STATISTICAL ANALYSIS IN EDUCATIONAL RESEARCH

Lindquist

STATISTICAL ANALYSIS IN EDUCATIONAL RESEARCH

E. F. LINDQUIST
Professor of Education
THE STATE UNIVERSITY OF IOWA

HOUGHTON MIFFLIN COMPANY
BOSTON · NEW YORK · CHICAGO · DALLAS
ATLANTA · SAN FRANCISCO
The Riverside Press Cambridge

The Riverside Press
CAMBRIDGE · MASSACHUSETTS
PRINTED IN THE U.S.A.

PREFACE

THE past twenty years or more have been a period of extremely rapid and significant development in statistical theory and practice. Yet, while many of the recent contributions — particularly those of R. A. Fisher and his students — appear to have almost revolutionary significance for educational research, research workers in this field have in general failed to recognize their amazing possibilities, or at any rate have not widely realized these possibilities in practice.

A part of this neglect has been due to the mistaken notion that it is seldom necessary to use "small" samples in educational research — that most of our samples consist of large numbers of pupils or of individual observations — and that, hence, "small sample" theory can be of relatively little practical interest or value to research students in education. In taking this attitude, we have overlooked the very significant fact that most of our samples, however large in terms of numbers of individual observations, are not simple random samples, but consist of relatively homogeneous and intact subgroups, such as the pupils in a single school or under a single teacher. The number of these subgroups, furthermore, is usually indeed small, and it is only through the use of small sample theory that we can accurately evaluate the results obtained.

Perhaps a more telling reason, however, for this continued neglect is that the only expositions of these techniques that have thus far been readily available, particularly Fisher's *Statistical Methods for Research Workers* and *The Design of Experiments*, have proved inordinately difficult for students in education to comprehend. This fact is due in part to the unfamiliar statistical notation and terminology employed; in part to the frequent and wide gaps in the sequence of logic which are left to the reader to fill but which can be readily supplied only by a reader with advanced mathematical training; and perhaps most of all to the fact that all illustrations given are in the field of agricultural experimentation and are concerned with "plots," "blocks," "yields," "treatments," etc., rather than with "schools," "classes," "scores," "methods," "pupils," etc.

The writer's primary purpose in this book, accordingly, has been to translate Fisher's expositions into a language and notation familiar to the student of education; to clarify the exposition further by presenting all steps in the logic, in a manner such that they may be followed by students with little mathematical training; and to point out specifically and illustrate concretely what seem to be the most promising applications of Fisher's methods in educational research. Particular emphasis has been placed upon the importance of more careful *design* in educational research. Many of the difficulties that have been met by educational research workers in the analysis of their results have arisen from their tendency to plan their experiments and investigations with little direct regard to the methods of analysis that are later to be employed, or even to postpone any very careful consideration of analytical procedures until the investigation itself has been actually concluded. One of the most valuable features of the methods of analysis of variance is that they recognize that the problem of design is inseparable from that of analysis, and that their use makes it difficult to ignore the maxim that no investigation should be actually initiated until the analytical procedures have been thought through to the last detail.

This book, however, has not been restricted to techniques which are directly attributable to R. A. Fisher. Its purpose, more generally stated, has been to make more readily available and comprehensible to students in education any of the more recent developments in statistical theory and practice which seem likely to prove of value in educational research but which thus far have received little or no attention in the standard introductory texts in educational statistics. In particular, an effort has been made to bring the student up to date on the logic of statistical inference and to make him more keenly conscious of the constant need for very critical consideration of the *assumptions* underlying any statistical technique he may employ.

The writer is convinced that the development of a genuine understanding of the nature of statistical inference and a thorough training in the use of the methods of analysis of variance should be considered an absolute essential in the general preparation of research students in education. It is hoped, therefore, that this book will prove usable as a textbook in advanced courses in educational statistics or in the second half of a required full year introductory course. If so used, it will require considerable supplementation by other references, since (with a few minor exceptions) no attempt has been made in this book to discuss any problems that have already been adequately treated in the standard texts. It has been the writer's experience, however, that adequate consideration of the problems here treated will require a considerable share of a typical three-hour course.

It may be noted that the omission of a set of exercises for the student has not been accidental. The methods here considered have been so little used in educational research as to make impossible, for the present, the collection of a set of exercises or examples based on actual data in the field of education. Rather than attempt to prepare a set of exclusively artificial examples, the writer has preferred to postpone the publication of exercise material until a later date. He does, however, intend to prepare and publish as soon as possible an exercise book for the student somewhat along the lines of the *Study Manual* accompanying his *A First Course in Statistics.* Such an exercise book will also provide a means of supplementing this book with discussions of later contributions, and of drawing attention to any changes in emphasis or content in the present volume that later experience may prove desirable but which may not justify a new edition of this book. The omission of lists of references or supplementary readings at the end of each chapter is also deliberate. Most of the general references which could be given would be of doubtful value to the students to whom this book is addressed, for the reasons already indicated in the case of Fisher's books. Furthermore, many of the original papers which have been consulted in the preparation of this volume have appeared in journals which are not readily accessible to students in education. (The more important of these, however, have been cited in footnote references.)

The writer has been extremely fortunate in obtaining an unusual amount of assistance in the preparation of this book. He is grateful, first of all, to the students in his own classes who used the book in its preliminary mimeographed edition and who directed attention to many typographical errors and to instances in which the lucidity could be improved. Various parts of the preliminary manuscript were read by Dr. C. H. McCloy of the State University of Iowa, Dr. Marian Wilder of the University of Minnesota, Dr. John C. Flanagan of the Co-operative Test Service, New York City, Dr. Edward E. Cureton of the Alabama Polytechnic Institute, and Dr. Jack Dunlap of the University of Rochester, all of whom offered many valuable suggestions. The entire manuscript was carefully read by Professor Allen T. Craig of the Department of Mathematics of the State University of Iowa. Special acknowledgments are due Professor G. W. Snedecor of Iowa State College, who gave very generously of his time in consultations with the writer in the earlier stages of the book and whose *Mathematical Statistics* contributed greatly to the writer's own understanding of the possibilities in Fisher's methods.

The major acknowledgment is due Dr. W. G. Cochran, formerly of the Rothamsted Experimental Station, Harpenden, England, and now of the Statistical Laboratory of Iowa State College. Dr. Cochran read the entire manuscript most painstakingly and offered a very large number of concrete and constructive suggestions, all of which the writer found it desirable to observe in the final revision of the manuscript.

Grateful acknowledgment is made to Professor R. A. Fisher, and to his publishers, Oliver and Boyd, for permission to reproduce the tables of t and of χ^2, and the normal probability table from *Statistical Methods for Research Workers*, as well as a part of

the table of random numbers from *Statistical Tables for Biological, Agricultural, and Medical Research* (Fisher and Yates). Similar acknowledgment is due Professor G. W. Snedecor and the Collegiate Press for the use of the table for *F* from Snedecor's *Statistical Methods.*

E. F. LINDQUIST

CONTENTS

I. Fundamental Concepts in Sampling Theory

1. Introductory 1
2. Definitions of Important Terms 2
3. Testing Statistical Hypotheses 10
4. Tests of Significance: The Null Hypothesis 15
5. Limitations of Standard Error Formulas Containing Population Parameters: Need for a Special Small Sample Theory 18
6. The Problem of Sampling in Educational Research 21
7. The Technique of Random Selection 24

II. The Use of the χ^2 Distribution in Testing Hypotheses

1. Introductory 30
2. The Meaning of χ^2 31
3. Degrees of Freedom 33
4. The Sampling Distribution of χ^2 34
5. Tests of Goodness of Fit 37
6. Tests of Independence 41
7. Tests of Homogeneity 43
8. Combining Probabilities from Independent Tests of Significance 46

III. Small Sample Error Theory

1. An Improved Estimate of the True Standard Deviation . . . 48
2. The Sampling Distribution of "t". 51
3. The Significance of the Mean of a Small Sample 54
4. The Significance of a Difference in the Means of Independent Small Samples 56
5. The Significance of a Difference in the Means of Related Measures 58
6. The Significance of a Difference in Variability for Small Samples: The "F" Test 60
7. The Significance of the Mean of a Sample Consisting of Randomly Selected Intact Groups 66
8. The Significance of a Difference in Means for Samples Each of Which Consists of Relatively Homogeneous Subgroups . . . 72

IV. The Importance of Design in Educational Experiments

1. The Need for Measures of Precision 76
2. Sources of Error in Educational Methods Experiments . . . 78
3. Types of Experimental Designs 80
4. Valid Methods of Analysis of Experimental Data 84

V. ANALYSIS OF VARIANCE
1. The Fundamental Theorem in Analysis of Variance 87
2. Analysis into Two Components: Analysis of Results in a Simple Methods Experiment 93
3. Rules for Analyzing the Results of a Simple Methods Experiment 100
4. Analysis into Three Components: Analysis of Pooled Results of Duplicated Experiments in Randomly Selected Schools (Groups of Uniform Size) 104
5. Analysis into Four Components: Analysis of Pooled Results of Duplicated Experiments in Randomly Selected Schools of Unequal Size 114
6. Rules for Analyzing Pooled Results of Duplicated Experiments in Randomly Selected Schools 119
7. Modifications of Design and Analysis in Methods Experiments Including Several Schools 127
8. The Assumption of Homogeneity of Variance 132
9. The Effect of Heterogeneous Variance Within Schools Upon the "F"-Test of Significance in Methods Experiments 139
10. General Possibilities of Analysis of Variance in Educational Research 145
11. Simple Factorial Designs in Methods Experiments 163
12. Factorial Design Duplicated in Randomly Selected Schools . . 173

VI. ANALYSIS OF COVARIANCE
1. Introductory 180
2. The Essential Nature of Analysis of Covariance 180
3. Derivation of Basic Formulas 183
4. Analysis of Covariance in a Simple Methods Experiment . . . 191
5. Analysis of Covariance in Duplicated Experiments in Randomly Selected Schools 196
6. Analysis of Covariance to "Hold Constant" Other Variables than Initial Status 204
7. Statistical Control of More than One Concomitant Variable . . 205

VII. MISCELLANEOUS PROBLEMS IN CORRELATION ANALYSIS
1. Introductory 208
2. The Significance of a Product-Moment Correlation Coefficient . 210
3. The Reliability of a Product-Moment Correlation Coefficient . 211
4. The Significance of a Difference Between *r*'s Obtained from Independent Random Samples 214
5. Combining Correlations from Several Samples 218
6. The Effect of School Differences upon Correlation Coefficients Obtained from Heterogeneous Samples Consisting of Intact School Groups 219
7. The Estimated Validity of a Test *n* Times as Long as a Given Test 228

8. Correlation Coefficients Corrected for Attenuation 233
9. Test for Linearity of Regression 235
10. Test for Significance of a Non-Linear Relationship 238
11. The Correlation Ratio 239
12. Biserial Correlation 241
13. Tetrachoric Correlation 243
14. The Rank Correlation Coefficient 246
15. Partial Correlation 248
16. Multiple Correlation and Regression 253

APPENDIX
Table 17. Absolute Deviations from the Mean of a Normal Distribution Exceeded by Given Proportions of the Cases . . . 261
Table 18. Table of Random Numbers 262

INDEX 265

LIST OF TABLES

1. Distributions of Individual and Mean Scores of Ninth-Grade Pupils on the 1935 Iowa Every-Pupil Test in English Correctness for 24 Schools 23
2. Table of χ^2 36
3. Table of t 53
4. Table for F 62
5. Distributions of Scores on the 1938 Iowa Every-Pupil Test in English Correctness for the Ninth-Grade Pupils in 11 Iowa High Schools . . 67
6. Criterion Scores in a Simple Methods Experiment 93
7. Analysis of Variance of Class Means in a Methods Experiment Involving 5 Schools and 3 Methods 107
8. Analysis of Variance of Pupil Scores in a Methods Experiment Involving 3 Methods and 5 Schools 115
9. Distribution of Variances of Scores of Seventh-Grade Pupils on a Comprehensive Test of Achievement in Arithmetic (Test D of 1938 Iowa Every-Pupil Tests of Basic Skills) in 94 Schools (20 Pupils from Each School), Compared with Distribution Expected on Hypothesis of Random Sampling 136
10. Distributions of Variances of Scores on Various Educational Achievement Tests, Each Compared with Distribution Expected on Hypothesis of Random Sampling 138
11. Distributions of Observed and Theoretical Variances of Scores of Sixth-Grade Pupils on a Test of Silent Reading Comprehension in 104 Iowa Schools (15 Pupils from Each School) 141
12. Scores on Initial (X) and Final (Y) Tests for Pupils Who Studied Under Methods A and B in an Experiment in Fifth-Grade Language Instruction 198
13. Values of Correlation Coefficient Required for Significance at the 5 per cent and 1 per cent Levels for Samples of Various Sizes (N) 212
14. Values of z for Various Values of r 215
15. Distributions of Obtained r's and z's (Each Computed for a Random Sample of 30 Seventh-Grade Pupils in a Single School) for Scores on Certain Tests in the "1939 Iowa Every-Pupil Tests of Basic Skills" in 71 Iowa Schools 227
16. Values of p/z for Various Values of p (Normal Curve of Unit Area and Unit S.D.) 243
17. Absolute Deviations $\frac{x}{\sigma}$ from the Mean of a Normal Distribution Exceeded by Given Proportions of the Cases 261
18. Table of Random Numbers 262

CHAPTER I

FUNDAMENTAL CONCEPTS IN SAMPLING THEORY

1. INTRODUCTORY

NEARLY all experimental research in education, and most of that which is not experimental in character, involves the drawing of inferences about a population from what is known of a sample taken to represent that population. Accordingly, one of the most important of the technical problems faced by the research worker is that of determining just how much may confidently be said about a population from what is known of a sample, or of ascertaining the degree of confidence which may be placed in the inferences drawn. Closely related to this is the equally important problem of how to select a sample, or how to plan an experiment, so that it may yield the most dependable or precise information about the population involved, and so that it will permit an objective and valid estimate of the degree of precision attained.

In recognition of their extreme importance in the training of the research worker in education, a major share of this text will be devoted to a detailed consideration of these problems. Particular attention will be given to the *design of experiments*, to *small sample theory*, and to the *testing of statistical hypotheses*. While the contributions to statistical theory which have recently been made in these areas appear to be of revolutionary significance in educational research, they have thus far been available to the research worker in education for the most part only in the literature of agricultural research and mathematical statistics. The language and the setting in which they have there been presented have proved inordinately difficult for the student of education to comprehend, and have perhaps seriously retarded their much needed introduction into educational research practices. It is accordingly one of the major purposes of this text to interpret these contributions in a language and notation familiar to the student of education, and to discuss and illustrate their possibilities with specific reference to the types of problems and materials with which he will have to deal.

It will be assumed in these discussions that the student is already acquainted with those aspects of sampling theory which are usually presented in introductory texts in educational statistics. More specifically, it will be assumed that he is familiar with and able to interpret the basic standard error formulas designed for large random samples for which the errors of sampling are normally distributed. Rather than rest too heavily on this assumption, however, this first chapter will in part be devoted to a brief review of some of the concepts and techniques with which the student is presumably already acquainted. This review will then be supplemented by a discussion of the limitations of these techniques in actual practice, and by an introductory consideration of certain additional concepts which are fundamental in the later discussions.

2. DEFINITIONS OF IMPORTANT TERMS

A *population* may be defined as any identifiable group of individuals, or as any collection or aggregate of comparable measures. A *sample* is any number of the members of a population that have been selected to represent that population. In ordinary usage, populations are usually thought of as consisting of human beings; in the statistical sense, populations may consist of any kind of members whatever. For example, the assessed valuations of real property in the rural school districts of Iowa may constitute a population, as may the numbers of books in the school libraries, or the ages of the pupils, or the years of experience of the teachers.

Populations may be either *finite* or *infinite*, either *real* or *hypothetical*. A finite population is one all members of which may be counted; an infinite population is one of unlimited size. For example, all *possible* weights of eight-year-old children in this country would constitute an infinite population, while the actual weights of the eight-year-old children now living in this country would constitute a finite population. The derivations of nearly all sampling error formulas assume infinite populations, whereas the populations to which the formulas are applied are usually finite. This assumption offers little difficulty, however, since the populations actually involved are usually so large that they may be considered as practically infinite.

A real population is one that actually exists; a hypothetical population is one that exists only in the imagination. Many of the populations involved in educational research are hypothetical. For example, an experiment may be conducted to determine the relative effectiveness of two methods of instruction. For the purposes of the experiment, two groups of seventh-grade pupils are selected, one of which is taught by Method A, the other by Method B, and at the close of the experiment comparable measures of achievement are secured for all pupils. In interpreting the results, the pupils who studied under Method B, for example, are considered as a sample from a population of seventh-grade pupils, all of whom had been taught by this method. Since the pupils in the experiment may be the only ones who have ever been taught by this method, this population is of course hypothetical. It is nevertheless useful to recognize that the method might produce different results if used with other seventh-grade pupils, and that the experimental results must therefore be considered as only a fallible indication of the results that would be generally attained. In some instances, we may wish to select a sample from a real population, but find it impracticable to secure an unbiased sample from that population. In that case we may use the sample that is available to us, "construct" a hypothetical population from which the given sample *might* have been drawn at random, and restrict our generalizations to that hypothetical population.

A *random sample* is one selected in such a fashion that every member of the population has an equal chance to be selected. This means that each member must be selected independently of all others. It is useful also to think of a random sample as one so drawn that all other *possible combinations* of an equal number of members from the population had an equal chance to constitute the sample drawn. Suppose, for example, that we are drawing a random sample of 300 cases from all high-school pupils in Indiana. There is, of course, an almost unlimited number of different combinations of 300 pupils in this population. One of these combinations, for instance, might consist of 2 pupils from Terre Haute, 13 from Lafayette, 276 from Indianapolis, and 9 from Gary. If our sampling is random, this particular combination must have the same chance of being selected as any other. Emphasis is placed on this latter concept of random sampling, since it indicates quite clearly that the samples used in educational research are seldom simple random samples. In practice, *accessibility* or *feasibility* are often determining factors in sampling. If we were actually drawing a sample of 300 pupils from Indiana high schools, under the methods usually employed the particular combination described above would have *no* chance of being drawn. The only feasible procedure would be to secure the co-operation of a few schools that together would provide the 300 cases needed; we could not expect to select each pupil independently from the whole population, and then use the pupils so selected regardless of how they were scattered throughout the state. The extreme significance of such practical obstacles to random sampling in educational research will be made clear in a later section. The procedure that may be followed to draw a random sample when no such obstacles exist is also to be explained later.

A *biased* sample is one so drawn that *in the long run* samples so selected will differ systematically from the population in the characteristic studied. Otherwise stated, a sample is biased if, when more cases are added by the same method of sampling, a given group character (such as the mean) will become more stable, but will tend to approach a value differing from the corresponding characteristic of the population. A sample may, of course, be biased with reference to one characteristic and unbiased with reference to another, if the two characteristics are entirely unrelated. A sample of all university sophomores selected from those taking a sophomore course in psychology, for example, might be biased with reference to vocational interests, but their mean age might be an unbiased estimate of the mean age of all sophomores in the university. Freedom from bias is one of the most important characteristics of a sample, and random sampling is one of the surest ways of obtaining it.

A *stratified* sample is one which may be subdivided into groups, each of which may be considered a sample from the corresponding subdivision of the entire population. For example, we might subdivide the entire population of adult males in the United States into various income groups and select a random sample (of any size) from each income group. The total sample thus secured would be considered a stratified sample. Again, in selecting a sample of schools in a given state, we might classify all schools according to enrollment, and select any desired number of schools from each enrollment classification. The numbers constituting the subgroups in a stratified sample are arbitrarily determined, and need not be proportional to the numbers in the corresponding subdivisions of the population. Stratified sampling has much the same advantages as controlled sampling, which are discussed in the following paragraphs. Methods of estimating the standard error of the mean of a stratified sample will be presented later (pages 157 ff.).

A *controlled* sample is one in which the selection is not left to chance, or not entirely to chance, but in which the distribution of some selected characteristic is *made* to conform to some predetermined proportion. It is a stratified sample in which the subgroup numbers are proportional to the corresponding numbers in the population. For example, we may wish to study (in a given population of school children) some trait, such as weight, that is known to be related to sex, and may wish to insure that our sample does not by chance contain an undue proportion of either sex. Assuming that there are equal numbers of boys and girls in the entire population, we might then select a certain number of boys at random from all boys and the same number of girls at random from all girls. In other words, we would *make* our sample repre-