not be called a science. However, the method of handling the data of mental measurement statistically in order to determine how inexact they are and what degree of reliance may be put upon them constitutes one of the most exact of sciences. We may have used a very inadequate measure of the ability of a pupil in written composition, for example, but by the collection of a large number of such measurements we are able by statistical methods to determine the most probable true measure of a given pupil's ability and to know just how much reliance may be placed upon the results.

What kind of facts does statistical method reveal? The statistical method described in this book is equally applicable to the measurement of "intelligence," of the products of education and training, of character traits — indeed, of any kind of anthropometric measurements. The emphasis in this book, however, is laid on the application of the method to so-called mental and educational measurements.

Let us suppose that a school principal has administered a mental-ability test and an achievement test to the pupils of Grades 4 to 9 of his school. What may the scores reveal?

(1) The principal may determine, within a known degree of accuracy, the relative amount of the mental development of each pupil tested and the relative amount of knowledge he has attained during his school career.

(2) He may determine, within a known degree of accuracy, the average level of mental development of the pupils of any

grade in comparison with that of the pupils of any other grade, and determine whether the steps in average development from one grade to another are approximately uniform. He may do the same for the steps in average amount of knowledge possessed by the pupils of the several grades.

(3) He may determine how widely different in amount, or, as we say, how *variable*, the abilities of the pupils of a single class are, and may compare the variability of one class with that of another. Stating this in another way, we may say that the principal may determine the relative *homogeneity* of his various classes in mental ability and in knowledge (achievement).

(4) He may determine the degree of *overlapping* of grades and classes in mental ability and achievement. For example, he may find that 25 per cent of the pupils of the fourth grade exceed the average of the fifth grade in achievement, and that 5 per cent of the fourth grade exceed the average of the sixth grade.

(5) If other schools in his city have used the same tests, he may determine how his school stands in comparison with those schools with respect to the average mental development of the pupils and their average achievement.

(6) By means of the *grade norms* given in the manuals of directions, he may determine how his school stands in comparison with schools in other cities with respect to the average mental development of the pupils and their average achievement. (The norm for the fifth grade, for example, is the average score of fifth-grade pupils all over the country.)

(7) He may compare the mental ability of any individual pupil with that of the other pupils of the same grade or class in a number of ways. He may say, for example, that George has a degree of mental ability that places him in the upper quarter of fifth-grade pupils, or he may say that George exceeds in mental ability 85 per cent of the pupils of the fifth grade. There is a simple way of determining this percentage

by means of a device called the Percentile Graph, which is described in Chapter V. A pupil's relative position in achievement may be similarly expressed.

(8) A pupil's ability may be expressed also in terms of the age for which such ability is normal or average. Thus, the principal may find that George has a mental ability just normal for the age of 12 years, 8 months. (If this is true, George is said to have a mental age of 12 years, 8 months.)

(9) A pupil's achievement may be expressed, if desired, in terms of grade status; that is, if a pupil has made a score in the achievement test which is just average for the fifth grade at the end of the third month of the school year, we may say the pupil has a grade status of 5.3.

(10) The principal may also compare the achievement of each pupil with his mental ability and thus determine which pupils are doing school work that is on a par with their mental ability. The ratio between achievement and mental ability may be expressed mathematically as the *Accomplishment Ratio*, as is explained on page 172.

(11) The principal may find the degree of correspondence between achievement and mental ability by the method of *correlation*. This may be done by means of a Correlation Chart, which is described herein (pages 192 to 201). The chart is designed for use by persons wholly unfamiliar with the method of correlation. By the use of this method the principal may find also the degree of correspondence between achievement as measured by the test and achievement as judged by the teacher, or between any two abilities, either measured or estimated.

(12) If one form of the test is given at the beginning of the year and an alternative form at the end, the principal may determine the relative progress made in achievement by each pupil and by the various classes. By means of measures of progress, it is possible to compare the merits of various

teaching methods and devices. In scientific procedure of this sort rests the future development of educational method.

In enumerating the various items of information that a school principal may discover from the results of the administration of mental and educational tests, we have not mentioned the *application* of these, chief of which, of course, is the classification of pupils into homogeneous groups in order that instruction may be better adapted to the needs and abilities of the pupils. Also, it is not to be inferred, of course, that such discoveries may be made only by the principal. Any teacher or supervisor may make the same investigation as far as her own pupils are concerned.

The determination of any one of the facts mentioned above regarding one's pupils, however, requires the understanding of statistical method, at least to some small degree. It is the aim of this book to set forth in an elementary way those principles and procedures which will enable any teacher, supervisor, principal, superintendent, or other researcher to obtain regarding his pupils any of the kinds of information mentioned above, and many others, without his having made any previous study of statistical method.

It may be advantageous to read each chapter through as a whole without attempting to master it as you go along. In Chapter III, for example, several methods of finding a percentile rank are discussed, first according to a common conception, next a precise but rather impractical method, and finally an approximate method that is simple and practical. You should become familiar with the third method, and for that purpose an exercise is provided giving practice in finding a dozen percentile ranks. It is not necessary, however, to spend time studying the first two methods unless you are especially interested in these. In general those discussions not followed by exercises are of a type leading to more practical methods for which exercises are given.

CHAPTER TWO

The Central Tendency of a Distribution

Comparison of groups. The pupils of a certain school in New York were recently tested with an arithmetic reasoning test. The scores of the pupils of Grades 5A and 5B are given in Table 1.

Is it possible to tell from inspection of this table whether there is a tendency for the scores of one class to run higher "on the average" than the scores of the other? Make a guess from inspection of the table as to which grade has made the better scores "on the average." Guess about how much better the scores of one grade are than those of the other.

TABLE 1

Scores of Pupils in Grades 5A and 5B in the Arithmetic Reasoning Test

Grade	No. of Pupils	Scores
5A	36	11 8 7 8 7 7 5 4 8 10 11 9 10 12 9 6 15 7 8 5 12 14 11 5 5 9 10 12 4 8 7 10 8 8 10 9
5B	47	7 12 8 10 11 6 11 11 11 10 8 5 11 8 8 6 13 7 9 13 11 9 12 9 6 8 8 8 6 12 6 14 10 16 12 13 10 11 12 10 14 10 10 9 8 13 11

The mean. A common measure of the achievement of a class as a whole is the "average" score of the class. This is found by adding all the scores of the class and dividing the sum by the number of pupils in the class. Thus, the sum of the scores of the pupils in Grade 5A is 308. Dividing this sum by 36, the number of pupils, we get $8\frac{20}{36}$, which we can call 9 if we wish only the nearest whole number. Carried to two decimal places, the "average" for the class is 8.56